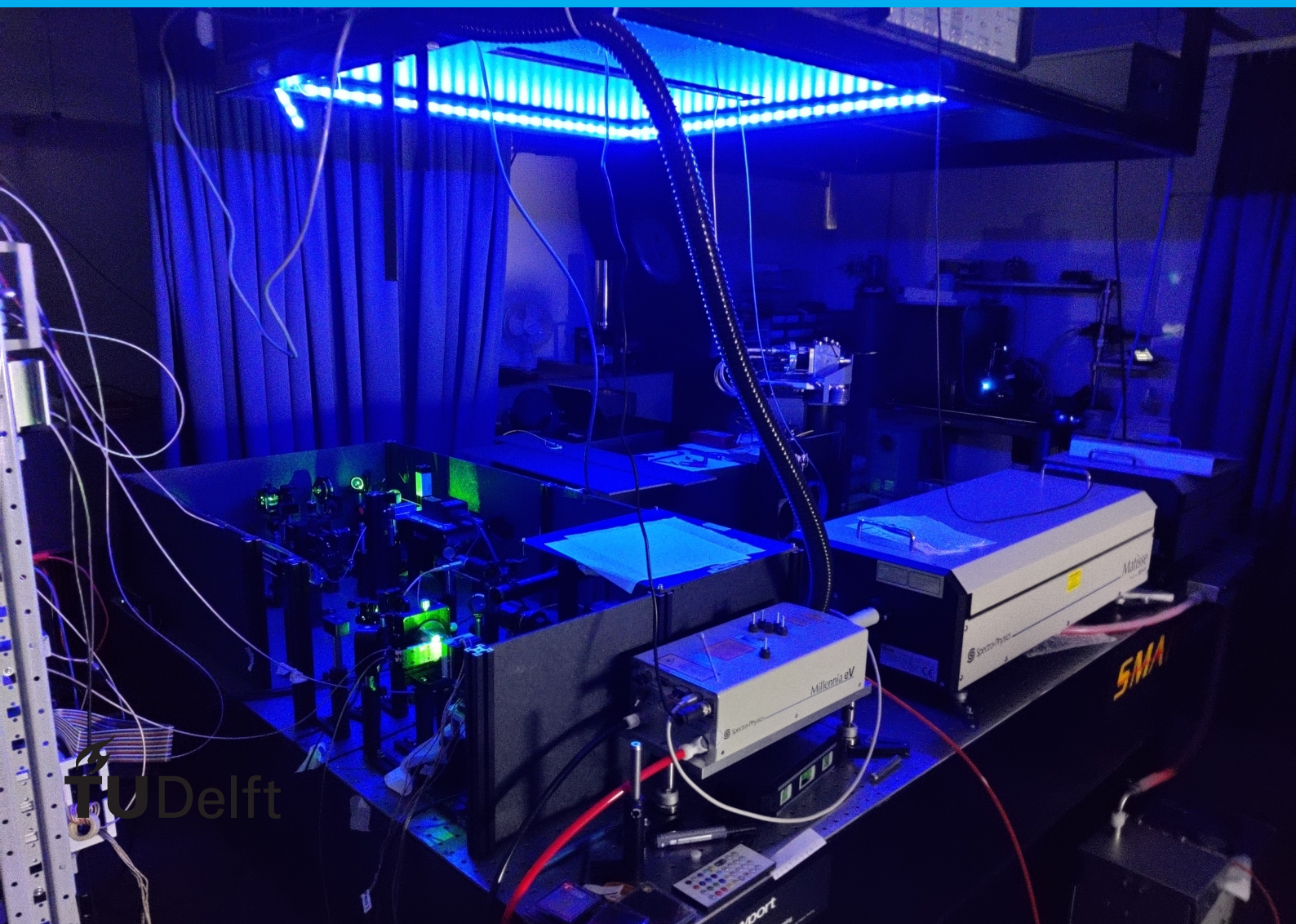
Characterization of Single Photon Avalanche Diodes and Integration with Diamond for Quantum Bio-sensing

Dylan Aliberti



Characterization of Single Photon Avalanche Diodes and Integration with Diamond for Quantum Bio-sensing

by

Dylan Aliberti

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended on Friday 22 November 2024.

Student number:	5170036	
Project duration:	January 1, 2024 - October 31, 2024	
Thesis committee:	Dr. R. Ishihara,	TU Delft, supervisor
	Dr. S. Goswami,	TU Delft
	Dr. S. Nur,	TU Delft
Daily supervisor:	I. Varveris	

TU Delft

Preface

I really liked this project. I got a lot of experience with physical work in the lab, and physical work can be very satisfying when things work. This is something I could do for the rest of my life. Do some lab work, accompanied by the occasional coding, and slowly building up systems for increasingly advanced experiments.

In order to find a master project, I was mostly looking through QuTech projects available, arguing that the field of quantum computers is suitable for me. Then I came across this project, which is about quantum sensing. It gave a nice perspective on how the same kind of technology can be used in things other than quantum computing. The context of specifically bio-sensing also spoke to me. Biology has something magical to it, and bringing that into images is always fun!

During the project, I realized that I am a ”Jack of all trades” kind of person, even more than I first thought. I made my hands dirty with all kinds of work. Besides the physics calculations, I had to design a PCB. It was my first PCB and I learned a ton about electronics. I even learned things in the domain of biology, as the project's end goal revolves around bio-sensing. I had to learn how to program an FPGA. This was a nice extension to my earlier programming experience. Before this project I had learned to write programs for both CPU, GPU and multiple CPUs together. The FPGA was an enjoyable next step, albeit technically being design rather than programming. I would say this project was a very good fit for me.

Apart from the content of the project, Yannis was an amazing daily supervisor. We had a lot of synergy and worked well together. Yannis is more than just a supervisor. We are friends now. A big thanks to Yannis for his support during the project! Also thank you to all the people from QIT, DEMO and QuTech who helped me with special tasks and advice! And not to forget my family and friends, who never stopped supporting and loving me.

A more elaborate acknowledgements section can be found towards the end of this thesis in Acknowledgements.

Dylan Aliberti
Delft, November 2024

Abstract

The quantum diamond microscope (QDM) is a device that uses nitrogen-vacancy (NV) color centers in diamond to detect magnetic fields using a technique called optically detected magnetic resonance (ODMR). NV centers are suitable for biological measurements because they can operate at room temperature and diamond is bio-compatible. This thesis works towards integrating this technology in a compact portable quantum bio-sensor, for which a chip with single photon detectors is a promising integration platform. Before this thesis started, a CMOS chip had been fabricated with single photon avalanche diodes (SPAD) on it including circuitry for digital communication. During this thesis, first, software was improved for an existing confocal ODMR setup, adding all necessary features for 2D scanning routines. Then, a new widefield setup was designed and established for testing the CMOS SPADs. Afterwards, the SPADs were electrically characterized and optically tested, the latter by using the widefield setup. The SPADs dit not function correctly and thus could only be partially characterized, leading to the CMOS design needing to be checked for errors. Finally, an attempt at bio-sensing was made. A sample was prepared by cultivating human embryonic kidney (HEK) cells on diamond, submersed in a 40 μg/mL solution of magnetic nanoparticles (MNP). During incubation the cells absorbed the MNPs, making them responsive to magnetism. This diamond has been scanned with the confocal setup while applying an external magnetic field of 1.07 mT. No HEK cells were found this way, even though the photo-luminescence intensity map hints at their presence within the scanning area, indicating that the external magnetic field and/or the concentration of the MNPs may be too low. More research is needed to confirm this.

Contents

1 Introduction

Today, many different forms of microscopy exist. Microscopes know many applications. One example is research on biological samples, which is essential for advancements in biological and medical science. In the context of bio-imaging, typically used microscope techniques are either optical, electron or fluorescence microscopy. For many biosamples, optical imaging is degraded due to autofluorescence and scattering in typically complex bodies (Glenn et al [12]). Instead, Glenn et al opted for using a quantum diamond microscope (QDM).

Glenn et al (2015) demonstrated magnetic bio-imaging by using a QDM, applied on immunomagnetically labeled cells. Another example of QDM usage is observing neuron pulses [14]. The QDM is a well-established method, used by many researchers, and is commercially available as a complete instrument [1].

Quantum diamond microscopy is a method that uses nitrogen-vacancy (NV) color centers in diamond to detect magnetic field strength. These NV centers have electronic quantum states that are sensitive to magnetic field strength, electric field strength, temperature and strain. This makes NV centers very versatile quantum sensors [11]. By providing the right excitations and collecting fluorescence light, one can gain information on the four previously mentioned aspects, which includes magnetic field strength, an essential quantity for bio-imaging.

1.1. Towards an integrated quantum diamond biosensor

QDM's are commonly made using an NV ensembles in diamond. Such ensembles collectively emit enough light to be detected by common photo sensors like charge-coupled device (CCD) [14] or photo diodes [21]. These methods can be used at room temperature and are thus suitable for bio-sensing on both dead and live cells. Note that room temperature is a key constraint here, as much more sensitive sensors are available at cryogenic temperatures. In order to improve magnetic field sensitivity for live bio-sensing, the sensitivity of the photo sensor at room temperature needs to be improved.

A promising candidate for improving sensitivity is the Single Photon Avalanche Diode (SPAD). This type of device can be used as a single photon counter, and works at room-temperature. It is compatible with CMOS technology, making it possible to fabricate using existing machines. This also allows for integrating additional circuitry together with the SPADs on the same chip. By optically coupling SPADs with single NV centers, higher magnetic sensitivity can be achieved compared to NV ensembles, as ensemble behavior broadens resonance peaks and reduces contrast in ODMR [18]. Such coupling would require fully integrating the sensor on a chip.

Here at the Quantum Integration Technology (QIT) / Ishihara lab, TU Delft, work is being done towards building such an integrated sensor. The concept consists of a CMOS chip with a grid of SPADs as well as all necessary circuitry to make them fully functional pixels. A diamond can be directly attached on top of it. Waveguides can be used to deliver excitation light directly to the NV centers. The diamond will contain a grid of single NV centers, for ultimate sensitivity. A biosample is grown directly on the diamond, making use of the bio-compatibility of diamond, and making sure the biofilm is as close as possible to the NV centers for good sensitivity. See figure 1.1.

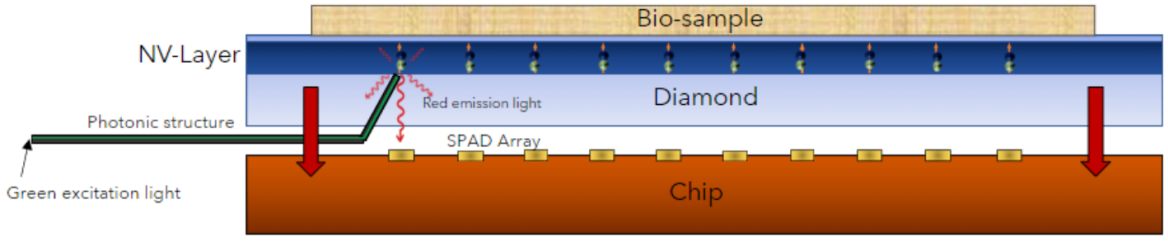


Figure 1.1: Concept of the sensor with diamond on a chip. The diamond has implanted NV centers at the top, and the biosample rests above the NV layer. Waveguides are crafted going through the diamond to deliver excitation light. The fluorescence light is captured by single photon avalanche diodes integrated on a CMOS chip. Image by I. Varveris

1.2. The starting point of this thesis

A CMOS chip with SPADs has already been designed and fabricated by earlier students [22] [4]. This chip contains 16 rows with 16 SPADs each, accompanied by digital circuitry, forming a 16x16 grid of SPADs ready for use. On top of that, three bare SPADs are included at the top for characterization. See figure 1.2.

Besides that, a confocal setup has already been built in the lab. This setup is tested and works, see Alfie's bachelor thesis [23]. Different diamonds, both with single NV centers, different sizes, and either with or without a bio-sample grown on top, are available to use for this thesis.

1.3. The goal

The end goal of this research is to develop an on-chip quantum image sensor with NV-centers in diamond integrated with SPADs for high-sensitivity and high-accuracy sensing of bio-samples. Towards this end, this thesis aims to achieve the following sub goals:

- **Improve ODMR software**
  Making the confocal ODMR setup fully functional and adding tweaks helps to familiarize with ODMR on NV centers at room-temperature and adds a reference when building a new setup.
- **Design a new setup for testing CMOS SPADs**
  With the obtained knowledge from tweaking the confocal ODMR setup, a new similar setup can be built for widefield ODMR, which serves to test the CMOS SPADs.
- **Characterize and test the CMOS SPADs**
  The SPADs need to be characterized both electrically and optically. The optical characterization is done with the widefield setup from the sub goal above. The characterization consists of measuring IV curves, CV curves, photon detection probability (PDP) and dark count rate (DCR).

The chip contains a range of different SPAD designs, so that they can be compared. Besides characterizing the SPADs, they need to be put to use in a test setup that performs bio-imaging, to demonstrate the working of the SPADs. During the course of this thesis, different components and setups have to be designed and developed in order to perform all tests and characterizations. This includes coding, designing a PCB to host the CMOS chip, and building/editing optical setups. Note that the waveguide idea from above is not part of this thesis, as it is unfeasible to iterate waveguides designs while also doing all of the above. Instead, the SPADs will be tested with an optical system to focus an image of the diamond onto the SPADs. The final milestone of the thesis is a demonstration of bio-imaging.

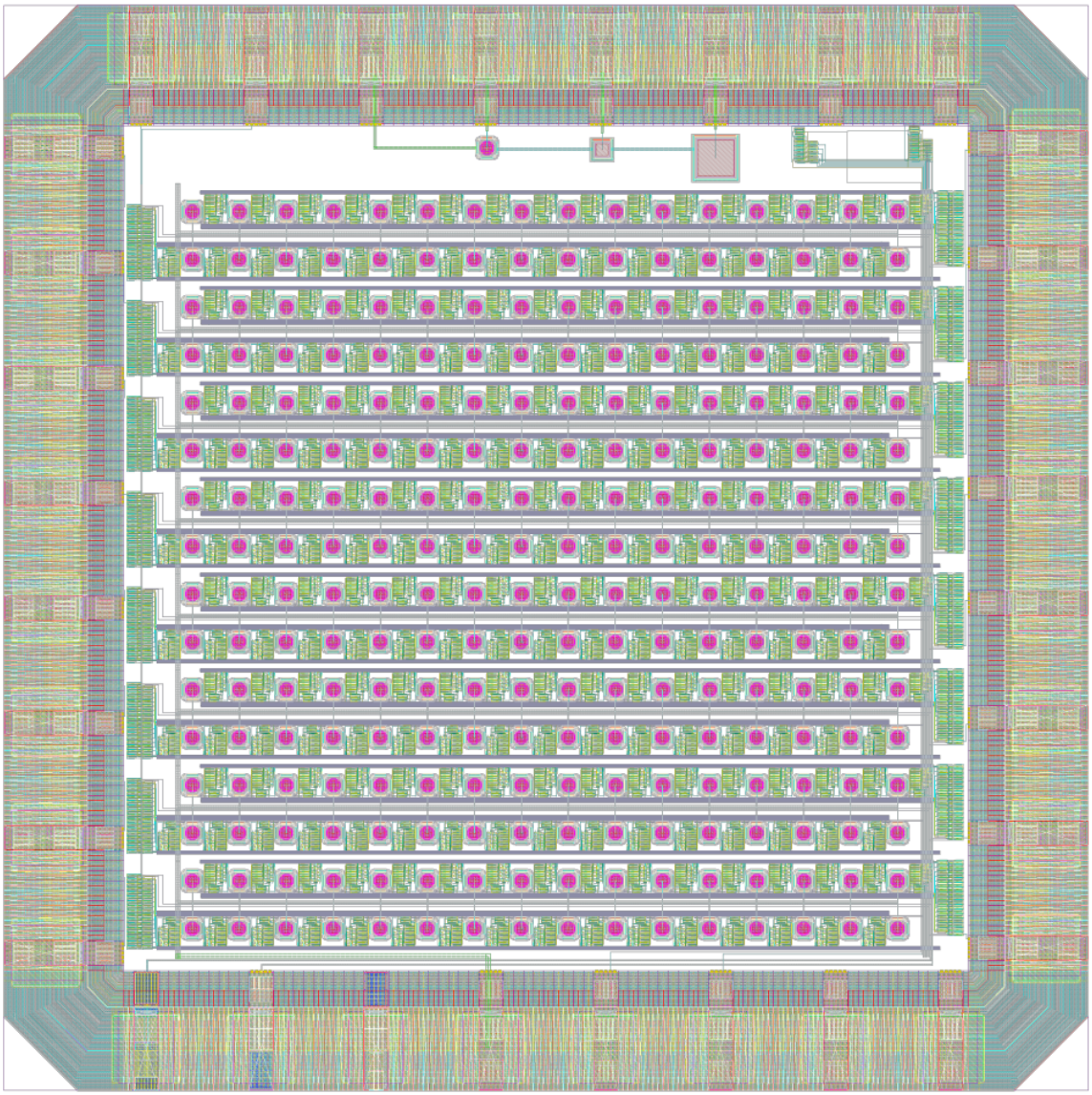


Figure 1.2: Full layout of the CMOS chip. It contains a 16x16 grid of single photon avalanche diodes, accompanied by the necessary digital circuitry to operate them end communicate the data. At the top, three bare SPADs are included for characterization. Taken from bachelor thesis [22].

2

Theory

The setup is essentially the quantum diamond microscope setup, but with single photon avalanche diodes (SPAD) instead of the more conventionally used charge-coupled device (CCD) imagers. Let us first take a look at how nitrogen-vacancy (NV) centers in diamonds work and how they can be utilized to obtain magnetic field information. Afterwards, let us see what other components are needed to build a functional quantum diamond microscope (QDM). Finally, let us consider the theory of SPADs, as this is the particular component of interest within the QDM setup.

2.1. The nitrogen-vacancy center

An NV center is a type of defect in a diamond lattice, where a nitrogen atom and a vacancy occupy the positions where two carbon atoms are supposed to be. See figure 2.1. The complex forms 4 covalent bonds with neighboring carbon atoms. In its neutral state, NV^0, it contains one free electron. It is able to absorb one electron from the conduction band, forming NV^-. This way, it has two free electrons, forming a spin-1 system. In the rest of this thesis, NV^- will simply be referred to as ”NV” for brevity.

2.1.1. Level structure

Figure 2.2 shows the energy levels of NV centers. Some transitions are highlighted that are needed for optically detected magnetic resonance, which is addressed in the next section. The ground state is the ${}^3A_2$ state, which will be denoted $|m_s, g\rangle$, where m_s represents the spin of the system, which can be -1, 0 or +1. Similarly, lets us write the excited state ${}^3E$ as $|m_s, e\rangle$.

The ground state is worked out in more finer levels in figure 2.3. The bottom of the figure also shows corresponding frequency sweeps, which will be explained in the next section. The effect of magnetic field and temperature is noted in the diagram, along with the zero-field splitting of 2.87 GHz and nuclear splitting, which depends on the isotope.

The NV center has good properties that make it a very versatile quantum sensor. It can be used at room temperature, and it is sensitive to temperature, strain, electric field and magnetic field. This can be explained by different contributions to the Hamiltonian, which is addressed in the next subsection.

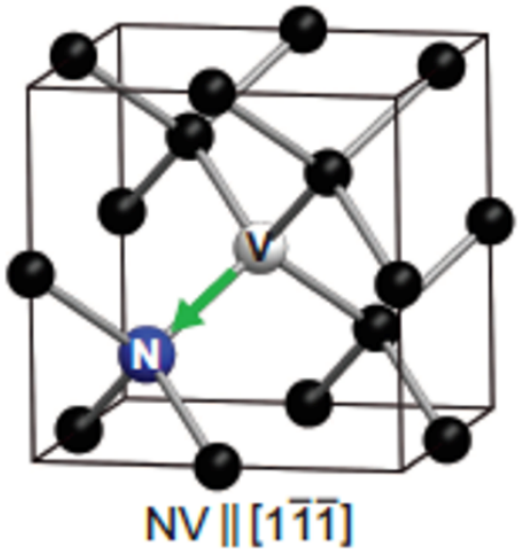


Figure 2.1: NV center in a diamond unit cell. A nitrogen atom and a vacancy (empty space) are located at the places where two carbon atoms are supposed to be in the lattice. The green arrow represents the intrinsic quantization axis of the NV center. Image taken from QNami [11]

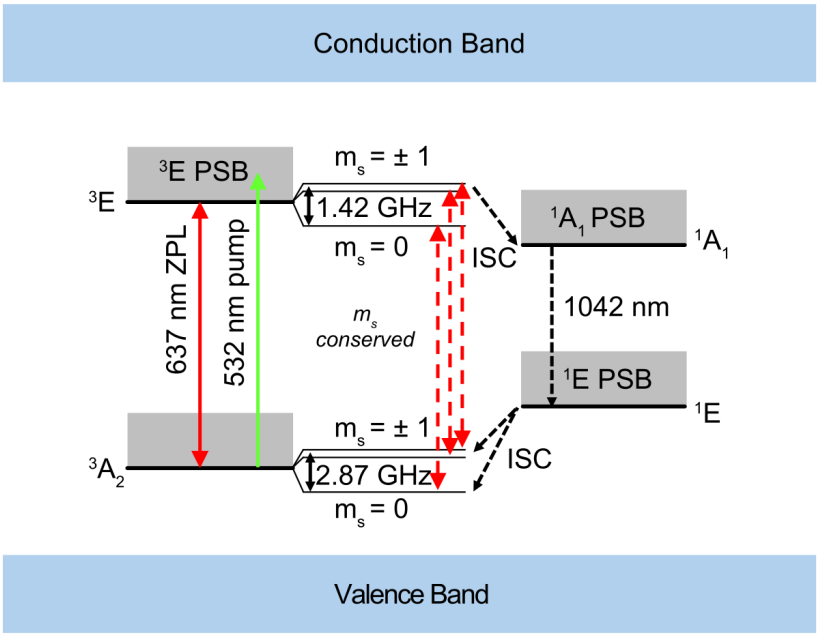


Figure 2.2: NV transition diagram. Starting from the $^3A_2$ state, the electron is pumped to the ^3E state. From there, if it has spin zero, the only way back is by emitting a red photon, getting it straight back to the $^3A_2$ state. If the electron has spin ±1, then it has an extra possible route through other states by non-radiative intersystem crossings (ISC). Due to this extra option, less of the spin ±1 states end up emitting red photons, resulting in a contrast difference with respect to the spin 0 electrons. Image taken from Levine et al [18].

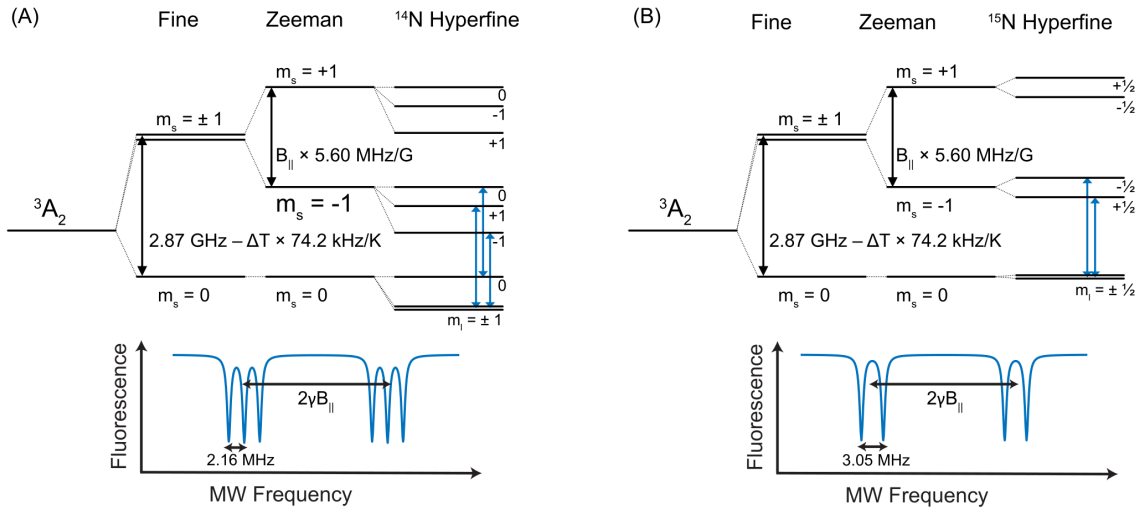


Figure 2.3: Energy level structure of NV centers for both nitrogen-14 and nitrogen-15 isotopes. On the bottom, the corresponding photoluminescence spectra of the microwave sweeps are shown. The different isotopes have different amounts of hyperfine splitting. The first order dependency of the energy levels on magnetic field and temperature are shown. Image taken from Levine et al [18]

2.1.2. Ground state Hamiltonian

The NV center's states are sensitive to many variables. equation 2.1 gives a summarized and simplified Hamiltonian that captures all relevant dynamics for quantum diamond microscopy [18]. The axes are chosen so that the z-axis is aligned with the NV's intrinsic axis. The paragraphs below explain where each term comes from.

$$\begin{aligned}\frac{\hat{H}}{h} =& \frac{g_e \mu_B}{h}(B_x\hat{S}_x + B_y\hat{S}_y + B_y\hat{S}_y) && \text{Magnetic Terms}\\ &+ d_{\parallel}E_{tot,z}\hat{S}_z^2 - d_{\perp}E_{tot,x}(\hat{S}_x^2 - \hat{S}_y^2) + d_{\perp}E_{tot,y}(\hat{S}_x\hat{S}_y + \hat{S}_y\hat{S}_x) && \text{Electric Terms}\\ &+ M_z\hat{S}_z^2 - M_x(\hat{S}_x^2 - \hat{S}_y^2) + M_y(\hat{S}_x\hat{S}_y + \hat{S}_y\hat{S}_x) && \text{Strain Terms}\\ &+ D(T)\hat{S}_z^2 && \text{ZFS, Temperature}\end{aligned} \tag{2.1}$$

Intrinsic interactions

The intrinsic interactions consist of zero-field splitting and hyperfine splitting. As the NV center is a spin-1 system, it is possible for the electrons to have the same spin, resulting in spin-spin interaction. This results in the zero-field splitting between the $|0, g\rangle$ and the $|\pm 1, g\rangle$ states, which is about 2.87 GHz. The zero-field splitting $D(T)$ is also dependent on temperature, as the spin-spin interactions scale with lattice size. Furthermore, interaction with the nucleus results in hyperfine splitting of 2.16 MHz and 3 MHz for nitrogen-14 and nitrogen-15 respectively. As the hyperfine splitting is fixed, it can be left out for the purpose of modeling NV dynamics in the context of quantum diamond microscopes.

External influences

The effect of an external electric field can be modeled by the Stark effect. The influence of strain can be modeled as an external electric field and thus follows a similar formula with different constants. Lastly, the effect of a magnetic field can be modeled by Zeeman splitting. A detailed explanation of how these three effects work on the NV center can be found in Doherty et all [7].

2.2. Optically detected magnetic resonance (ODMR)

2.2.1. Driving a cycle of transitions

In order to utilize the NV center for magneto-sensing, transitions need to be pumped. See figure 2.2. The NV center is excited using a green laser with wavelength 532 nm. If the magnetic moment m_s of the spin is zero, then the only way for the NV center to decay back is by emitting red fluorescence light around 637 nm. When $m_s = \pm 1$, then the NV center has an additional decay option through non-radiative intersystem crossings (ISC), passing through two other states without emitting visible light, and going back to the ground state without preserving spin. Due to this alternative decay route, decay of the $m_s = \pm 1$ releases less photons. The contrast in photoluminescence intensity can be up to 20% [8].

2.2.2. Driving spin transitions with microwaves

Transitions between the $|\pm 1, g\rangle$ and $|0, g\rangle$ states can be driven by applying microwaves around 2.87 GHz. When the NV center is simultaneously driven with 2.87 GHz microwaves and 532 nm excitation light, the fluorescence light becomes less than when only the 532 excitation light is applied, due to some of the $|\pm 1, g\rangle$ decaying through the non-visible route. When the microwave frequency is swept, this results in a visible dip around 2.87 GHz. Due to hyperfine splitting, the dip may actually appear as multiple peaks closely together. See figure 2.3.

2.2.3. From microwave sweep to quantifying variables

By sweeping microwaves, the energy levels of the $|\pm 1, g\rangle$ states can be accurately determined with respect to the $|0, g\rangle$ state. As we have seen in the Hamiltonian, the energy levels can be influenced by a number of external variables. Hence, external influences will show up in the microwave sweep by the dips being displaced. This can be compared to a reference with zero external influence, so that the difference in energy can be quantified and interpreted.

Magnetic field sensing

For magneto-sensing, the $|\pm 1, g\rangle$ states are expected to split due to Zeeman splitting, which is described with the Hamiltonian

$$\frac{\hat{H}}{h} = \frac{g_e \mu_B}{h}\left(\mathbf{B}^\dagger \mathbf{S}\right) = \bar{\gamma}_{NV} \mathbf{B}^\dagger \mathbf{S} \quad , \tag{2.2}$$

where h is the Planck constant, $g_e = 2.003$ the NV's electronic Landé factor, μ_B the Bohr magneton, and $\mathbf{S}^\dagger = (S_x, S_y, S_z)$ the dimensionless electronic spin-1 operator [11]. $\bar{\gamma}_{NV} = 28$ MHz/mT is the reduced gyromagnetic ratio of NV centers, which is more convenient to use in this case. The splitting of the $|+1, g\rangle$ and $|-1, g\rangle$ peaks scales with the axial component of the magnetic field. From equation 2.2 it can be derived that, for a given frequency splitting ν, the axial component of the magnetic field can be calculated by

$$B_\parallel = \frac{\Delta\nu}{2\bar{\gamma}_{NV}} \quad . \tag{2.3}$$

Note that Zeeman splitting is not the only term affecting the energy levels of the $|+1, g\rangle$ and $|-1, g\rangle$ states, see equation 2.1. Thus, by using equation 2.3, we assume that electric field and strain are negligible.

Magnetic field sensitivity

Magnetic field sensitivity is defined as the smallest possible difference in magnetic field strength δB that is distinguishable by the quantum diamond microscope in one second. A value being the smallest distinguishable is defined as having a signal-to-noise ratio of 1. Measurements of different durations have to be normalized. δB scales with $\sqrt{t_{meas}}$ where t_{meas} is the measurement duration. Hence, the sensitivity η_B of a QDM is defined as [18]

$$\eta_B = \delta B \sqrt{t_{meas}} \quad . \tag{2.4}$$

The magnetic field sensitivity can also be estimated from the contrast C, full width half maximum (FWHM) $\Delta\nu$ and count rate R of the ODMR dips when the sensitivity is limited by shot-noise. Let h

be the Planck constant, g_e the NV electronic g-factor, and μ_B the Bohr magneton. Then the shot-noise-limited magnetic field sensitivity can be approximated by [2]

$$\eta_B = \frac{4}{3\sqrt{3}} \frac{h}{g_e \mu_B} \frac{\Delta\nu}{C\sqrt{R}} \quad . \tag{2.5}$$

Here, the factor $\frac{4}{3\sqrt{3}}$ comes from the steepest slope in the standard Lorentzian curve. Recall that the NV gyromagnetic ratio $\bar{\gamma}_{NV}$ is defined by

$$\bar{\gamma}_{NV} = \frac{g_e \mu_B}{h} \quad . \tag{2.6}$$

Thus, the approximation for magnetic sensitivity can be simplified to

$$\eta_B = \frac{4}{3\sqrt{3}} \frac{1}{\bar{\gamma}_{NV}} \frac{\Delta\nu}{C\sqrt{R}} \quad . \tag{2.7}$$

2.3. Single photon avalanche diodes (SPAD)

To understand what a SPAD is, lets first take a look at the Avalanche PhotoDiode (APD). This is a p-n junction that is biased near its breakdown voltage. The ionization caused by the high electric field amplifies photon-generated charge carriers, resulting in avalanches.

A SPAD is an avalanche photodiode that is operated in reverse bias beyond breakdown voltage [10]. When operating in this voltage region, ionization can occur when a photon hits the photodiode, which frees up a charge carrier. Such carriers can free up more carriers, leading to an avalanche. An increasingly large current will flow until either the SPAD is destroyed by its own heat, or is quenched by an external circuit. Figure 2.5 illustrates this principle. Due to it's similarity with a Geiger counter, SPADs are sometimes also called Geiger-Mode Avalanche Photodiodes. The cross-section of a common SPAD design is shown in figure 2.4. This is also the design that is used in the CMOS chip that needs to be characterized, see chapter 1.

2.3.1. Quenching

When an avalanche occurs in the SPAD, the current is self-sustaining. This current needs to be stopped before the SPAD is ready to detect a new photon. The process of stopping the current every time an avalanche occurs is called quenching. Quenching can be done either passively or actively [15]. A simple passive quenching circuit is a resistor in series with the SPAD. The resistor needs to have a resistance R large enough such that V_{over}/R is smaller than the latch current of the SPAD [24], which is the minimum current to sustain an avalanche. V_{over} is the overvoltage defined as $V_R - V_{BR}$, where V_R is the reverse bias voltage applied on the SPAD and V_{BR} is the breakdown voltage. Active quenching has the potential to quench the SPAD faster than passive quenching, resulting in a higher maximum count rate. Such a circuit is more complicated in design.

In the CMOS chip with SPADs to be tested, see chapter 1, the arrays of SPADs are actively quenched. The analog SPADs on the top are passively quenched.

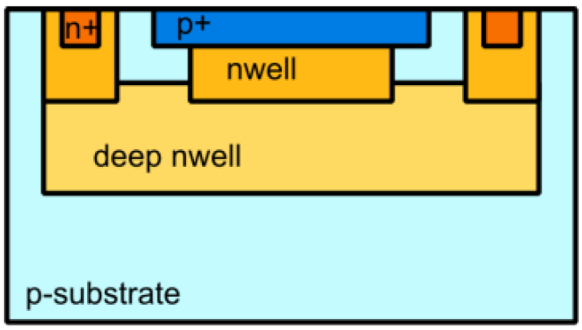


Figure 2.4: A common SPAD design. It uses a p+/nwell junction. A guard ring is included on the sides, which is separated from the p-substrate by a deep nwell. A guard ring helps to electrically isolate the SPAD from its surroundings and prevent premature edge breakdown [10]. The p+ region is the anode and the n+ region forms the cathode. Image taken from Prenzoni et all [20].

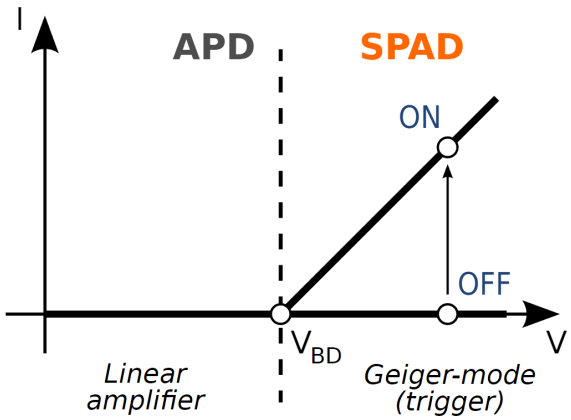


Figure 2.5: IV diagram of APD and SPADs. When the bias voltage is tuned beyond breakdown voltage, the APD turns into a SPAD. An effect called bifurcation starts to work. With bifurcation is meant that the SPAD has two branches in the IV diagram. When voltage is tuned beyond breakdown, no avalanche is yet occurring. Only after the first avalanche has been triggered, which can be a photon event or thermal noise, the SPAD jumps to the ”on” branch in the IV diagram. It stays there until it is quenched. To quench the SPAD, the voltage has to be temporarily turned below breakdown voltage. When the voltage is turned up again, the SPAD is ready to detect the next photon. Image taken from Wikipedia [6]

2.3.2. Characteristics

The performance of a SPAD depends on certain characteristics, which are the following:

- **Dead time**
  When a photon triggers an avalanche is in a SPAD, it needs to be quenched before it is ready for the next photon. During the quenching process, the SPAD will either not be sensitive to photons or have a lowered probability of triggering. The dead time directly determines the maximum count rate that the SPAD is able to detect. The quenching circuit is a key factor in the dead time [10].

- **Junction capacitance**
  The junction capacitance is the capacitance measured between the two ends of the SPAD. This capacitance is relevant for the recharge time of the SPAD and quenching. In passive quenching, the dead time is dominated by the time constant RC of the SPAD. Reducing the capacitance thus reduces the dead time [27].

- **Dark count rate**
  Sometimes an avalanche is triggered by a dark carrier instead of a photon. Such dark carriers appear due to thermal generation, which is dependent on temperature and operating voltage of the SPAD [26].

- **Photon detection probability**
  The probability that a photon in the active region of the SPAD triggers an avalanche is called photon detection probability. Not all photons that enter the SPADs active area trigger an avalanche. This is because part of the photons get reflected, and part of the ionization impacts do not result in an avalanche due to the charge carriers recombining before multiplication [17].

- **Time jitter**
  Time jitter is defined as the uncertainty in the duration from the moment a photon arrives to the moment that the avalanche is triggered. This uncertainty is important in applications that use the precise timing of SPAD pulses to do time-correlated single photon counting (TCSPC), e.g. in single photon telecommunication [13]. As jitter is not relevant for sensing applications that collect many photons over a collection time, measuring this characteristic is outside the scope of this thesis.

- **Afterpulsing**
  Afterpulsing is correlated noise. When an avalanche is quenched, there is a chance that a charge carrier gets trapped. When the SPAD is back beyond breakdown voltage, this carrier could cause

an unwanted avalanche. Research has been done on filtering out afterpulsing, as afterpulses can be recognized by their correlation with genuine photon pulses [9].

2.3.3. Why CMOS SPADs

For the purpose of building an integrated quantum bio-sensor, the photo-detectors need to operate at room temperature, to accommodate in-vivo experiments. CMOS is also a well-developped technology, making it a relatively inexpensive way of mass-producing SPADs. Furthermore, circuitry can be included on the CMOS chip, as CMOS technology has all necessary accomodations to make electrical circuitry. Thus, CMOS SPADs are a good candidate for building the integrated bio-sensor. The SPAD chip of this thesis to be characterized also contains a generous amount of circuitry that manage active quenching, analog-to-digital conversion and multi-channel serial communication for FGPA's.

3

Experimental setup

The work of this thesis has been done in both a confocal setup and a widefield setup. The widefield setup is just meant to test the SPAD chip and does not have the diamond directly on top of the chip. As mentioned in chapter 1, the final goal is to make an integrated quantum diamond-on-chip sensor, and the SPAD chip is one of its components that need to be tested.

This chapter will first introduce the specific diamonds and SPAD used in this thesis. Afterwards, the setup schematics are presented.

3.1. Diamond samples

During the course of this thesis, measurements have been done on many diamonds. In chapter 5, only a selection of measurements is shown to demonstrate the working of relevant processes. The diamonds that were used for these measurements are listed below.

- **Triangular sample**
  This diamond is the result of a failed attempt to bond it to SIO_2. The diamond broke in a triangular piece, and it has trapped particles underneath.
  Details: Provided by Almax, general grade, 10 micron thick, roughly 4x6mm^2, orientation (100), [N] $\sim$ 1 ppm. Bonded on SiO_2/Si film.

- **Etched sample**
  A diamond that was etched from 300 to 257 micron thick to make it thin enough for measurements on the back surface of the diamond. The diamond has to be thin enough so that the top surface does not hit the microwave antenna. This diamond was going to be bonded on SiO_2/Si film in the future, but problems were encountered.
  Details: Provided by EDP, general grade, 257 micron thick, orientation (111), [N] $\sim$ 6 - 8 ppm.

- **Bio-sample**
  The diamond contains magnetically labeled human embryonic kidney (HEK) cells grown on a diamond submersed in 40 μg/mL solution of magnetic nanoparticles (MNP). This is further explained in the following subsection 3.1.1.
  Details: Provided by EDP, 300 micron thick, 10x10mm^2, orientation (100), [N] $\sim$ 6 - 8 ppm.

- **Maarten's sample**
  A diamond that was already in the lab's inventory at the start of this thesis. Despite being old and dirty, it still gives a good amount of counts of about 1 Mcounts/s at 1 mW laser power.
  Details: Provided by CCL, 300 micron thick, 4x4mm^2, orientation (100), [N] unknown, estimated to be between 5 - 10 ppm.

- **Blank biosample**
  An older bio-sample with a blank solution (meaning no cells). This is from a different batch than the newer bio-sample from above in this list.
  Details: Provided by EDP, 300 micron thick, 3x3mm^2, orientation (111), [N] $\sim$ 6 - 8 ppm

- **The EDP sample**
  A diamond glued on a 1 inch silicon plate, which is used in the widefield testing setup.
  Details: Provided by EDP, 300 micron thick, 10x10mm^2, orientation (100), [N] $\sim$ 6 - 8 ppm

3.1.1. Preparation of the biosample

The preparation of the biosample was done by Qiangrui Dong, from the group of ImPhys of TNW, under the supervision of Daan Brinks. The biosample was prepared by growing human embryonic kidney cells of type 293T (HEK293T cells) directly on diamond. Diamond is bio-compatible and thus will not cause harm to the cells. During incubation, the diamond was submersed in a 40 μg/mL solution of magnetic nanoparticles (MNP). The MNPs are small enough ($\sim$ 5 nm) such that they can be absorbed through the cell membranes. After incubation, the MNP solution is removed from the diamond and the sample is dried, thus killing the cells. The result is a diamond sample with a layer of HEK cells on top of it, ready to be tested under the quantum diamond microscope. More about this process can be found in Wei et all [25].

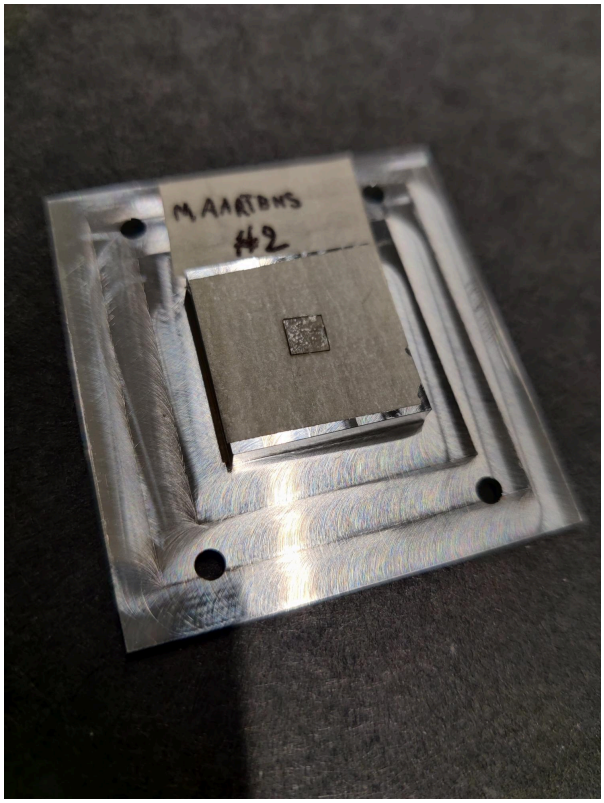


Figure 3.1: Photo of one diamond sample on the typical mounting plates that are used in the lab. Normally no tape is applied between the diamond and the metal.

3.2. The SPADs

Of the CMOS design by the previous bachelor students, four variants were made [22]. These four variants each contain three different SPADs at the top that are directly accessible through the bonding pads, meant to be characterized. The table below gives an overview of all the SPAD variants to be tested.

Chip	SPAD 1	SPAD 2	SPAD 3
Chip 1	round, 12 μm diameter	square, 12 μm size	square, 36 μm size
Chip 2	round, 4 μm diameter	round, 14 μm diameter	square, 16 μm size
Chip 3	round, 18 μm diameter	round, 20 μm diameter	square, 26 μm size
Chip 4	round, 8 μm diameter	round, 10 μm diameter	square, smaller build, 10 μm size.

Table 3.1: Overview of all SPAD variants to be tested.

3.3. Confocal ODMR setup

The confocal setup already existed when this thesis project started, and improvements were done only on the software side. In order to excite the diamond, a 532 nm laser is focused on it. The red fluorescence light, which is mostly between 600 - 800 nm, is separated from the green light using a dichroic mirror. It goes through an additional filter that is only transparent above 633 nm, and is then guided into an SPCM (Single Photon Counting Module). ODMR is performed by applying a microwave on the diamond, which is swept over frequency, and counting the photons at each frequency. Figure 3.2 shows a schematic of the confocal setup. Figure 3.3 as well as the title page show the real confocal setup in the lab. An overview of the devices used to control the setup is given by figure 3.4.

Note that unwanted light coming from the wrong angle is filtered out by having a single-mode fiber. Only the transverse mode can pass through, which corresponds to a photon with a Gaussian profile moving parallel through the fiber. Only photons that enter the fiber straight have a high probability to enter the single mode fiber. The same effect can be achieved by using a pinhole and two lenses, but this method is harder to focus and align.

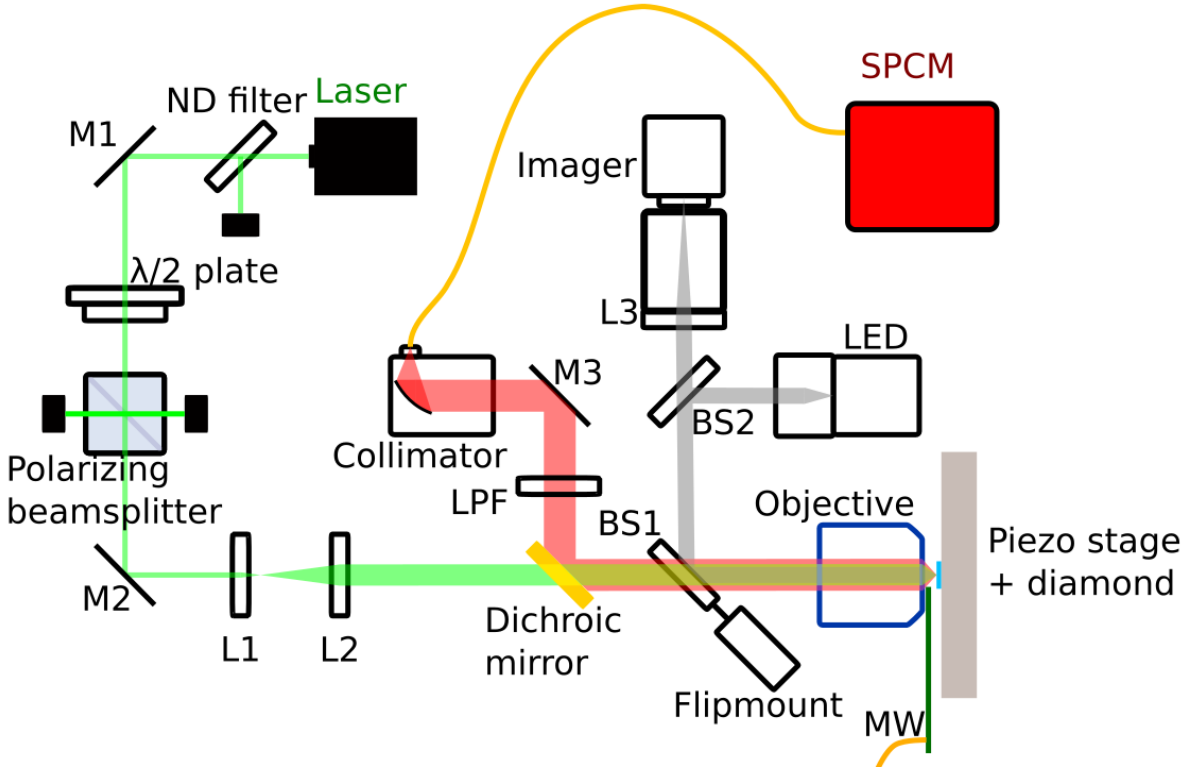


Figure 3.2: Schematic of the confocal ODMR setup (top view). The left half has the sole purpose of getting a laser beam with the right beamsize, power, offset and direction. The Natural Density (ND) filter is used to attenuate the laser by two orders of magnitude. The combination of the halfwave (λ / 2) plate and polarizing beamsplitter serves as a way to finely tune the laser power. The combination of L1 and L2 serves as a beam expander.
The right half of the schematic shows both the path of the fluorescence light (red), and components to see the diamond (represented with gray beams). The fluorescence light is reflected by the dichroic mirror into a long pass filter (LPF). Although the dichroic mirror already effectively filters the red light, the LPF serves as an extra measure to further dampen any scattered or reflected green light. The filtered light is finally collected into a fiber by the collimator. By using a single-mode fiber, unwanted light coming from the wrong angle is filtered.
BS1 is a 45:55 R:T pellico beamsplitter. Notice that it is mounted on a flipmount. When a diamond needs to be mounted, the imager and LED are used to observe a digital image of the diamond through the objective, so that it can be aligned. With BS1 lowered, the imager and LED reach the diamond through BS2, which is a 10:90 R:T beamsplitter. The laser spot is then visible on the diamond. Before starting a measurement, BS1 is lifted from the laser beam, so that the full amount of fluorescence photons reach the SPCM. A microwave antenna is mounted between the objective and the piezo stage to deliver microwaves.

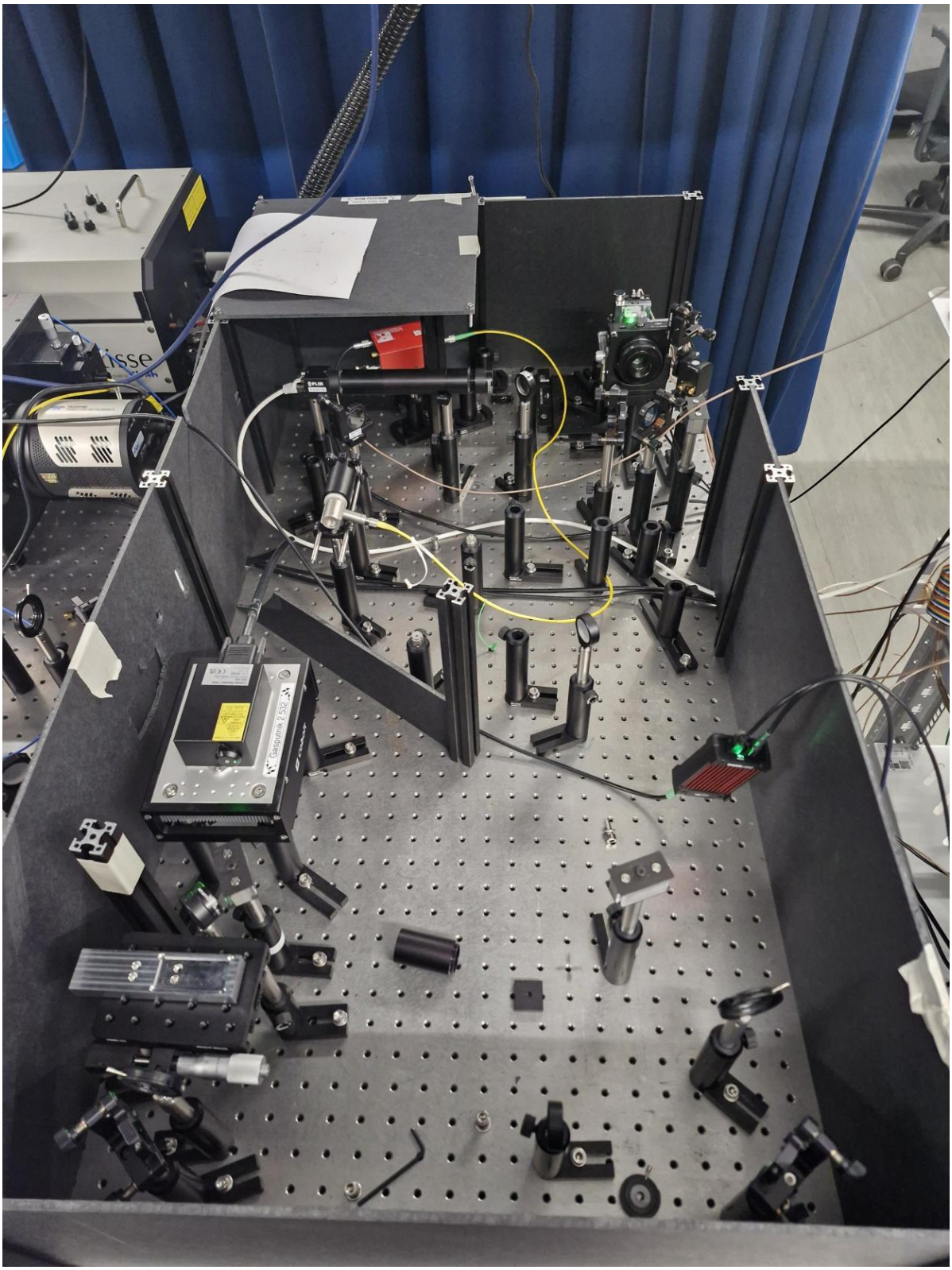


Figure 3.3: A photo of the confocal setup in the lab. The relevant components are mostly at the edges of the box, and the clutter in the middle can be ignored for the sake of this thesis.
The laser is the black box on a heatsink on the left. The objective is at the top right of the image. The single photon counting module (SPCM) is the red box at the top left, which is fiber coupled with the collimator. The beam expander was not installed when this photo was taken.

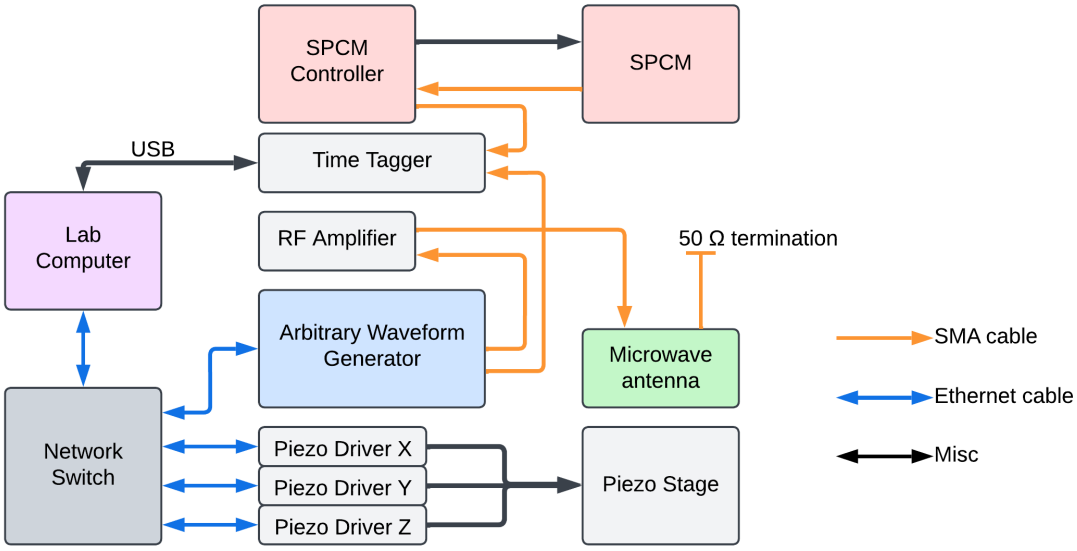


Figure 3.4: An overview of the electronic devices used to drive the confocal ODMR setup. The Arbitrary Waveform Generator (AWG) Generates microwaves which get amplified and emitted by the microwave antenna. The AWG also sends markers to the Time Tagger to signal when to start keeping track of counts. The Single Photon Counting Module (SPCM) is driven by the SPCM Controller, which passes all the counts towards the Time Tagger.

3.3.1. Magnetic shielding

During the experiments it was discovered that the piezo stage has a magnetic field of its own that causes Zeeman splitting of about 10 - 15 MHz. Simple rectangular magnetic shields were made to mitigate this, which is shown by figure 3.5. The results are shown in chapter 5. These shields are mounted as an extra plate between the piezo stage and the mounting plate with the diamond on it.

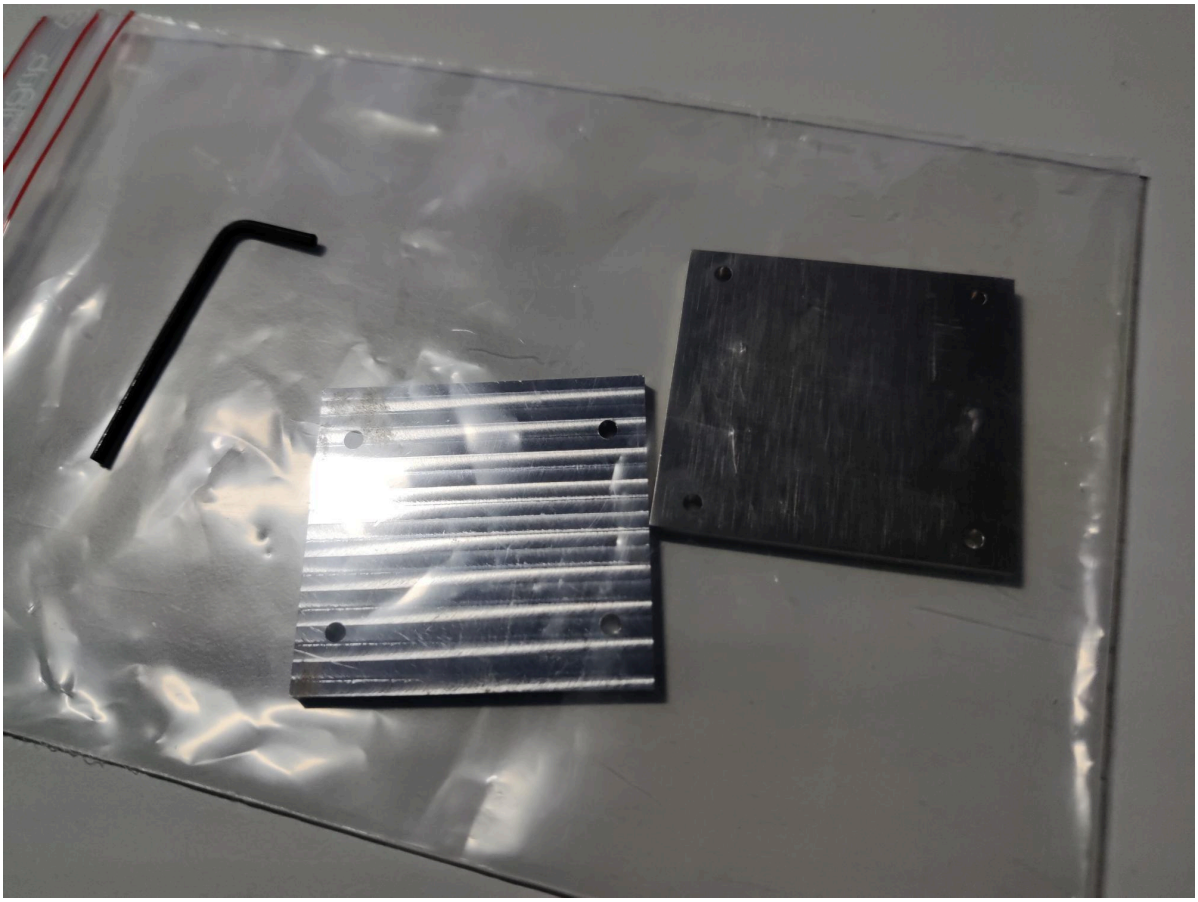


Figure 3.5: Photo of the magnetic shields. Left: Steel. Right: MuMetal.

3.4. Widefield test setup for the SPAD chip

In order to test the SPAD chip, a new setup had to be built. The chip contains a surface with pixels. A square image on the diamond should get imaged onto the surface of the chip, allowing to perform ODMR on many pixels simultaneously. The chip is wire bonded on a PCB that handles all connections between the chip and the FPGA and power supplies. Figure 3.6 shows a schematic of this new setup. It can be considered a miniature version of the confocal setup (see figure 3.2), but with the reverse kind of dichroic mirror, and with the laser beam focused in such a way that it lights up a circular surface on the diamond. An overview of the devices used to control this setup is given by figure 3.7.

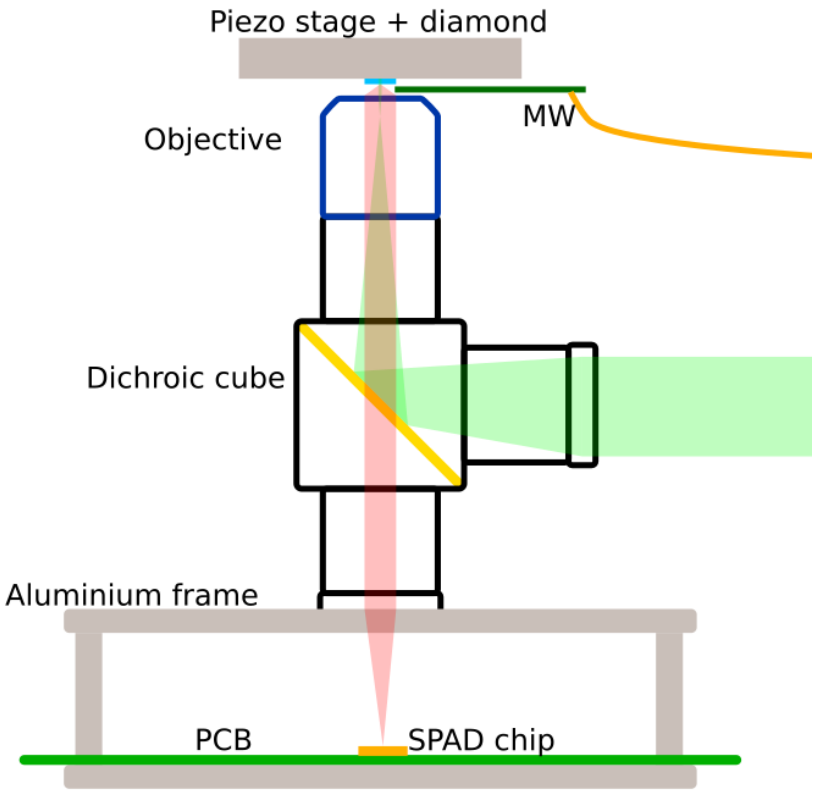


Figure 3.6: Schematic of the widefield setup for testing the SPAD chip (side view). It uses the same principle as the confocal setup, where a dichroic mirror is used to separate (green) excitation light and (red) fluorescence light, while they both go through the same objective. The difference is that the green laser beam has an extra lens that focuses it right before the objective, such that it hits the diamond as a collimated beam. This way, a circular area of the diamond is excited as opposed to one single spot, which a confocal setup does. A microwave antenna rests on the objective to deliver microwaves to the diamond.
Note that the green laser beam needs to be expanded and attenuated before entering the depicted optical subsystem. This is done in a similar way as in figure 3.2, but has been omitted in this schematic.

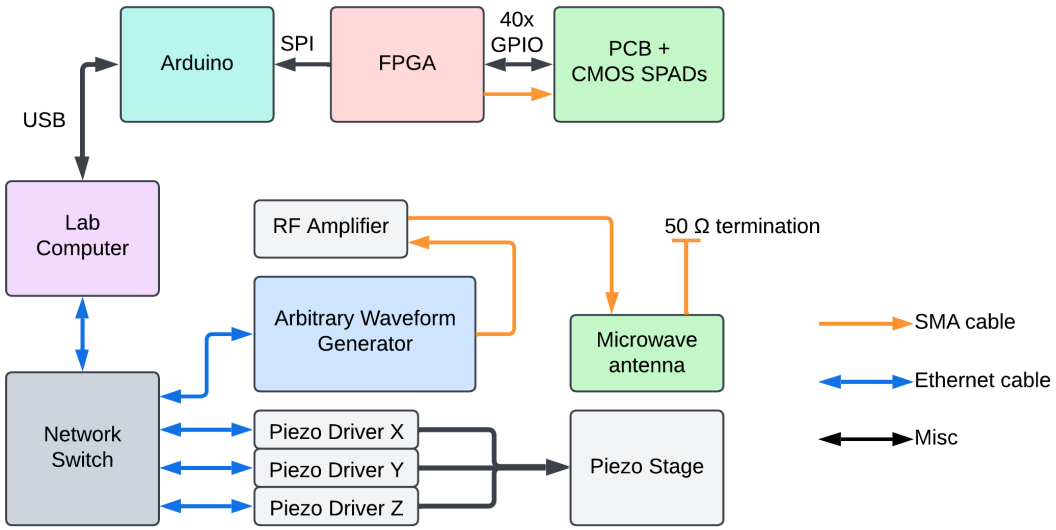


Figure 3.7: An overview of the electronic devices used in the widefield testing setup. Compared to the schematic of the confocal setup in figure 3.4, the Single Photon Counting Module (SPCM) and Time Tagger have been replaced with the CMOS-PCB-FPGA-Arduino chain. This chain is more elaborately described physically in section 3.6 and on the software side in chapter 4.

3.5. 2D imaging

How do we go from being able to measure counts to performing ODMR and extracting 2D images? In order to do ODMR, a microwave field needs to be applied, which is swept over frequency. The sweep is done from 2.85 - 2.89 GHz. For most measurements done during the course of this thesis, this frequency range was just enough to fully contain the dip with a little margin so that a double Lorentzian dip can be fitted robustly. For each frequency, the counts are measured for 0.2 seconds. When plotting count rate vs frequency, Lorentzian dips can be seen where electron spin resonance occurs with the microwave. By fitting a double Lorentzian dip, the following values can be extracted:

- Contrast
- Peak splitting
- Frequency shift

These values are then used to further process into either strain or magnetic field, depending on which measurement is being done.

In order to obtain 2D images of the diamond, the ODMR measurement needs to be performed on many pixels. The confocal setup can only measure one spot at a time, and thus the piezo stage is used to move in a grid, doing each measurement sequentially. The widefield setup can inherently measure many pixels simultaneously, so moving the piezo stage is only needed to either bring the diamond in focus, or move between different areas. The measured ODMR data will be of shape (M, N, F) with M and N the number of pixels in the x and y directions respectively, and F the number of frequencies in the sweep. This 3D data is processed into multiple 2D images of shape (M, N) containing the different values mentioned above.

3.6. Electrical connections

At the start of the thesis, the chip had been designed, fabricated and delivered. The part of the electrics beyond that was part of the experimental work of this thesis.

The chip is wirebonded on a PCB, which is connected to an FPGA. The FPGA handles the fast signals, adds them up, and passes a slower signal to an Arduino. The Arduino has built-in modules for digital communication, which it uses to pass all received data forth to the computer. The computer can then run Python code to process the data into an image. See figure 3.8 for a schematic overview.

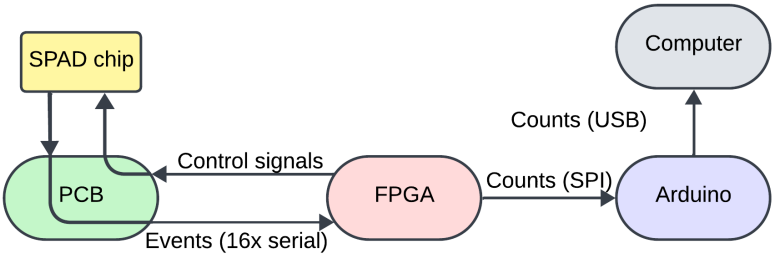


Figure 3.8: Overview of the communication chain between the chip and the computer. The chip is wirebonded on the PCB, which passes the connections straight to the FPGA. The FPGA controls the chip by providing a clock signal and multiple control signals that react to single pulses. The FPGA also decodes the serially communicated photon events and counts them. A slower, more summarized signal is passed to the Arduino through SPI protocol. The Arduino forwards everything it receives to the computer via USB without any processing.

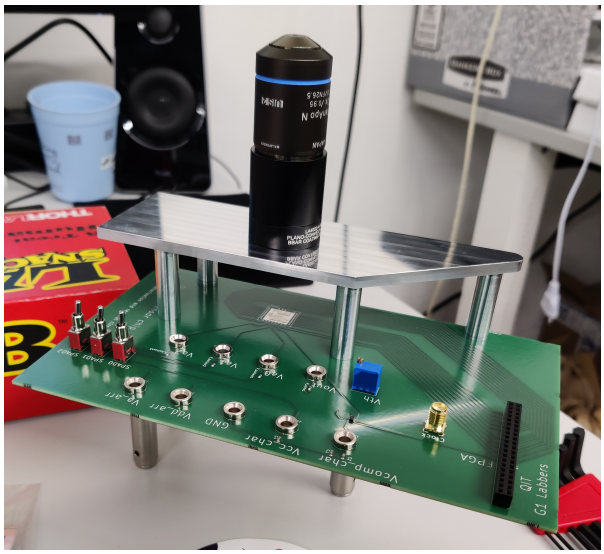


Figure 3.9: The PCB clamped in its custom metal table. Optics are mounted above, showing how the table is compatible with Thorlabs SM1 components. The heights are calculated so that the image from the objective is focused on the square where the SPAD chip will be wirebonded. Unfortunately, optical measurements with this PCB never happened due to problems within the SPAD chip.

Figure 3.9 shows how the PCB would fit in an optical system. Figure 3.10 shows the final setup with everything included. Note that optical parts on the table are visible that are not part of the setup. Focus on the part that is mounted on or besides the little aluminium table.

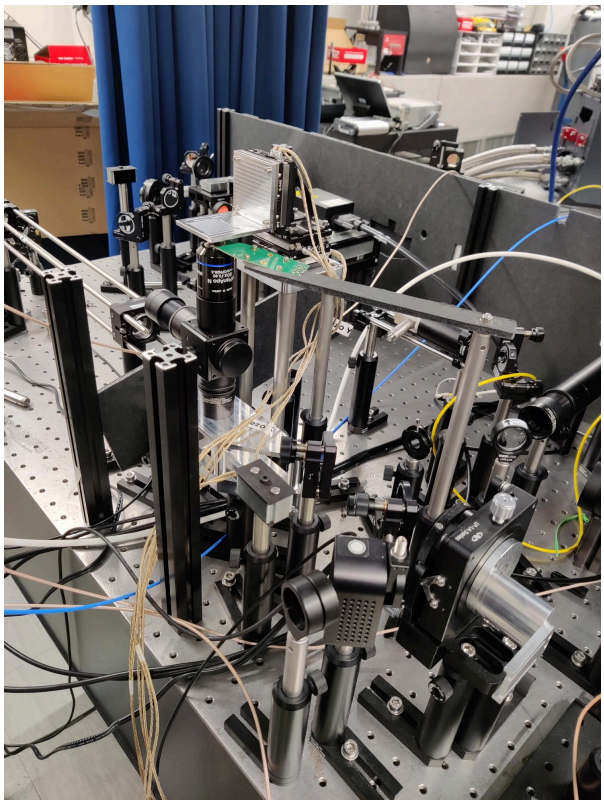


Figure 3.10: The final setup, for doing widefield tests. The little aluminium table is meant to hold the PCB in place, but also contains a slot for mounting the imager. An optical fiber can also be mounted on the bottom for tests with a fiber-connected SPAD. This setup is the realized version of schematic in figure 3.6.

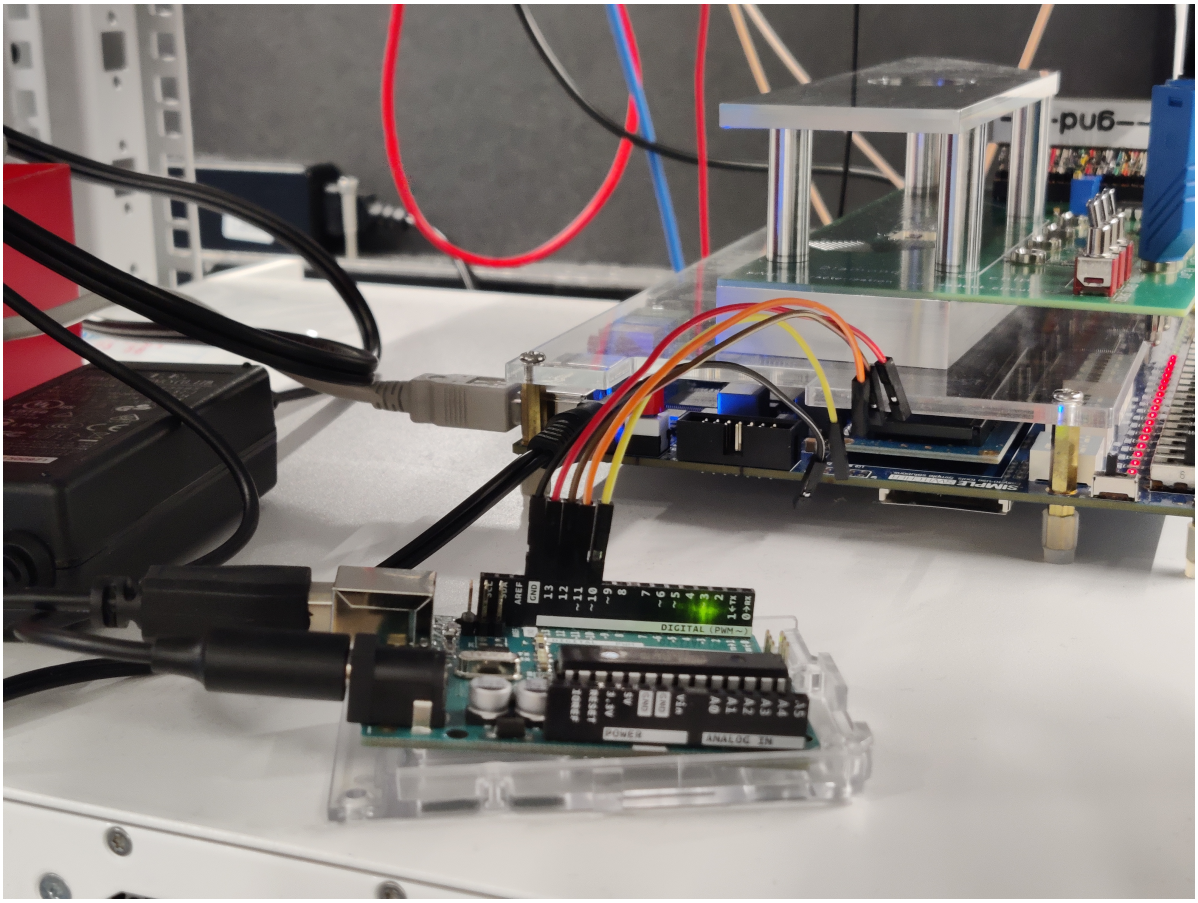


Figure 3.11: A photo of the FPGA and Arduino, with the PCB on top of the FPGA. The SPI connection between the FPGA and Arduino can clearly be seen, although it was temporarily disconnected when this photo was taken. In the background, the adapter is visible that connects the FPGA and the PCB, with the label ”–gnd–” on it (note that not all cable are ground. It was to indicate that the two ground cables present were on that side of the cable). The aluminium table for mounting the optical system onto the PCB is mounted on the PCB.

3.7. CMOS design

As mentioned in the introduction, previous bachelor students of electrical engineering have designed a CMOS chip with arrays of SPADs, see figure 1.2. This section is meant to give a short summary of what the chip does, so for the full details, please refer to the bachelor theses [22] [4].

3.7.1. Physical features

This chip contains 16x16 SPADs integrated with digital circuitry, along with 3 exposed SPADs that can be directly accessed for characterization purposes. The chip is made with 40nm CMOS technology, by TSMC, which allows it to have all necessary circuitry in just 1x1mm^2 space. The chip is made into four different variants, so that multiple SPAD designs can be tested. The SPADs are built as in the schematic in figure 2.4.

3.7.2. Digital communication

The chip is designed to communicate with an FPGA, rated for 1 GHz clock frequency by simulation. The FPGA has 22 connections with the chip, which consists of a clock signal, control signals, and 16 channels for serial communication of the photon counts. The FPGA constantly polls the chip, which is essentially asking to all SPADs ”Did you see a photon in the last X nanoseconds?”, and then the SPADs respond by sending 16x16 ”yes”es or ”no”s over the 16 serial channels.

3.7.3. Characterization

As previously mentioned, the chip is made into four different variants. Each chip contains three different SPAD designs on the top, to be accessed directly with electrical equipment for characterization. This makes for a total of 12 different SPAD designs to be tested. These top SPADs are not equipped with their own circuitry like the arrays, and instead need to be quenched by an outside circuit. This circuit is contained by the PCB on which the chip is wirebonded.

3.8. PCB design

As the bonding pads of the chip are too small to be hand-soldered, it had to be wirebonded. A custom PCB was designed for the chip to be wirebonded on. The wirebonding was done with 25-micron aluminium wires at the QuTech sample preparation room in the TU Delft. See figure 3.12. The PCB mainly has two roles: To provide the necessary circuitry for electrical and optical characterization, and to pass all the digital connections to the FPGA through a convenient connector. The design is shown in figure 3.13. Let us take a closer look at these roles.

3.8.1. Characterization circuit

The PCB contains an adaptation of the characterization circuit from the bachelor thesis [4], see figure 3.14. This can be used for optical characterization of the three analog SPADs on each of the four chips. Electrical characterization is done by plugging electrical equipment into the test points that were included in the design.

3.8.2. Digital readout

The part of the PCB carrying digital signals is trivial. Except for the clock signal, all necessary connections are directly tracked to a 2x20 GPIO connector, so that the FPGA can be connected on it. The clock has a separate SMA connector. One of the cables is used as ground connection rather than a signal cable, so that the FPGA and chip share a common ground. This is essential to ensure the voltages are right and the signals stay predictable.

3.9. Voltage mismatch

The SPAD chip’s logic voltage is 1.1V. However, the lowest logic level that the FPGA can use is 1.5V. Thus, FPGA output needs to be lowered, and chip output raised. The proper way to solve this is with a level shifter, but, due to time constraints, simple voltage dividers were used. These lowered 1.5V to 1.085V, ensuring the chip does not get overwhelmed by the FPGA’s voltage. This method distorts high-frequency signals a bit, and the FPGA clock had to be lowered to 1 MHz in order to get good stable signals through. This clock frequency is not enough to meet the originally set requirement of being able

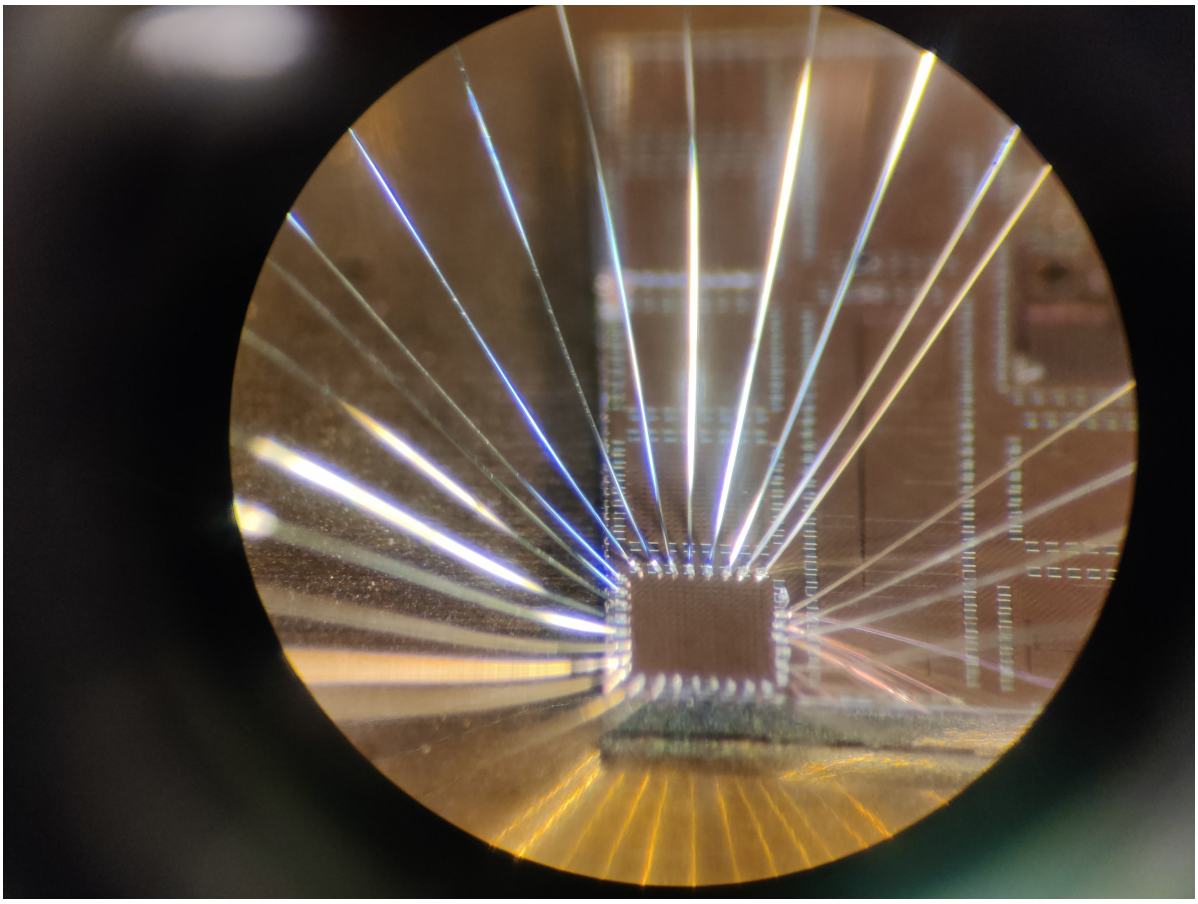


Figure 3.12: The CMOS chip containing the SPADs, wirebonded on the PCB. The first chip from the multi-project wafer was bonded. You can clearly see the aluminium bonds with a diameter of 25 micron.

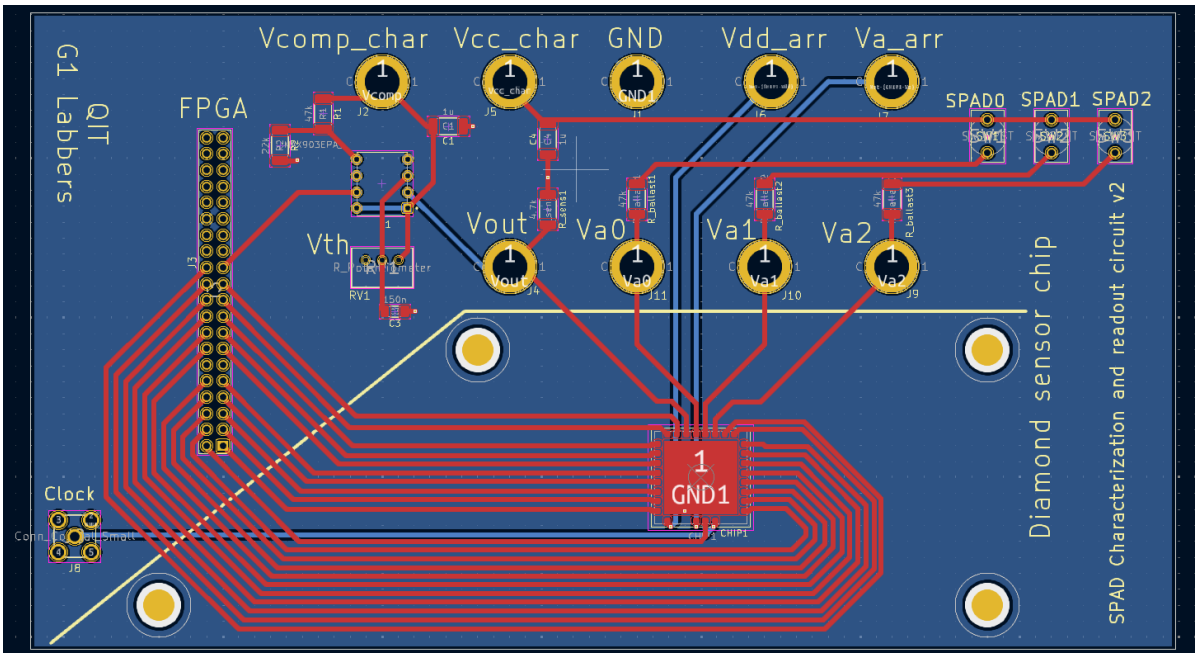


Figure 3.13: The PCB Design. The PCB consists of an SPAD characterization circuit and connectors for connecting power and the FPGA with the chip. Four holes are included for mounting the PCB on a custom metal part. This design was made in KiCAD and can be found on Github together with the code (more on that in the Appendix).

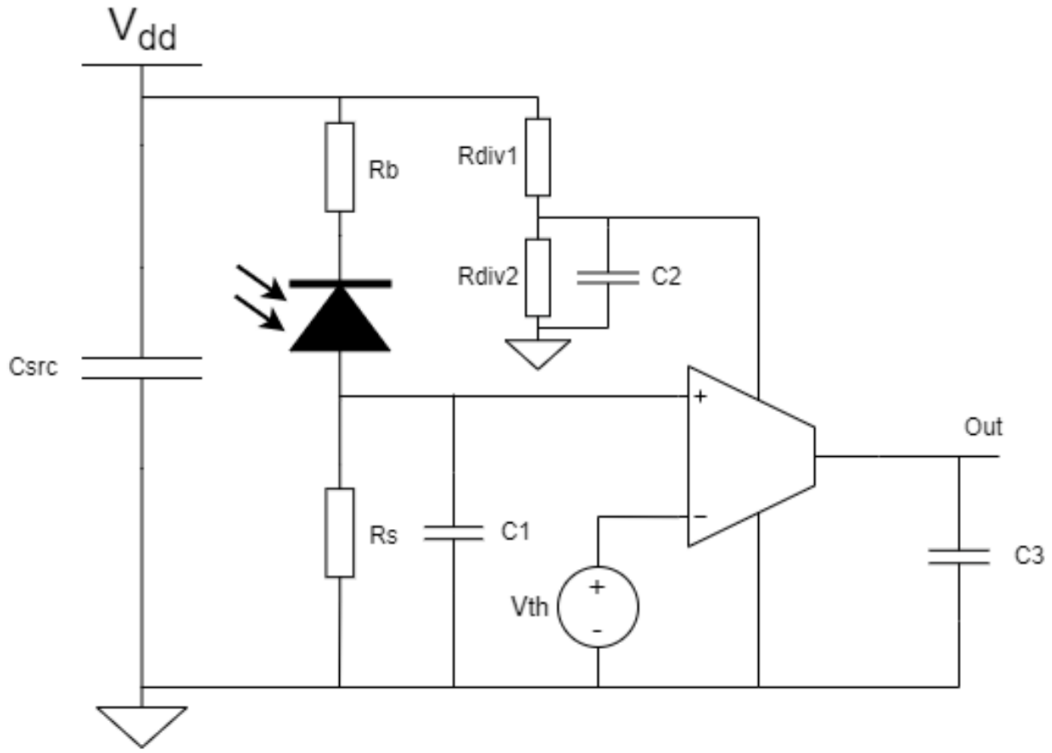


Figure 3.14: Schematic of the characterization circuit, taken from [4]. Passive quenching is achieved by having a ballast resistor in series with the SPAD. A sensing resistor is used to be able to calculate the current based on voltage measurements within a reasonable voltage range. A comparator acts as an ADC that converts the signals of photon counts into hard digital pulses, ready to be sent to the FPGA. Taken from [4]

to handle 3 Mcounts/s, which was imposed in the bachelor projects that were involved in designing the chip [22] [4]. Nonetheless, it can still serve as proof-of-concept, and a proper level-shifter can be made after determining that the principle works.

Regarding the chip output, the FPGA interprets everything above 975 mV as a logical high, and 1.1V is above that. Note that this significantly lowers the margin for noise and/or voltage drop, but for saving time it was worth a shot. Unfortunately, the moment to test this never came, due to a problem with the chip, see chapter 5.

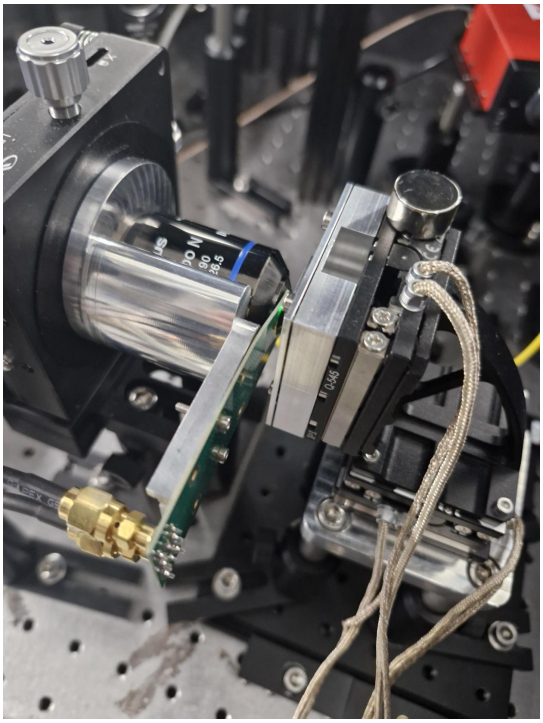
FPGA design and other software

This chapter has established how the communication chain of chip-PCB-FPGA-Arduino-PC is physically setup. The latter three also require software to run, which is addressed in chapter 4.

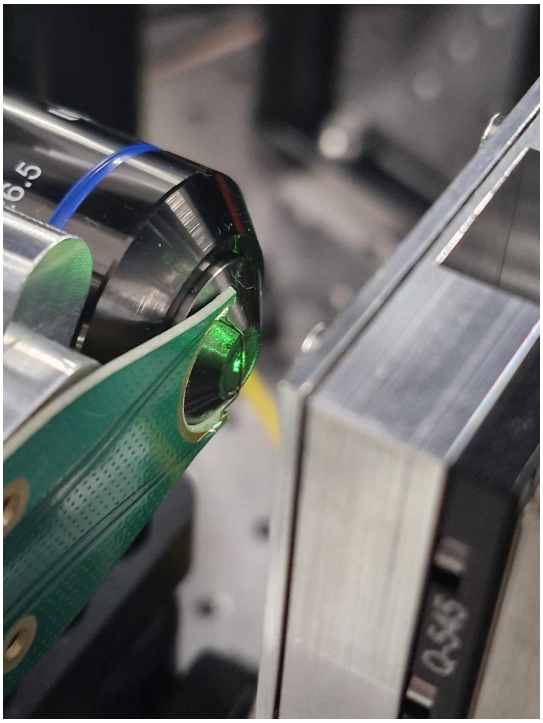
3.10. Microwave antenna designs

At the start of this thesis, a microwave (MW) antenna was included in the confocal setup, which is a small PCB which holds a wire that hangs inbetween the objective and the diamond. This ensures optimal microwave delivery as the wire is as close to the diamond as possible.

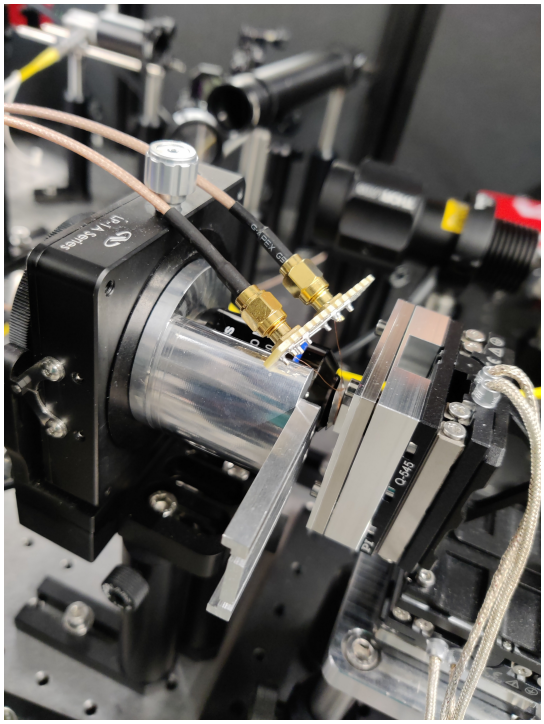
Despite the good microwave delivery, this MW antenna design can stand in the way when there is a need to focus on the backside of the diamond (such as for our bonding experiments, where strain needs to be measured at the back side of the diamond where the bonding has happened). In such a case, a MW antenna design that does not occupy space between the objective and the diamond is desirable. Hence, a new MW antenna was designed and tested. See figure 3.15c.



(a) The microwave antenna on a PCB designed by TNO. The PCB holds a small wire between the objective and the diamond, ensuring the wire is as close as possible to the diamond for optimal microwave delivery.



(b) Close-up of TNO's MW antenna design. The piezo stack has been moved away in order to show the wire.



(c) The new microwave antenna. It consists of a copper ring that goes around the tip of the objective, so that it does not occupy space between the diamond and the objective. This goes at the trade-off of having less microwave power at the diamond surface.

4

Software for ODMR

The previous chapter explained the physical setup of both the confocal and widefield setups. This chapter is dedicated to the software side. The full code is available on Github (see appendix chapter A). This chapter explains what the code does and how it was designed.

4.1. Software improvements for confocal ODMR

As explained in the introduction, an existing confocal ODMR setup had already been built by previous students before this thesis started. Although the setup has been functional before, it hadn't worked for months for unknown reasons. Besides this, the setup was not yet ready to perform 2D scans. A great deal of my master thesis has been improving the existing ODMR setup, besides making the new widefield testing setup.

4.1.1. Getting the setup working again

Before the start of this thesis, effort was made by previous students to migrate towards the automation of measurements, using code that was shared by TNO and improved on. TNO collaborated with us with the goal of building and improving automated single color-center characterization at room-temperature. The code runs without errors, but initially no ODMR contrast was visible. With previous code from Alfie's bachelor thesis [23], the setup turned out to work! Upon closer investigation, there seemed to be an incompatibility between the way TNO initializes the arbitrary waveform generator (AWG), and the way it was done in our old code. As a quick patch, a separate initialization script was setup to run anytime TNO's ODMR code runs.

4.1.2. Setting up 2D scanning

In TNO's code, the piezo stack is used by measuring counts while the stack moves at a constant rate. The measurement is shortly interrupted when the stack has to move to a new row, and then starts again. In order to do this, the three axes have to be synchronized among themselves and with the timetagger, in order to make smooth movements and measuring with the right timing. So in order to be able to use TNO's code, this synchronization had to be fixed, and any necessary adaptations of the code had to be made. I gathered information on how to do this, but ultimately decided it was outside the scope of this thesis. Fixing this would require quite some work, while an easier solution was already available: The method of performing separate steps with measurements in between. This was how the stack was controlled in Alfie's code [23], so I recycled his code to quickly make this work. Note that, in the future, we should opt for completely merging with TNO's code, as it is more efficient and organized, and facilitates closer collaboration. Regarding time-efficiency, note that TNO's scan should be faster because scanning and moving is done simultaneously. In the experiments done during this thesis, typically a dwelling time of 0.2 seconds was used, and the time needed for the piezo stack to take a step is also roughly 0.2 seconds. As a rough estimate, this means that TNO's version is expected to run about twice as fast without quality loss.

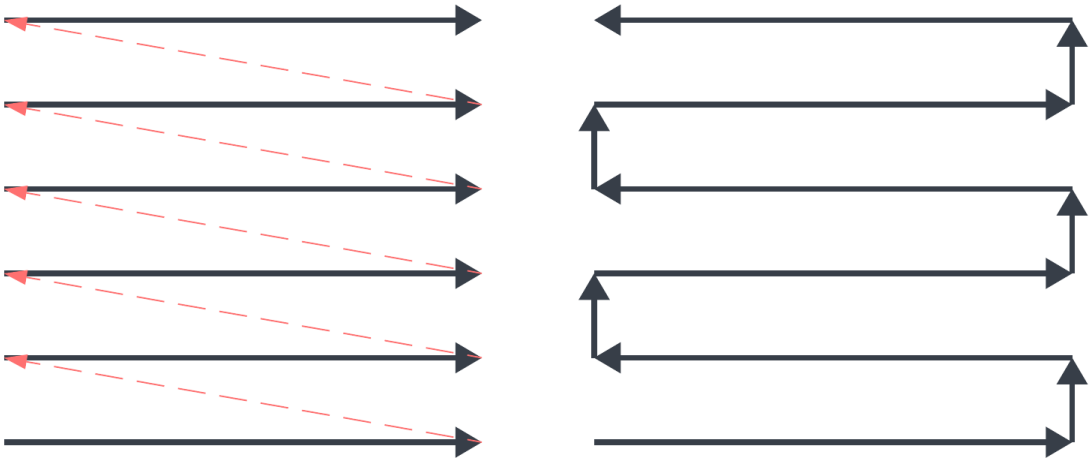


Figure 4.1: Scanning patterns. Left: Row by row left-to-right. In the diagonal movements no measurements are done. Right: Serpentine scan. No unnecessary movements are made this way.

4.1.3. Upgrades for 2D scanning

During the course of this master thesis, multiple upgrades have been done on the code and parameters driving the 2D scans, in order to accommodate different needs and problems that came up later. This section serves to explain the different modifications that were done on the 2D scanning program.

Serpentine scan

A simple way to reduce unnecessary movement of the piezo stack is by moving in a serpent pattern. Compared to the previously used left-to-right row-by-row pattern, this roughly halves the total movement. See figure 4.1. Besides reducing movement, the serpentine scanning pattern reduces discontinuity in the image. In other words, small errors accumulated during each steps become less prominent, due to avoiding the large movement when changing rows.

Adjustment of the notch filter

Sometimes, the piezo stack would start vibrating a lot, making it difficult for the system to move, often getting it stuck. By advice of Physiks Instrumente, the brand of the piezo stack, the notch filter was adjusted, which is a parameter that can be used to dampen certain frequencies in the movement of the stack. The purpose of this filter is to avoid resonance when the load is at the resonance mass. By eye it was hard to realize resonance was happening, but this turned out to be the case. It worked like a charm, and the piezo never vibrated like that again. Nonetheless, there was another problem also getting the stack stuck, see next subsection 4.1.3.

Periodic autozero

During long scans (of at least multiple hours) the piezo stack sometimes got stuck, freezing the experiment. Even after adjusting the notch filter (see previous subsection 4.1.3), this problem was present, only this time less frequent, and without vibrations. The stack knows it is not at target position, but fails to move. This typically happens when the piezo stack has performed in the order of tens of thousands of steps since its last autozero. Autozero means the stack performs a reference move in all three axes to recalibrate its zero position. In order to avoid getting the piezo stack stuck, a periodic autozero was added. It was set to activate every thousand steps, and serves to keep the piezo stack well-calibrated during the whole measurement. It was effective. The piezo stack never got stuck ever again for months after. One could argue this creates overhead in the measurement time. The autozero takes roughly ten seconds, and each step (assuming small steps) takes almost 0.2 seconds. Thus, autozeroing every thousand steps means that every 200 seconds, an additional 10 seconds of overhead is added, resulting in +5% scanning time. I find this a very acceptable amount of overhead.

Triangle selection

Albeit a bit of a niche feature, this was a quick fun easy-to-code feature to add. A diamond broke into a triangle piece, but there was still interest in mapping it. Hence, I added a feature that allows to define three corners (x and y coordinates). The serpentine scanning pattern still goes as usual, but any grid points that are not inside the triangle are skipped. For most triangular shapes this results in approximately half the scanning time as it would be when scanning the full bounding rectangle. Refer to chapter 5 to see what it looks like.

Tilt correction

In the real world, unfortunately things can never be mounted perfectly. This is also the case with the piezo stack. While only moving through x and y, the diamond appeared to become dim on one side of the image, which turned out to actually correspond to the surface getting slightly out of focus.

The cause of the tilt

Upon further investigation, unorthogonality among the xyz stages cannot be the cause of the problem, as long as the surface of the piezo stage is parallel with its moving directions. Tilt starts to be a problem when the diamond surface is not parallel with the moving directions of the piezo stage. This is illustrated in figure 4.2. In fact, the diamond is not mounted perfectly flat on the piezo stage. This is very likely due to the way we mount the diamonds. We glue them with PMMA on a mounting plate, and bake the PMMA for 2 minutes at 120 degrees Celcius. Typically the diamonds get a tilt in the order of a couple microns in the z direction for each mm in the x or y direction. The tilt usually is dominant in one of the four sides. The specific side seems to be random. Furthermore, the tilt remains consistent when remounting the same diamond (in the same orientation), while it differs among different diamond samples. This indicates that the cause of the tilt must be the way the diamond is mounted.

Digital correction for the tilt

In order to correct the tilt, a feature was added in the code to define a plane in xyz space. This is defined by a zeroing point (x_0, y_0, z_0) and tilt coefficients a_x and a_y. The tilt coefficients can be determined by looking at the diamond through the imager. To do that, the piezo stage needs to be moved such that the edge of the diamond is on the screen and in focus. This has to be repeated for all edges. The coordinates of the edges need to be noted down, after which the tilt coefficients can be calculated by $a_x = \frac{\Delta z}{\Delta x}$ and $a_y = \frac{\Delta z}{\Delta y}$. Note that this assumes the diamond surface to be a perfectly flat plane. In fact, the diamonds in the lab's inventory whose flatness is known is generally below 1 micron. Measuring Δz over a span of multiple millimeters gives values of typically multiple micron at least. Thus, the diamond are sufficiently flat to effectively act as a flat surface for the tilt calculation. Under large strain however, localized height differences in the diamond surface have been seen, see chapter 5.

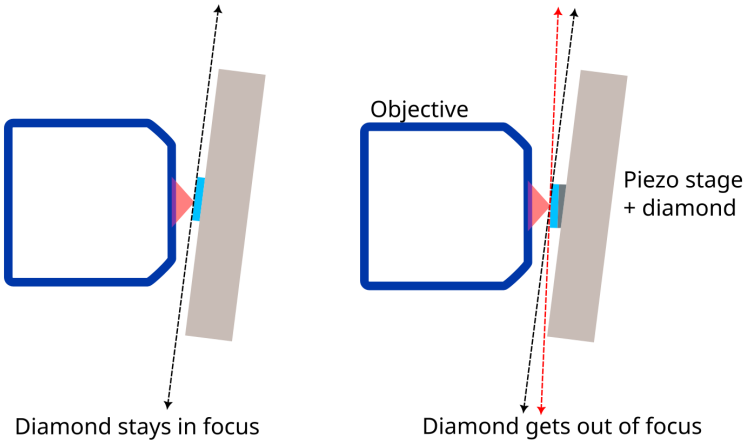


Figure 4.2: The effect of two types of tilts in the diamond and piezo stack. In a confocal microscope setup, only one focus point matters. This eliminates issues caused by one type of tilt.
Left scenario The diamond is perfectly mounted on a tilted piezo stage. The stage moves parallel with the diamond surface. The diamond surface thus follows the black dashed arrow. Notice that the focus point remains exactly on the diamond surface as it moves along the black line. Hence, this type of tilt does not affect the focus.
Right scenario The diamond is glued at an angle compared to the piezo stage, which may also have a tilt of its own. The black line is the direction in which the piezo stage moves. The red line indicates the angle of the diamond surface. Notice that the moving direction and the diamond surface are not parallel to each other. In this example, if the piezo stage moves downwards, the diamond surface will get closer to the objective, thus going out of focus. The amount by which the diamond is out of focus is equal to the difference between the red and black arrows. Hence, anything that causes the diamond surface to not be parallel with the moving direction causes problems with tilt during scans. The glue that is used for attaching the diamond to the piezo stage dominates the angle between the diamond and the piezo stage.

4.1.4. Z-scan and 3D photo-luminescence mapping

Following advice from TNO, a z-scan was added. This is a scan that moves the focus through the depths of the diamond instead of moving along the surface. This way, a graph can be obtained of photo-luminescence vs depth. This can give information on the exact location of the top and back surface. When the focus point gets exactly on the surface, more excitation light than usual is reflected into the SPCM. This increase is visible despite the dichroic mirror and lowpass filter reducing the reflected excitation light reaching the SPCM. This is the most precise way to determine the exact location of a surface.

3D mapping

With the z-scan operational, it became possible to perform 3D PL scanning. In other words, for each xy point in the grid, a whole z-scan gets done. This scan takes much longer than 2D mapping, but gives much more information. Instead of defining a plane that roughly corresponds with the surface, we get exact data on where the surface is and its PL for each pixel. This can be post-processed in both PL maps and height maps of the surface.

4.1.5. Post processing

When all work on the scanning code was finished, it was possible to perform a multitude of different kinds of measurements:

- 2D photo-luminescence scan
- 2D ODMR scan
- z-scan
- 3D photo-luminescence scan

Note that, in theory, 3D ODMR scan is also possible, but the time such a scan would take is beyond the patience of even the strongest willed people, and has thus been omitted in the code.

2D photo-luminescence map

This was the first kind of measurement added, and requires no post-processing. The data is already in a 2D array and can directly be plotted.

2D ODMR

A two-dimensional ODMR scan has lots of information packed in a 3D array. The three dimensions of the array correspond to x, y and microwave frequency. A multitude of 2D plots can be obtained from this information:

- **PL map**
  Take the mean of each xy pixel

- **Raw contrast**
  Take the difference between max PL and min PL for each xy pixel. Note that this overestimates the true contrast due to noise broadening the range of values. This is only meant to be used as a rough way to see observed contrast when fitting the Lorentzian functions fails.

- **Fit contrast**
  Fit the double-Lorentzian function and give each pixel the value of the fitted contrast. Provided that the fitting went well and makes sense, this is a more accurate estimation of the contrast than raw contrast.

- **Peak splitting**
  Fit the double Lorentzian function and give each pixel the value of the distance between the two peaks. Peak splitting scales with the perpendicular component of both magnetic field, electric field and strain.

- **Frequency shift**
  Fit the double Lorentzian function and give each pixel the value of the mean frequency of the two peaks. Frequency shift scales with the axial component of strain and electric field and is also affected by temperature.

By imposing further assumptions, magnetic field and strain can be calculated, as explained in chapter 2.

z-scan

The z-scan consists of a 1D array with photo-luminescence vs z. This can thus be directly plotted without further processing. The typically obtained shape in the graph clearly displays a region of high photo-luminescence, corresponding to the region where the diamond is.

3D photo-luminescence map

A three-dimensional photo-luminescence map can be processed into both PL maps and height maps. By combining the top and back height map, the observed diamond thickness can also be computed.

Extracting a 2D PL map

In order to extract a 2D PL map, a slice has to be taken out of the 3D data. In principle this can be done by defining a plane. The current version of the code automatically takes the top and back surfaces estimated by the height maps and gives their corresponding PL slices.

Height maps

Height maps can be defined as follows: The region where the diamond is located gives much more photon counts than the space around it. Define a threshold, and the region where counts get above the threshold is interpreted as the diamond. When doing this for each pixel, a map of the top and back surface can be computed.

Observed diamond thickness

By substracting the top and back height maps, the observed diamond thickness can be obtained. This should be equal to the true thickness of the diamond divided by the refraction index, which is approximately 2.4. For this computation, the choice of the threshold mentioned above is crucial. This is because, as the focus point is moved away from a surface, the counts smoothly go down instead of in a sudden discontinuous way. Thus, a low threshold can overestimate the thickness. On the other hand, a high threshold might make the program miss the start or end of the diamond, making it underestimate the thickness. Normally the thickness of the diamond is known beforehand, so conversely this can be used to finetune a good threshold for the height maps.

Improved fitting

Let us take a closer look at the fitting process of the double Lorentzian dip. At first, the triple dip fit from TNO's code was modified to fit a double dip. However, noise in the data sometimes confuses the fitter. During 2D scans, the quality of the ODMR data is in direct conflict with resolution and time. Hence, there was a need for a more robust filter that can handle lots of noise. The following subsections will summarize what the first fitter does, and how it was improved.

The TNO-based fitter

The TNO-based double dip fitter has three main steps that are performed to find the optimal fitting parameters, which consist of filtering, peak finding and least squares optimization.

- **Initial filter**
  A Butterworth filter of order 5 is used to filter out high-frequency noise. This helps with peak finding later.
- **Peak finding**
  At first, the Scipy peak finder attempts to find two peaks (Scipy is a commonly used scientific library for Python, see https://scipy.org/). If the peak finding fails, the halfway point is calculated based on cumulative sum, which is likely to be located roughly in or inbetween the dips.
- **Nonlinear least squares**
  After finding the initial estimates for the locations of the dips, nonlinear least squares is invoked to finely optimize the parameters.

The new fitting method

The goal of making a new fitting script is to improve robustness to handle higher amounts of noise. A secondary goal was to make the fitting process suitable for GPU acceleration, in order to be able to process large images. One of the ideas was to perform ODMR using the imaging camera of the confocal setup that contains millions of pixels, which would take too much time if just a simple double loop in Python was used. See below the steps listed in this method. These steps are also illustrated by figure 4.3. More results of the fitter in specific scenarios can be found in appendix chapter

- **Baseline estimates**
  The baseline intensity is estimated based on the first and last 5 pixels in the graph. This fitting method assumes that the graph fully contains the peak. With those same pixels, the intensity of the noise is estimated, so that it can be accounted for in further estimates.
- **Amplitude estimate**
  The amplitude (or contrast, if the data was normalized) is estimated by taking the difference between the maximum and minimum values present in the graph, and subtracting two times the standard deviation. High amounts of noise would normally lead to overestimation of the amplitude, but by subtracting two standard deviations, this is accounted for.
- **Basic peak finder**
  As the Scipy peak finder cannot be called from GPU, a simple peak finding method had to be written from scratch. This peak finding was done by defining a low and high threshold. A peak is defined as going under the lower threshold and all the way back above the high threshold. These thresholds are calculated by taking the minimum value of the graph, and adding predetermined

factors multiplied by the noise standard deviation. These factors are taken to be 3 for the lower threshold, and 5 for the higher one.

- **Width and frequency estimations**
  The step above will yield an integer amount of peaks, which can be anything from 0 to half the number of measurement points. Usually this will be 2, and then the peak locations can be taken as initial estimates for the Lorentzian peak locations. The number of measurement points that the peak occupies under the threshold serves as an estimate for the peak width.
  When the number of peaks found is 1, then the distance between the peak locations is guessed to be 3 MHz, which is roughly the size of nuclear splitting.
  When the number of peaks is 0, then the peak locations cannot be estimated, and a default value is passed that assumes two peaks at 25% and 75% of the graph, hoping that the optimization step will take care of it.
  When the number of peaks found is 3 or higher, The center frequency is estimated as the average position of all dips. The width can be estimated by the average width of all dips. The splitting between the peaks is estimated as the standard deviation in the location of each dip. Although 3+ dips give no direct information about a supposedly double-peak structure, these estimates are made with the hope that the optimization step can take care of it.

- **Finetuning parameters**
  An Adam optimizer is called to fine-tune the parameters, using least squares as the loss function. Adam is a type of gradient descent method, and more information can be found in Kingma & Ba et al [16]. The parameters are passed in logarithmic scale to ensure that all values remain positive.

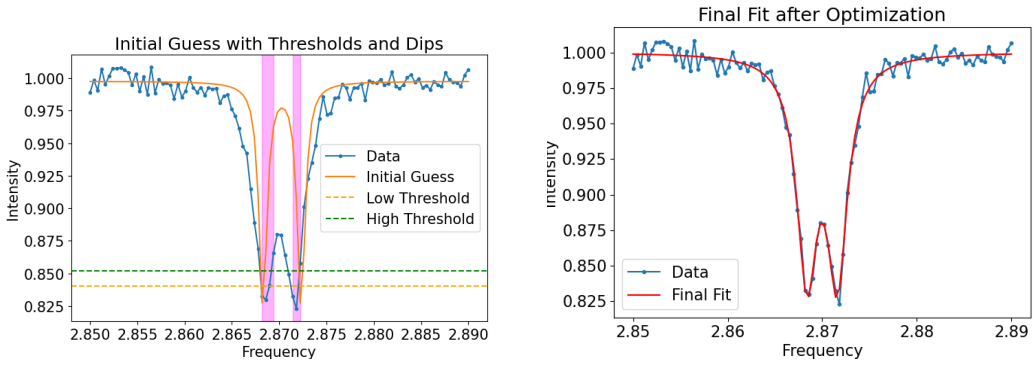


Figure 4.3: Demonstration of the new fitter with simulated data.
Left The initial guess. A low and high threshold for dip detection are dynamically determined based on the standard deviation of the left and right tails of the signal. The thresholds are set at the lowest data point plus predetermined factors times the standard deviation. A dip is defined as a signal that passes both thresholds at the sides of the dip. This is indicated by the magenta rectangles. An initial guess for all parameters of the double Lorentzian dip are determined using the found regions. For more detailed information, refer to the main text above.
Right The final fit after going through the Adam optimizer. The fit was a success.

4.2. Software for the widefield testing setup

As mentioned in chapter 3, the FPGA, Arduino and computer need software to process the signals received from the SPAD chip. This section explains the software of these components.

4.2.1. FPGA design

The main job of the FPGA is to act as a fast parallel counter. It decodes the incoming 16 channels of serial data and sends the total number of counts for each SPAD through a slower signal that microcontrollers and computers can handle. See figures 4.4 to 4.7 for a quick visual overview of the VHDL modules designed for the FPGA. In the subsections below some more explanation is given on the general workflow of the FPGA.

The readout cycle

The FPGA keeps polling the chip. This is done by sending a pulse to the chip, indicating that it is time to send the values of all SPADs. The chip then proceeds to send all 256 values over 16 channels, which is decoded by the FPGA and added to its internal counters. The values on the chip are cleared, and time is given for new photons to get counted before the next readout. The internal counters of the FPGA are 32 bit, thus having room for more than 4 billion counts in total. Hence the readout cycle can be repeated up to 4 billion times. Above that, the counters are in danger of integer overflow.

Passing on the counters

When all readout cycles have been done, all counts are sent to the Arduino. 256 counters of 32 bits (4 bytes) make a total of one KiB to be transferred. This is done by using SPI, for which a VHDL implementation for FPGA's is openly available [5]. The FPGA is configured as the SPI master, while the Arduino acts as the slave. While the FPGA runs at 50 MHz and can do many things simultaneously, the Arduino only runs at 16 MHz and does things sequentially, typically needing multiple clock cycles for one operation. Hence, the SPI clock must run much slower than the FPGA's internal clock. This was found to be roughly two orders of magnitude lower, and also giving space inbetween the bytes for the Arduino to process it, making it effectively three orders of magnitude lower. This speed was found to still be fast enough for measurements longer than 1 second. To save time, the counts are stored in a 1 KiB buffer to be sent, while the next set of readout cycles can already continue. Each buffer starts with a 0xFFFFFFFF value to indicate the start of the frame, followed by the 32-bit count numbers joined back-to-back.

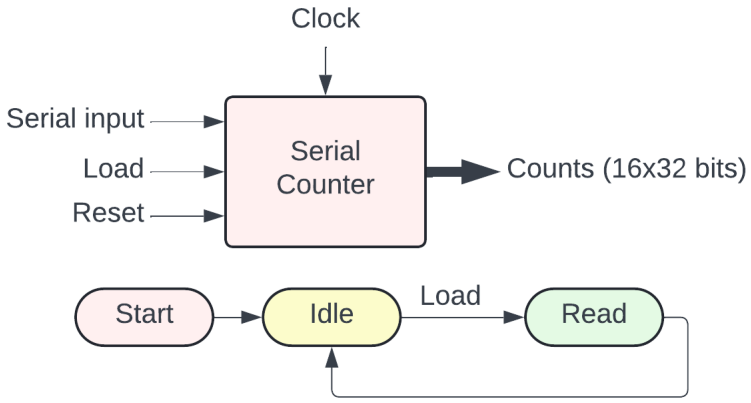


Figure 4.4: Simplified schematic of the serial counter module in the VHDL code for the FPGA. It idles until a load signal is sent to the SPAD chip. Once that happens, it verifies that the flag bit arrives at the right time, and goes into the read state. In the read state, on every rising edge, a bit is received from the chip, and needs to be added to the corresponding counter. When all 16 data bits have arrived, it goes back to the idle state.

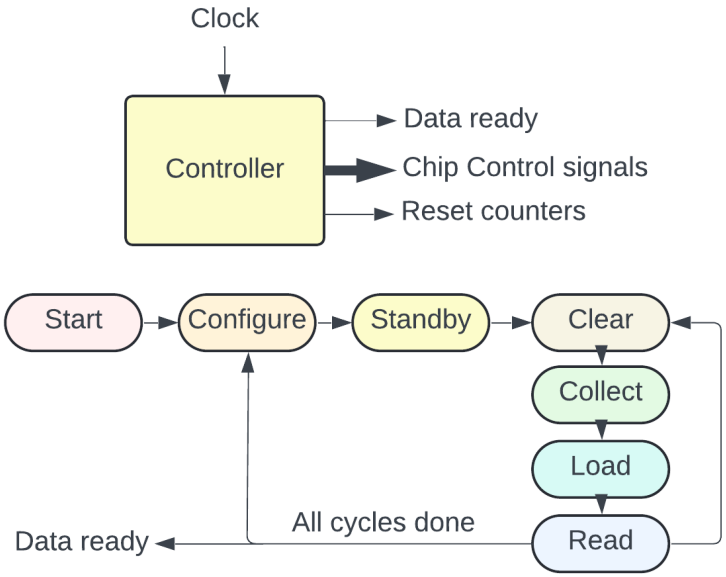


Figure 4.5: Simplified schematic of the controller module in the VHDL code for the FPGA. It starts by configuring the chip, meaning it sends sequences of pulses to the SI_w and SI_d pins of the SPAD chip to configure the timing of the quenching. When configuration is done, it goes into standby state. Originally the idea was to have it start on command. Due to time constraints however, the controller is made to keep on measuring. The standby state therefore automatically advances to the readout states. The readout states are a specific sequence: First send a ”clear” pulse to clear the shift registers on the SPAD chip. Then, wait for a defined time for the SPADs to collect photon events. Then, send a ”load” pulse to fill the shift registers with the new photon events. Then wait for the serial counter module to handle the readout before ending the readout cycle. After the readout cycle has been repeated a set amount of times, a ”data ready” pulse will be sent to the buffer sender to start transmitting data.

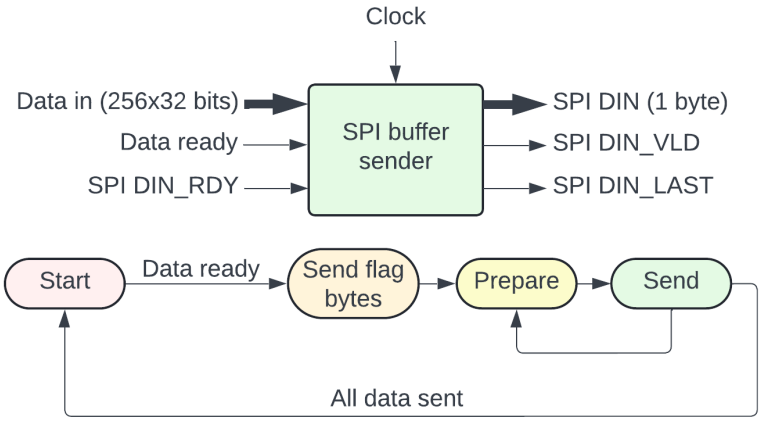


Figure 4.6: Simplified schematic of the buffer sender module in the VHDL code for the FPGA. It accepts all the bits of all the counters as input, and gives the data byte-by-byte to the SPI master module, which handles the low-level transmission of each byte to the Arduino. This is achieved as follows: When it receives the ”data ready” signal from the controller module, it starts sending the flag bytes. This consists of four bytes with value 0xFF. Then, it advances to the prepare state, where it loads the first byte of the counters. After proceeding to the send state, it gives the signal to the SPI master module to start sending the byte. While the SPI is handling that transmission, the buffer sender returns to the prepare state to load the next byte. It then waits until the SPI master module signals it’s ready for the next byte, and then goes to the send state to send the next byte. It goes on and on until all data has been sent, and then returns to the start state.

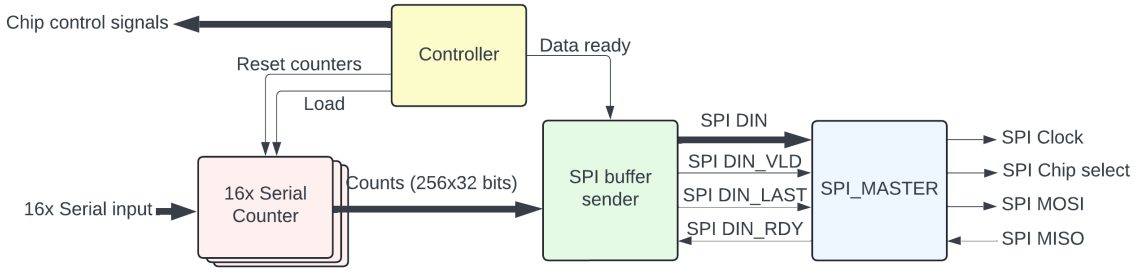


Figure 4.7: Simplified schematic of the top level entity in the VHDL code for the FPGA. The controller keeps repeating readout cycles. The serial counter keeps track of decoding the serial data and updating the counts. The SPI buffer sender module handles byte-by-byte transmission of the data when a set amount of readout cycles have been done. The next measurement can already start while the current one is still being transmitted.

4.2.2. Arduino

The purpose of the Arduino is to facilitate communication between the FPGA and the computer. It communicates with the FPGA through SPI. All bytes that it receives are immediately forwarded to the computer through USB connection, without any additional processing. To this end, the code of the previous bachelor students [4] could be reused unmodified.

4.2.3. The lab computer

The computer is the end of the communication chain, receiving bytes from the Arduino continuously. The computer waits for the 0xFFFFFFFF signal, then starts interpreting every quartet of bytes as a 32-bit unsigned integer. Each such integer is put in its place within an 2D array, so that it ends up with a complete image of all the counts. This image can be shown live on screen, or used within the ODMR code for performing simultaneous ODMR.

5

Results

Many results have been gathered during this thesis. This chapter will first present verification that the ODMR software improvements work. Then it will show the results on magnetic shielding. Afterwards the tests with the SPAD and the FPGA are presented. Finally, the bio-imaging results are shown.

5.1. ODMR software improvements

Multiple improvements to the ODMR software have been made, see chapter 4. In this section, some results will be shown to demonstrate the improvements in action.

5.1.1. Tilt correction

One of the software improvements was tilt correction. A diagonal plane is defined in xyz space, and the piezo stage moves along that plane, as opposed to the xy plane. Figure 5.1 shows a measurement without tilt correction. The top left corner is slightly out of focus. All other diamond images in this chapter are tilt-corrected, and serve to demonstrate its effectiveness. This tilt correction can handle an arbitrary amount of tilt, only limited by the diamond hitting the objective. This is roughly 100 micron in the z-direction, and translates to maximum tilts roughly between 100 micron/mm for 1x1mm^2 diamonds to 10 micron/mm for 10x10mm^2 diamonds. Different objective working distance and/or microwave antenna positioning can impact this limit.

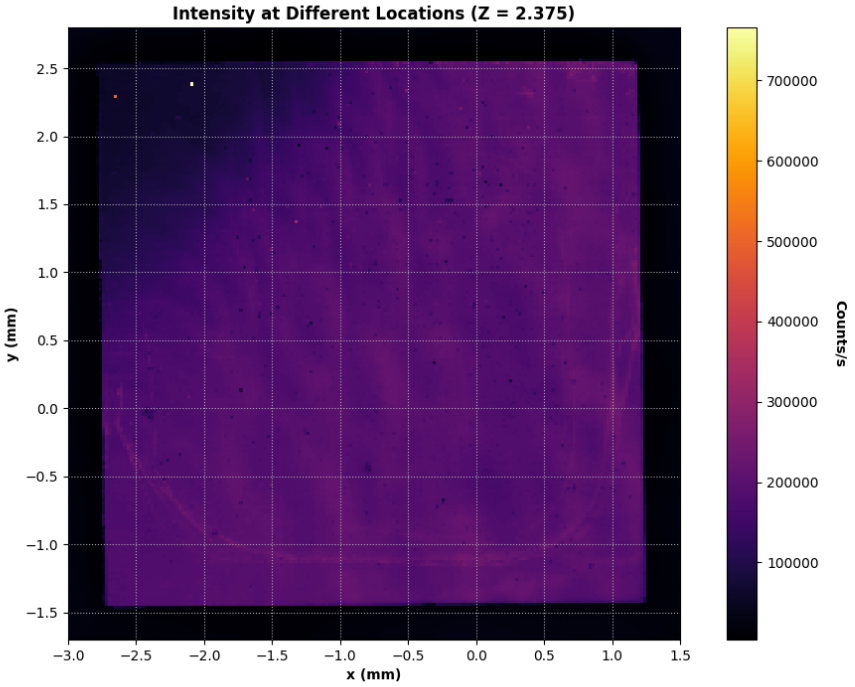
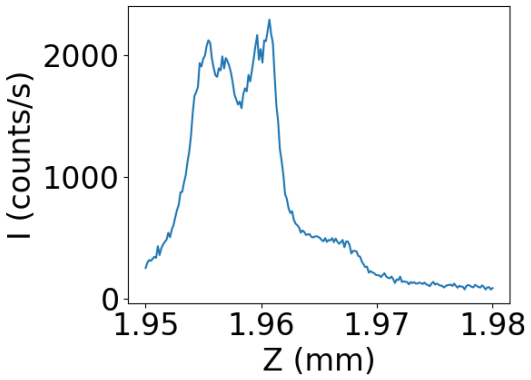


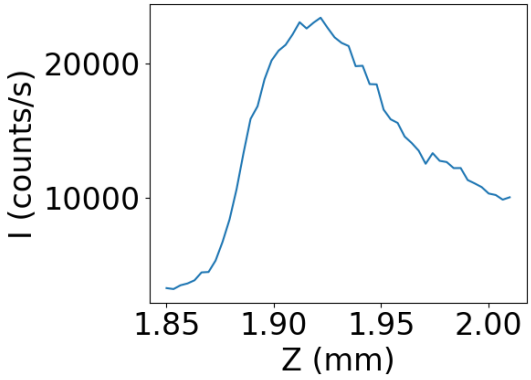
Figure 5.1: Photo-luminescence scan of a diamond using the confocal setup. This was made before tilt correction was implemented. The top left corner of the diamond is out of focus. This measurement acted as the inspiration to implement a tilt correction. After this, no more measurements had problems with parts of the diamond getting out of focus.

5.1.2. z scan

To demonstrate the z-scan in action, two graphs have been selected. Figure 5.2a shows one of the better z-scans, showing the expected shape. Figure 5.2b shows a z-scan with a shape deviating from expectations, but which can be explained by its surface being etched and thus not flat enough to show reflections. Most z-scans show features of size 0.5 - 1 micron and larger. This loosely matches with the diffraction limit which effectively puts a minimum size on the laser spot.



(a) Z-scan of point A on the triangular sample. The region where the diamond is can clearly be seen by the increased counts. Both the front and back surface show a spike, which likely corresponds to some green excitation light perfectly reflecting back and leaking through the filters.

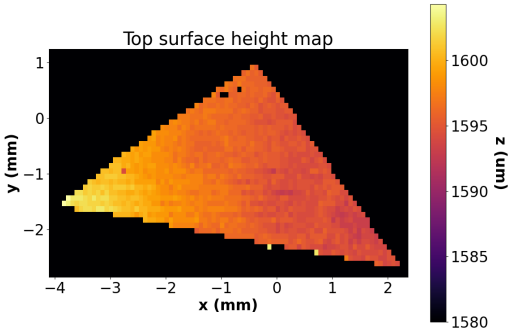


(b) Z-scan of the etched sample. As opposed to the z-scan in figure 5.2a, this z-scan does not show clear spikes at the surfaces. This is likely due to the etched surface not forming a perfect flat surface to show reflections.

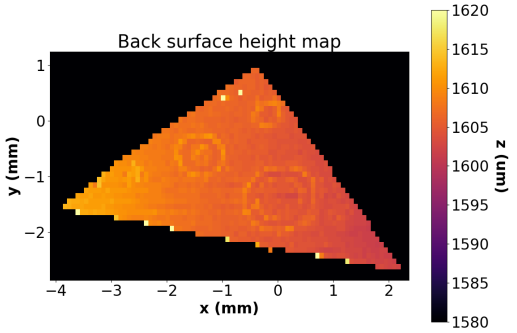
Figure 5.2: Two z-scan results from different diamonds.

5.1.3. 3D PL

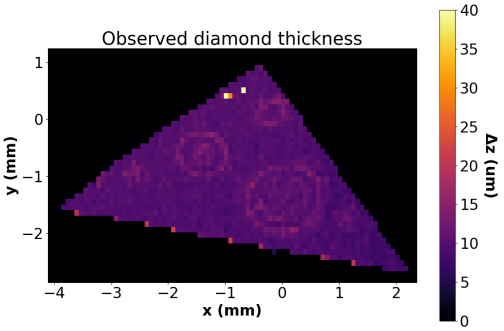
To demonstrate 3D PL measurement, once again the triangular sample is used, as it has a well defined structure. The triangular sample is the result from a failed bonding process from diamond on SiO_2 [temporary note: need to confirm this is the material], likely due to some small dirt particles trapped underneath the diamond, which would cause the diamond to bend in rings around those particles. Figures 5.3a - 5.3c show the result of the 3DPL scan. The ring-shaped bends can be seen in the measurement, indicating that the 3DPL measurement works correctly for measuring height differences on both the front and back diamond surfaces. Note however that, especially for calculation of the diamond thickness, a good choice of threshold is crucial. It matters not so much for relative differences, but more so for the thickness numbers in absolute terms. For the triangular sample, a threshold of 1000 counts/s was used, which was decided after manually inspecting the z-scans. Also note that the diamond looks thinner from the perspective of the objective due to the high refraction index. Due to this optical effect, the distance between the front and back of the diamond are effectively shrinked by a factor 2.42, which is the refraction index of diamond. In order to calculate the true physical thickness, the observed thickness needs to be multiplied by this refractive index. By processing data from the 3D PL measurement, height differences in the order of microns can be distinguished.



(a) Height map of the top surface of the triangular sample. A slight tilt can be seen. The surface is approximately flat.



(b) Height map of the back surface of the triangular sample. Besides a tilt, also rings can be seen on the triangle, indicating that the back surface is bent in ring-like shapes. This supports the hypothesis that the triangular sample became like this due to small particles trapped under the diamond during the bonding process, thus bending the diamond.



(c) Observed diamond thickness of the triangular sample. This is calculated by subtracting the height maps from figures 5.3a and 5.3b. Since the diamond is slightly bent along the rings, the confocal microscope is thus looking at an angle through the surface. That can explain why the diamond looks like it has non-uniform thickness, even though it is supposed to be a uniform slab.

Figure 5.3: Results for 3D PL scans for the triangular sample.

5.1.4. The fitter

During the bio measurements, the fitter reached a success rate of about 95.4%. In the failed fits, it most often assigns both dips to the same true dip. In the results of the bio measurements that are shown later in this chapter in figure 5.18, the failed fits are easily distinguished because of their outlying values. The fitter was fine-tuned on simulated data, but real data has one aspect that differs from simulated data: The two peaks may have slightly different amplitudes. The cause of this difference is unclear. The new fitter performs particularly bad when such asymmetry in the peaks appear, even mild ones. Of the 4.6% of pixels that had a failed fit, likely all of them failed in this specific way. An illustration of one real measured pixel where the fitter got confused can be found in appendix chapter E.

5.1.5. 2D ODMR

To demonstrate 2D ODMR, the measurement of the blank biosample is a good fit. This was from before the magnetic shielding was done (which is addressed in the next section), and shows peak splitting due to the piezo stack's magnetic field. The peak splitting varies roughly between 12 - 14 MHz, which corresponds to a magnetic field strength of about 0.43 - 0.5 mT, assuming negligible strain and electric field. For a sense of scale, compare that with the Earth magnetic field strength in the Netherlands, which is about 48 μT.

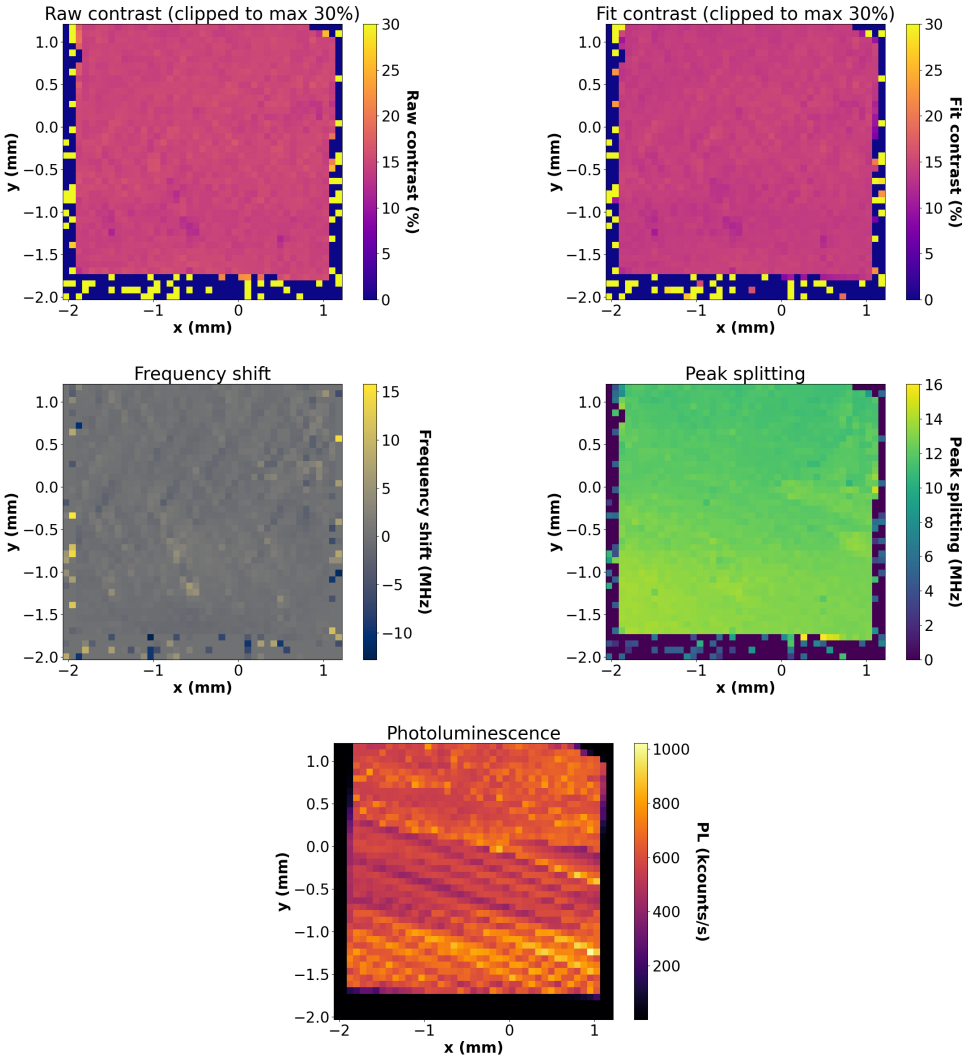


Figure 5.4: 2D ODMR results of the blank biosample. Raw contrast is the relative difference between the maximum and minimum intensity measured for each sweep. Fit contrast is the contrast of the ODMR dips according to the fitted function, which should be slightly lower than raw contrast when the fits are successful. Frequency shift and peak splitting denote the shift and split of the two resonance dips respectively. The peak splitting translates to about 0.43 - 0.5 mT magnetic field strength. Finally, the bottom image shows the average photoluminescence across each sweep.

5.2. Effectiveness of magnetic shielding

The piezo stage emits a static magnetic field that causes peak splitting roughly between 10 - 15 MHz when performing ODMR. In order to mitigate this, two simple magnetic shields were made. One is made of steel, the other from MuMetal. The steel one is 3 mm thick, the MuMetal one 3.1 mm. Both are of rectangular shape that matches the shape of the piezo stage.

In order to test the effectiveness of the magnetic shields, ODMR has been done at the same location on the same diamond without shield, with steel shield, and with MuMetal shield, for which Maarten's sample was used. Figure 5.5 show the results. Both the steel and MuMetal shields significantly block the magnetic field of the piezo stack. The shields block enough of the magnetic field that the resulting splitting at most of the same size as hyperfine splitting. Hence, the magnetic field is blocked sufficiently.

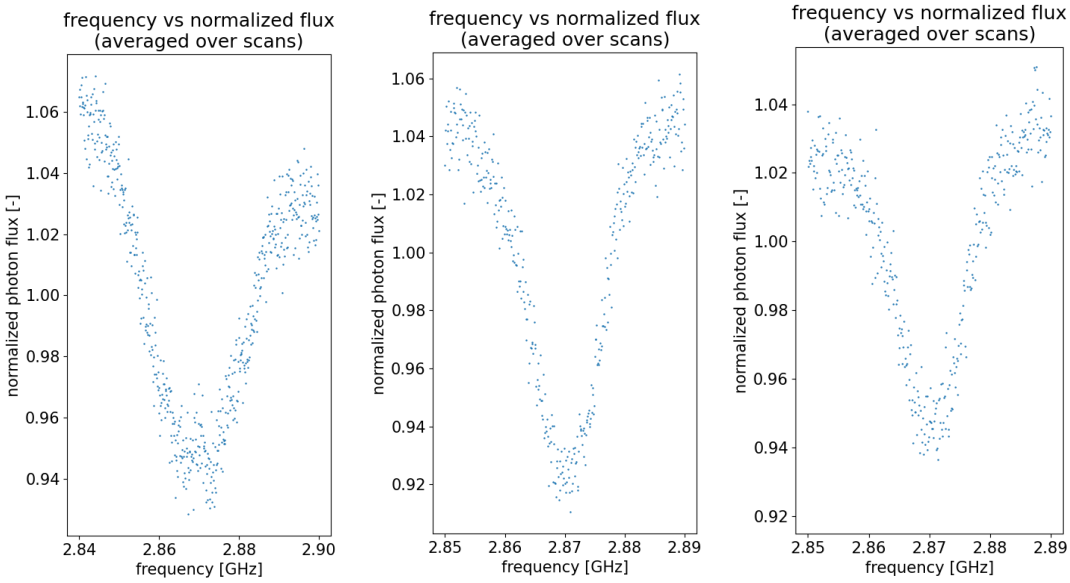


Figure 5.5: ODMR results with magnetic shielding. Left: no shield. Middle: Steel shield. Right: MuMetal shield. A clear difference is visible between the measurement without a shield and the measurements with a shield. The left graph has a splitting of about 11 MHz, which corresponds to roughly 0.4 mT. For the shielded measurements, no clear distinction can be found between two peaks. In fact, the peaks are close enough together that the splitting due to magnetic field is no bigger than hyperfine splitting, which is between 2 - 3 MHz. Thus, the magnetic field is reduced to below the sensitivity of our system, and therefore the shields block the magnetic field of the piezo stack sufficiently.

5.3. Performance of the new microwave antenna

In chapter 3 two MW antennas were presented: One on a PCB from TNO, and one on a ring-like structure. The new antenna is designed to not occupy space between the diamond and the objective, but at the trade-off of delivering microwaves less efficiently. This was tested in practice by doing the same measurement with 5 sweeps with both antennas. The old antenna showed 8% contrast, while the new antenna only showed about 1 - 2% contrast. This confirms the trade-off between space and MW delivery.

5.4. SPAD characteristics

By manually changing the reverse bias voltage of the power supply, the breakdown voltage of the SPADs was determined. This was determined to be roughly between 15 and 16 V, with exception of one SPAD that had breakdown at 17 V. This agrees roughly with S. Pellegrini et al [19], who also measured with SPADs on 40 nm CMOS technology. Later, by using a probe station, full IV and CV curves were measured. figures 5.6 - 5.9 show the IV curves of all the SPADs, grouped by chip. The CV curves are too many to show here. Three representative curves have been selected and are shown in figure 5.10. They generally agree with typical diode curves, although SPADs only show a frequency dependency and not a voltage dependency. Compare for instance with Bourim & Han [3]. All other measured CV curves resemble one of these representative three. The full list of CV curves can be found in the appendix.

In order to try to see pulses, the SPADs were connected to an oscilloscope using the appropriate connectors on the PCB. The SPADs were powered above breakdown voltage (varying voltages up to 28 V). Against expectations, no pulses were visible. Instead, the SPADs remained in a continuous avalanche, or in other words, never got quenched. As the quenching is done by a passive circuit, various resistors in the order of $10^5 - 10^6 \Omega$ were connected in series to try slower more robust quenching, but the result did not change. An unrelated SPAD with its own passive quenching circuit was connected to the oscilloscope to make sure it is able to detect the individual pulses, and the pulses were in fact visible. The passive quenching circuit itself only consists of a resistor. As different resistors were tried without success, the problem is unlikely to be an issue in the passive quenching circuit. Hence, there is likely a problem with the SPAD design in our SPAD chip, preventing them from getting quenched.

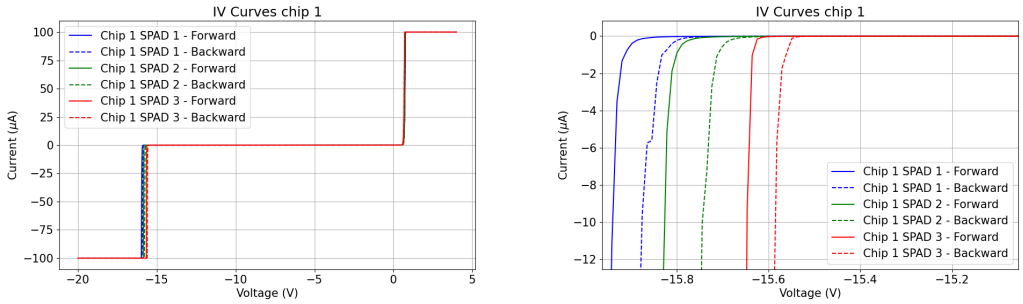


Figure 5.6: IV curves of the SPADs of chip 1. The right graph is a zoomed-in version of the left graph. A small hysteresis can be seen between the forward and backward sweeps. The breakdown voltages are roughly between -15.5 and -15.9 V.

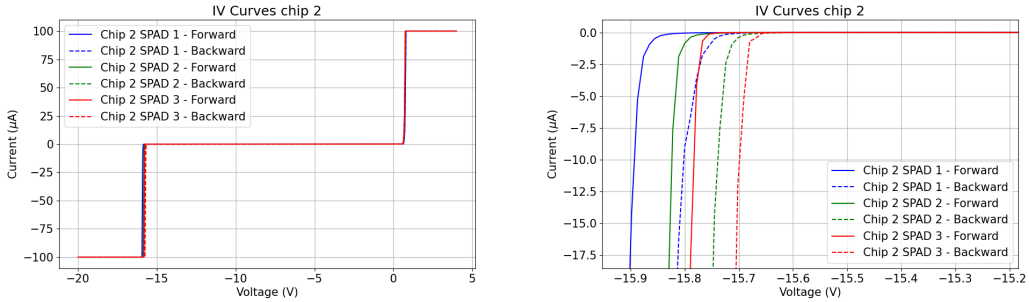


Figure 5.7: IV curves of the SPADs of chip 2. The right graph is a zoomed-in version of the left graph. A small hysteresis can be seen between the forward and backward sweeps. The breakdown voltages are roughly between -15.7 and -15.9 V.

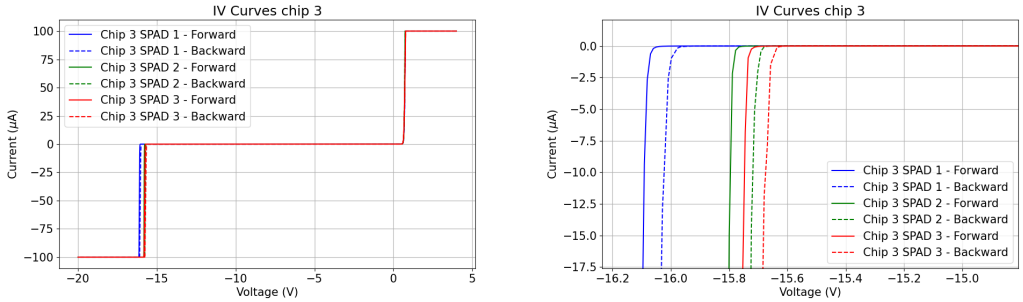


Figure 5.8: IV curves of the SPADs of chip 3. The right graph is a zoomed-in version of the left graph. A small hysteresis can be seen between the forward and backward sweeps. The breakdown voltages are roughly between -15.7 and -16.1 V.

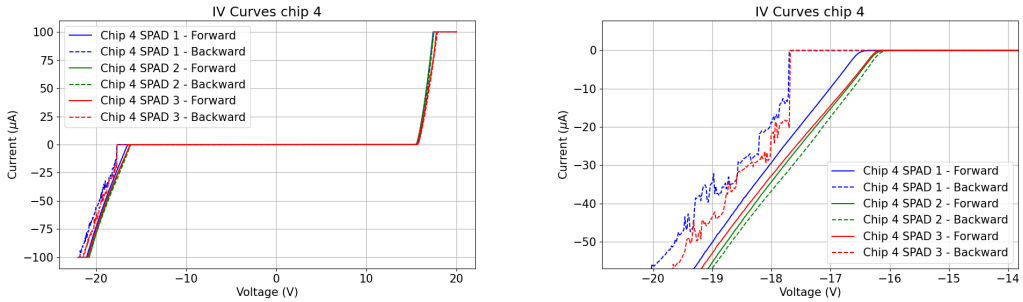


Figure 5.9: IV curves of the SPADs of chip 4. The right graph is a zoomed-in version of the left graph. Besides having the same amount of hysteresis as the other chips, the IV curves are also much less steep compared to the other three chips. Furthermore, the backwards sweep turned out more chaotic than the other chips. The breakdown voltages are generally between -16.1 and -16.6 V, although two of the backwards sweeps only show breakdown at roughly -17.7 V.

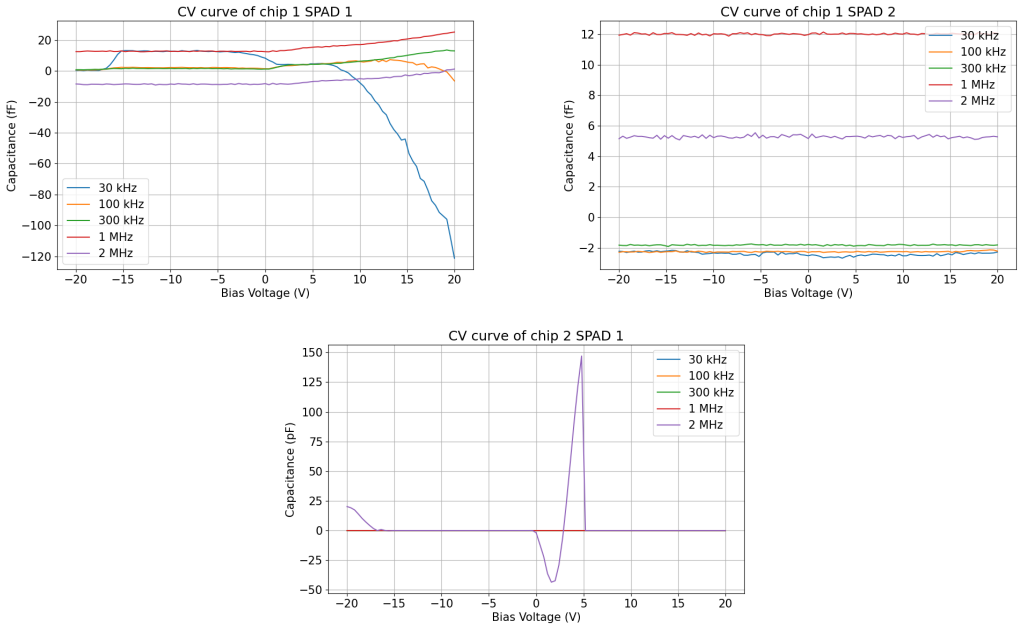


Figure 5.10: Three representative CV curves of the SPADs. A current limit of 100 μA is set, which distorts the sides of the graphs.
Chip 1 SPAD 1 shows a positive trend when approaching the forward bias from below for most frequencies. Beyond forward voltage, the graph makes no sense because the current limit of 100 μA is reached. Notably, at a frequency of 30 kHz, a plateau forms in the CV curve that mostly lies within the bias region where the SPAD acts as a blocking diode.
Chip 1 SPAD 2 Shows a capacitance that remains constant across all voltages, but does show frequency dependency, the highest capacitance being at 1 MHz.
Chip 2 SPAD 1 shows wild behavior both under the breakdown voltage and above the forward bias voltage. The spike around forward bias is common behavior for diodes, see e.g. Bourim & Han [3]. The dip that has negative capacitance could be an artifact of an imperfect measurement setup and/or calibration.

5.5. SPAD chip signal tests

Before mounting the SPAD chip in the widefield testing setup, a check had to be done to see whether we get any signal at all from the chip. Unfortunately, the answer is no. Multiple tests have been done to come to this conclusion, which are explained in subsections below.

5.5.1. Artificial test signal

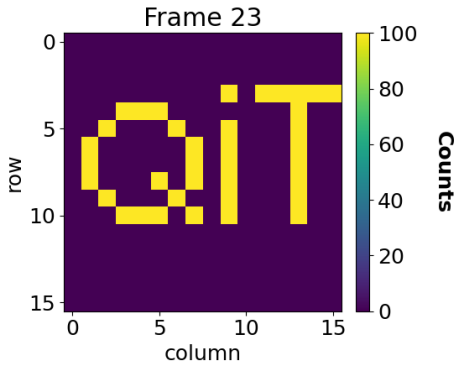
An FPGA design was loaded that creates an artificial signal rather than reading the signal from outside input. This signal was defined such that it mimics core behaviour of the SPAD chip, and gives signal as if light hits the SPADs such that they spell ”QIT”. This was successful, indicating that the FPGA-Arduino-computer communication works as expected. See figure 5.11a.

5.5.2. Test with powered chip

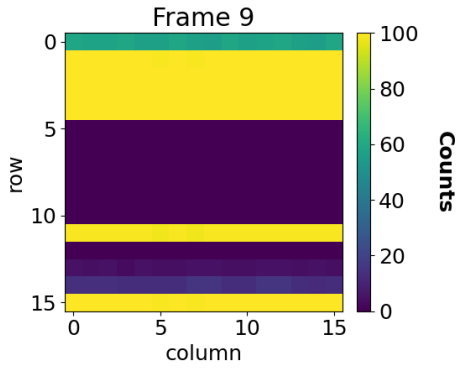
After success of the FPGA-Arduino-computer communication was established, the chip was connected to the FPGA and powered, but its SPADs, which require a separate higher voltage, not powered yet. This yielded undefined signal. The rows randomly switch between fully saturated counts and zero counts, but with a bias for the top and bottom to be saturated, while the middle is most often zero. Figure 5.11b shows a snapshot of this signal, but keep in mind it changes every frame. Powering the SPAD did not change the signal visibly.

5.5.3. Electrical tests on the chip

By the two previously mentioned tests, we now know that the FPGA-Arduino-computer communication works, but interaction with the real chip does not work. That means it is time for the oscilloscope to shine. In order to test the signals coming from the chip, only the LOAD and CLOCK wires were connected between the FPGA and the SPAD chip. The chip was also powered. The FPGA feeds a clock signal to the chip, and periodically sends LOAD pulses to the chip. When receiving LOAD, the chip is supposed to send a flag bit (a one) over all 16 serial wires, and then send the photon events encoded serially behind the flag bit. With the oscilloscope, it was confirmed that the CLOCK and LOAD signals arrive all the way up to and including the bonding wires. Unfortunately, the chip does not drive the wires. No flag bits at all are visible. And when connected to the FPGA, the values stay around 420 mV, which is the standard value that the FPGA input pins are pulled to when they are not driven. As an additional test, the FPGA’s input pins were pulled to ground and to 1.1V with a resistor of 390 kΩ. The measured voltages were respectively 1.07 V and 50 mV for the pulled up and pulled down cases. The signal that the FPGA transmits also matched the inputted values. This indicates that the FPGA’s input pins behave as expected, but that really the SPAD chip is somehow not driving the serial wires, which it is supposed to do.



(a) Test result with an artificial signal spelling ”QIT”. The FPGA mimics behavior of the chip with a separate module, which is then read out as if it was the SPAD chip. The counts are then physically passed to the Arduino and on to the computer. This indicates that the FPGA-Arduino-computer communication works. ”Frame 23” indicates the 23th frame of the live image, and can be ignored for the purpose of this test.



(b) Test result with the real SPAD chip connected to the FPGA. Undefined signal is visible, which changes every frame. The top and bottom rows are biased to mostly give ones, while the middle most often gives zeros. ”Frame 9” denotes the 9th frame in the live signal. This is arbitrarily chosen as a snapshot.

Figure 5.11: FPGA test results.

5.6. Testing the final setup

In chapter 3 the design of the new widefield ODMR testing setup for the chip was described. By the logic of the section above, the chip does not work, and thus no widefield ODMR can be done with it. This unfortunately breaks the main goal of the thesis. Nonetheless, the final experiment was proceeded with two different types of photo-sensors, such that the working principle of the testing setup could still be demonstrated. The first sensor is a single SPAD on a PCB, borrowed from NXP. The second is the same imager that is also used for aligning the diamond in the confocal setup (see chapter 3). Although the imager is not single-photon, it is still a sensitive device, and it has many pixels. So many in fact, that the pixels cannot be processed in a feasible time with a sequential program. Fortunately, the GPU-accelerated fitter from chapter 4 is able to do this processing. This computation took 2.5 hours on the lab computer.

5.7. Tests with alternative SPAD

In order to utilize the borrowed SPAD from NXP, an SMA cable had to be adapted to the banana inputs on the main PCB where the comparator and FPGA connectors are. That way, the FPGA can be used to count the photon events, and the whole setup essentially stays the same as planned, except for having one borrowed SPAD instead of the chip with 16x16 SPADs.

In order to reduce noise, a custom adapter was soldered to minimize the cable length that the signal would have to travel unshielded. Figure 5.12 shows this adapter connected to the main PCB. Despite the effort to minimize signal loss, the main PCB received too much noise, and the SPAD signal could not be reliably amplified. Figure 5.13 shows the measurements done with the oscilloscope that lead to this conclusion. As indicated by the figure, the SPAD signal looks perfect when connected to the oscilloscope with an SMA-BNC cable. In order to connect the SPAD to the PCB, the SMA cable needs to be slit in banana cables, which introduces too much noise while also losing voltage of the SPAD signal (50 mV noise vs 20 mV SPAD pulses). Still, inbetween the waves of noise, SPAD peaks could be distinguished. Nonetheless, this signal is not suitable to be amplified. Besides using the comparator of the PCB, an external RF amplifier has been given a try. The SPAD can be plugged in with an SMA cables, thus keeping the signal integrity well. The output of the RF amplifier did show amplified pulses, but with multiple oscillations at each pulse. As the number and voltage of these oscillations was not consistent, this signal is also not suitable to be digitalized. With more time ($\sim$ weeks), better electronics could be prepared to make this work. Sadly this experiment was done at the end of this thesis and thus no more time was left.

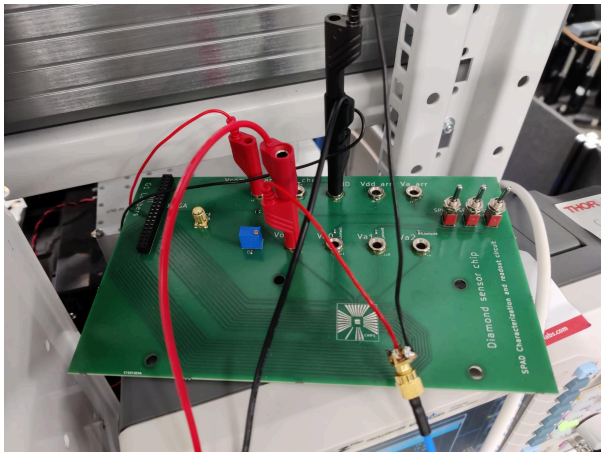
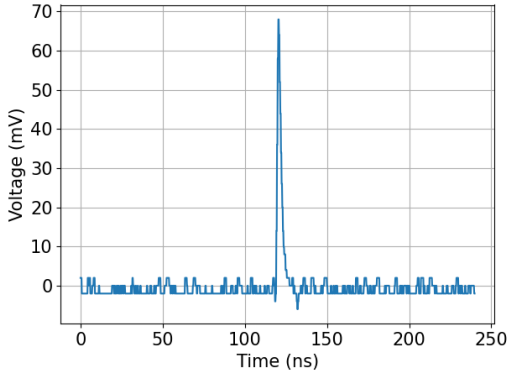
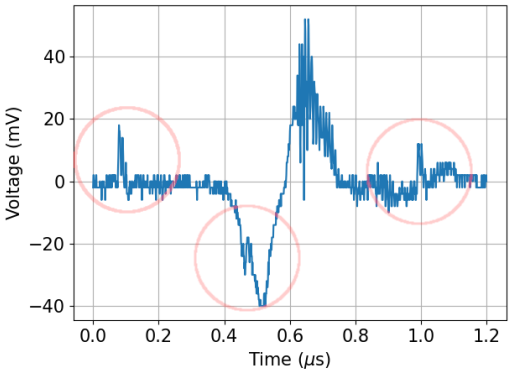


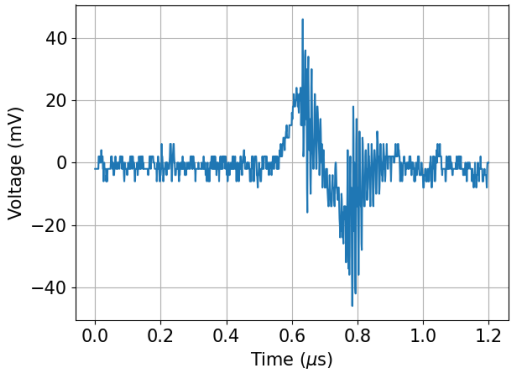
Figure 5.12: Picture of the main PCB. The SPAD borrowed from NXP is connected with an SMA cable to the main PCB. Different power supplies are connected to power the SPAD, the comparator and to give them common ground. Unfortunately, the SPAD pulses enter the PCB weakened, and the cables pick up external waves that reach higher voltages than the SPAD pulses, thus making amplification not feasible.



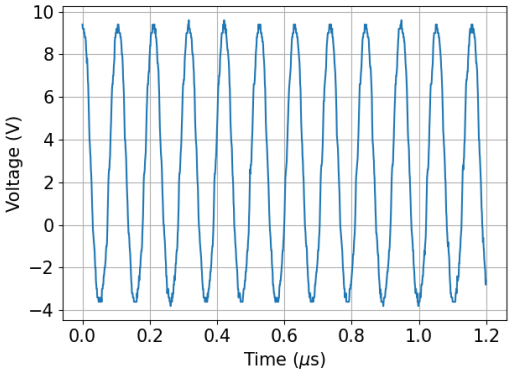
(a) The SPAD signal when connected directly to the oscilloscope with an SMA cable. The SPAD consistently gives pulses of the same shape, reaching almost 70 mV.



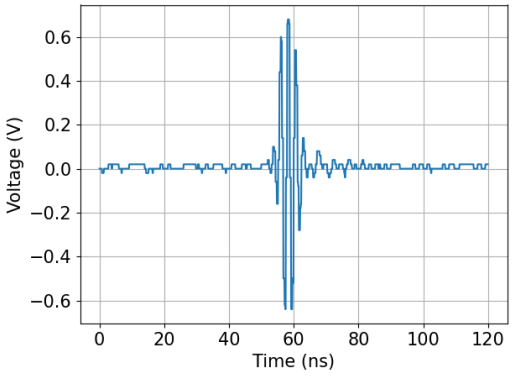
(b) The SPAD signal measured at the connection point where it enters the PCB. Comparing this with the same measurement while the SPAD is off, see figure 5.13c, indicates that the smaller pulses are weakened versions of true SPAD pulses, engulfed in bigger pulses induced by external noise.



(c) Measurement of stray signals at the connection point where the SPAD signal enters the PCB, with the SPAD off. Comparing this with figure 5.13b shows that the little peaks are probably weakened versions of true SPAD pulses, engulfed in bigger pulses induced by external noise. Any recognized pulses are marked with a red transparent circle.



(d) Output signal of the comparator integrated in the PCB. The comparator only gives nonzero output when the SPAD is on, indicating that a form of amplification is happening. Nonetheless, its output is almost a perfect harmonic wave. Note that the expected shape is random square pulses every time a SPAD pulse comes in. The voltage range is also much bigger than what the comparator should theoretically give, which is between 0V and 5V, which are the two logic states.



(e) Amplified signal of the SPAD using an external RF amplifier instead of the one integrated in the PCB. This amplifier has SMA connectors, thus the SPAD signal is kept relatively noise-free until it enters the amplifier. Just like the PCB-integrated amplifier, this RF amplifier shows oscillations when the SPAD is on. Each pulse seems to cause some oscillations that stop after roughly 10 ns. This is already more promising than the PCB-integrated comparator, but still not suitable to connect with the FPGA, as it expects one square pulse for each photon event. If every photon event consistently had X oscillations, that could have been accounted for, but the number of oscillations is not consistent between pulses. Furthermore, the voltage is not high enough for the FPGA's lowest voltage mode.

Figure 5.13: Oscilloscope measurements with the alternative SPAD. Unfortunately, the available electronics in the lab were not sufficient to be able to properly amplify the SPAD signal. That is a necessary step to count the photon events.

5.8. Tests with imager as main sensor

The imager has been mounted in the widefield setup. The little aluminium table that was originally meant to hold the main PCB has been edited with a square slot to hold the imager with its photosensitive area at the same location where the CMOS SPADs would be. A 30mW laser was aimed into it. 30mW was chosen because the laser gets imaged to a circular area instead of a single spot, so it needs more power to get the same power per unit area. This resulted in a lot of specks, making it hard to distinguish what was being seen. The setup was also tested with a white LED instead of the laser, which produced a much clearer image. The images obtained are shown in figure 5.14. Unfortunately, there was no suitable microwave antenna available that fit between the objective and piezo stack the way they were mounted. Later the antenna got made, see chapter3, but it had very bad signal.

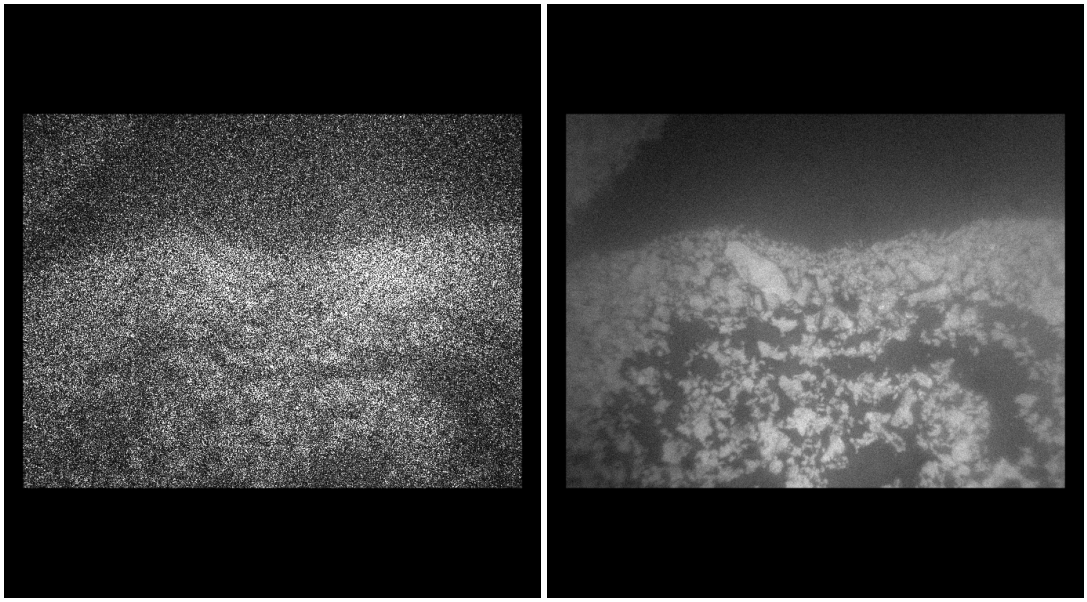


Figure 5.14: Camera images through the widefield setup. Left: Diamond is excited by a 30 mW green laser, and the imager is seeing the red fluorescence light which is filtered by dichroic mirror. Right: A white LED shines where the laser is supposed to be. The imager is seeing a clear image through the dichroic mirror.

In order to assess whether the imager is sensitive enough to be able to do ODMR, the imager was tested in a slightly modified version of the confocal setup so that the fluorescence light is redirected towards the imager. Figure 5.15 shows the view of the imager with 0.2 seconds collection time, which is the same as what is used in other ODMR experiments. The imager is capable of seeing the red dot coming from NV fluorescence, even if it is only up to 25 counts. Note that, not being a single-photon detector, every count on the imager consists of many photons reaching it. 2D ODMR was attempted by sweeping the frequency while acquiring frames from the imager. Due to the low amount of counts, fitting a double-dip does not work. Nonetheless, an ODMR dip is visible around 2.87, with a contrast of about 10 - 15%. This is shown in figure 5.16. This indicates that ODMR with the imager is possible if light is collected for a longer duration.

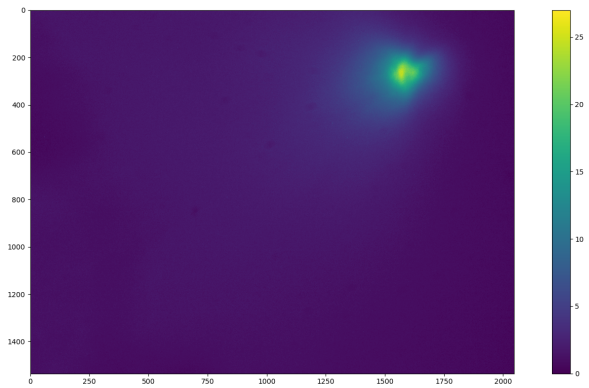


Figure 5.15: Fluorescence light collected by the imager. The focused spot where NV centers get excited and emit fluorescence light is clearly visible, with the brightest spot being 25 counts.

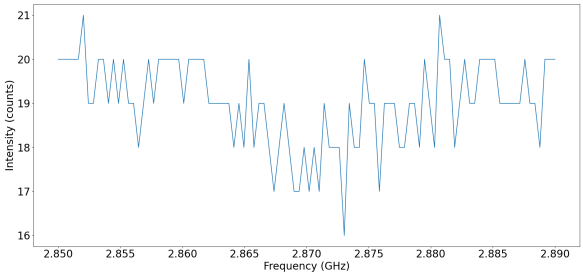


Figure 5.16: The ODMR graph of one of the brighter pixels in the camera image. Due to the low amount of counts, heavy discretization effects occur, preventing normal fitting routines from properly fitting Lorentzians. Nonetheless, a significant dip of about 10 - 15% contrast is present at 2.87 GHz, indicating that proper ODMR is possible with the imager if light is collected for a longer time.

5.9. Bio imaging

Finally, at the end of the thesis, bio-sensing was attempted on the bio-sample diamond. The biofilm on the diamond contains HEK cells ($\sim$ 12 micron in size) injected with magnetic nanoparticles with concentration 40 μg/mL. These particles are paramagnetic, so by applying an external magnetic field, small localized disturbances in the magnetic field should become measurable where the cells are. A magnet has been placed on the piezo stack, as shown in figure 5.17. This was chosen such that the ODMR peaks completely split apart, but are also not unnecessarily far apart, as this would slow down the sweeps. This resulted in around 30 MHz splitting across the whole scanning area, which corresponds to a magnetic field strength of 1.07 mT.

The bio-imaging has been done with the confocal setup, but without the beam expander. The confocal setup had to be slightly modified multiple times by other students, and this resulted in having to align the setup many times. The beam expander was omitted to reduce the alignment time to under half an hour. Figures 5.18 and 5.19 show the peak splittng and PL results respectively of the 2D ODMR on the biosamples. The effect of the external magnetic field is clear, being about 30 MHz splitting. No clear localized magnetic fields of the right size stand out above the noise, thus no HEK cells are detected. On the photo-luminescence graph, dark blobs of the right size are present that could indicate HEK cells. This indicates that the parameters for bio-imaging have to be revised and tried again in the future. Besides peak splitting and photo-luminescence, 2D ODMR results a few other parameters. The full results of all bio-measurements can be found in appendix chapter C. The average measured peak splits are all between 30 - 31.6 MHz. As this is smaller than the standard deviation of the splitting (see subsection 5.9.1), there is no statistically significant difference between the average magnetic field strengths among the experiments.

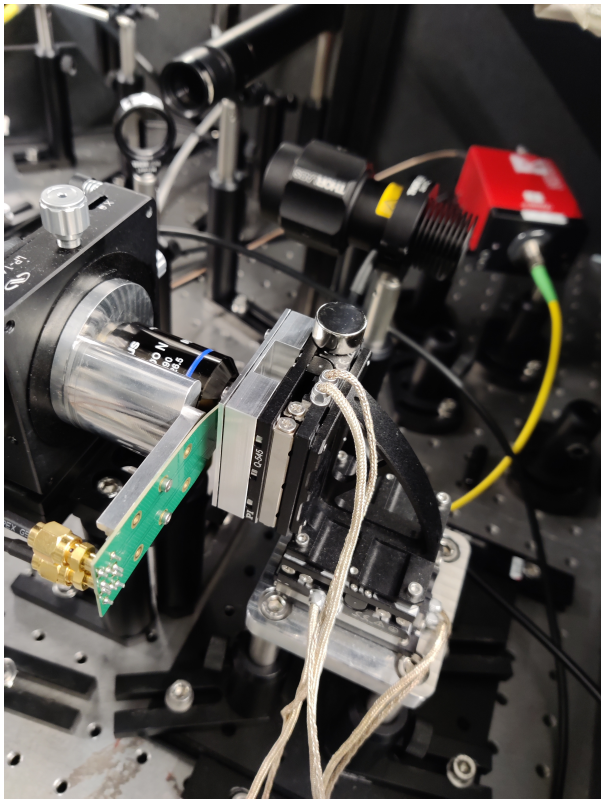
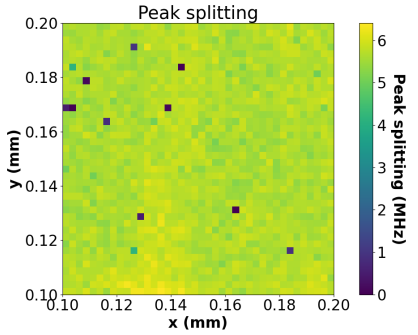


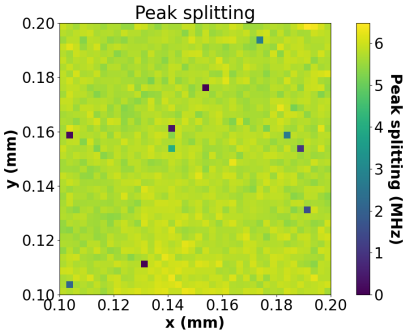
Figure 5.17: A magnet is placed at the corner of the piezo stack. Due to magnetic interaction, the magnet tightly snaps on the corner of the stack. The magnet has a horizontal distance of 27mm and vertical distance of 25mm to the diamond center. This results in a peak splitting in the ODMR graph of about 30 MHz, which corresponds to 1.07 mT magnetic field strength.

5.9.1. Magnetic sensitivity estimation

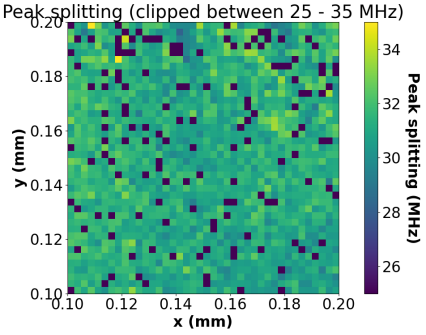
Equations 2.4 and 2.7 provide two ways of calculating the magnetic field sensitivity. The first equation is the definition. In order to use the definition, an estimate for the standard deviation has to be known. The bio-measurements are 2D scans with a lot of data points, but each data point is a different location. In order to properly estimate the magnetic standard deviation, multiple measurements of the same location have to be available. The standard deviations of the bio-measurements as a whole are all around 1.8 MHz, which corresponds to 54 μT. Taking into account the sweeps take about half a minute, the sensitivity would be 293 μT/$\sqrt{\text{Hz}}$. However, the standard deviation of 1.8 MHz is largely overestimated because of other factors such as non-uniformity in the external magnetic field and inherent differences between the pixels. Hence, 293 μT/$\sqrt{\text{Hz}}$ is a largely overestimated upper bound, and another method is needed to make a better estimate of the sensitivity of the system. The second of the above-mentioned equations is an approximation for magnetic sensitivity that holds when shot-noise is the limiting factor of the sensitivity. By inspecting the graph (see appendix chapter E), the shot-noise is in fact a dominant contributor to noise in the obtained magnetic field, as the noisy data points can cause small shifts in the exact optimum that the double-dip fitting process finds. The ODMR graphs from the bio-measurements show a contrast of about 5%. The full width half maximum (FWHM) of the dips are around 12 MHz. The typical countrate was 140 kcounts/s. Using equation 2.7, a magnetic sensitivity is estimated of 15 μT/$\sqrt{\text{Hz}}$. Note that the bio-measurement is not tuned for maximum sensitivity. The laser power was set to 0.3 mW which is lower than the optimum. This way, the risk is lowered of cells getting burned. Furthermore, the confocal setup has to excite diamond through the bio-film, which introduces optical inefficiencies. Other non-bio-measurements with the same confocal setup have been performed with magnetic field sensitivities < 10 μT/$\sqrt{\text{Hz}}$.



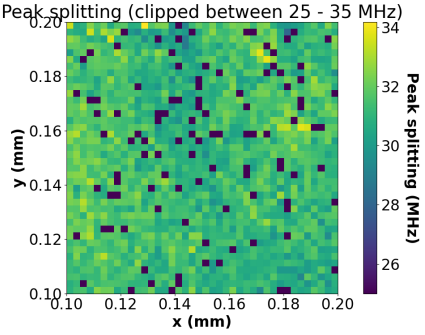
(a) No external magnetic field, on the diamond surface.



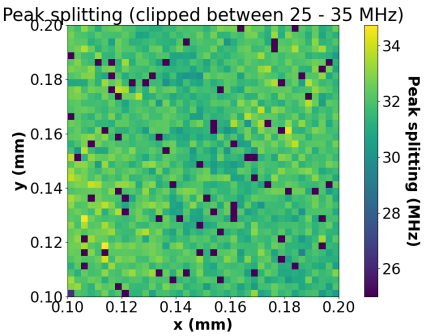
(b) No external magnetic field, 10 micron under the diamond surface.



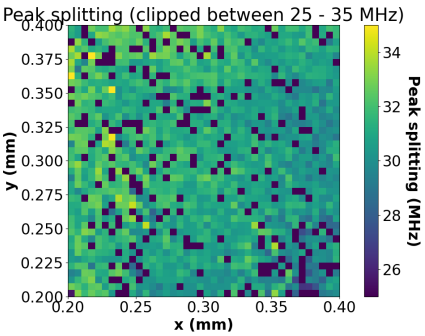
(c) With external magnetic field, on the diamond surface.



(d) With external magnetic field, 10 micron under the diamond surface.

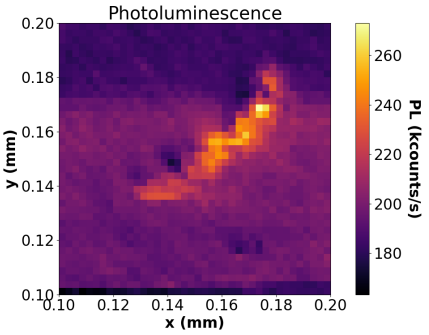


(e) With external magnetic field, 50 micron under the diamond surface.

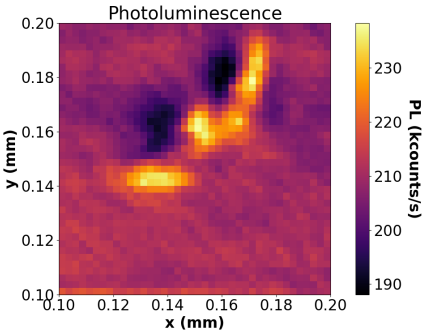


(f) With external magnetic field, larger 200x200 micron scanning area, on the diamond surface.

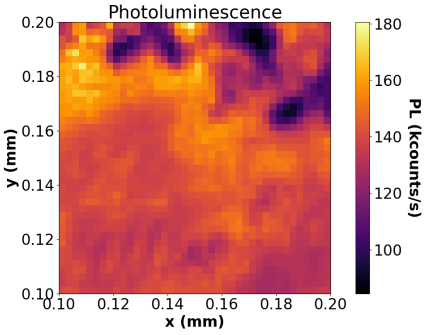
Figure 5.18: The peak splitting graphs from 2D ODMR on the biosample. An external magnetic field was applied that resulted in about 30 MHz splitting, which corresponds to about 1.07 mT magnetic field strength. Notice that the images have specks that have significantly outlying values compared to most pixels. These have been partly inspected, and are due to unsuccessful fits.



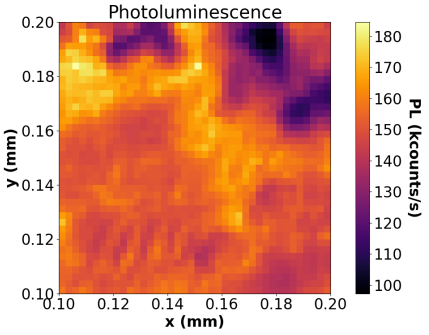
(a) No external magnetic field, on the diamond surface.



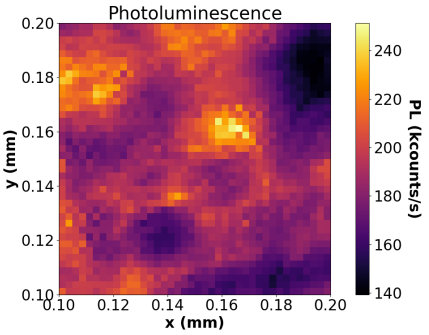
(b) No external magnetic field, 10 micron under the diamond surface.



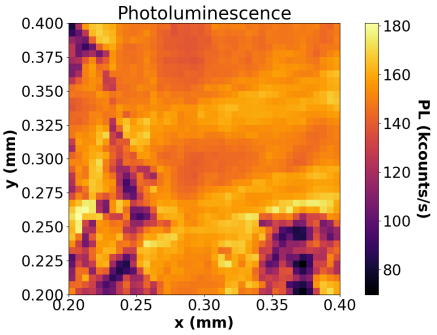
(c) With external magnetic field, on the diamond surface.



(d) With external magnetic field, 10 micron under the diamond surface.



(e) With external magnetic field, 50 micron under the diamond surface.



(f) With external magnetic field, larger 200x200 micron scanning area, on the diamond surface.

Figure 5.19: The photo-luminescence graphs from 2D ODMR on the biosample. Notice the dark spots. The average HEK cell should appear with a diameter of about 4 pixels, except for the bottom right plot, where the cells should have diameter of 2 pixels. Some dark spots are a reasonable size to be a HEK cell that might be blocking part of the light.

6

Discussion

During the course of this thesis, many challenges needed to be overcome, sometimes needing improvised or otherwise sub-optimal solutions. In this chapter, a closer look is taken to these solutions and their implications for the validity of the experiments.

6.1. Inconsistent placement of microwave wire

In both the confocal and widefield test setups, the microwave field is driven by a wire soldered on a PCB, which has a special holder to hold it in front of the objective. Before mounting the MW PCB, the wire needs to be bent in a trapezoid shape such that it forms a straight line going just barely outside the field-of-view of the objective, in between the objective and the diamond. Bending the wire in a trapezoidal form by hand is hard, and it will always end up with a wavy imprecise shape. By applying small forces on the wire, it could be bent straighter, but this does not fully solve the issue. Also, screwing the PCB at the right place is not trivial. To mount the MW PCB, it is first put in view of the imager. Then it is slowly moved away until it is fully out of the screen, at which place the MW PCB is screwed tight. Despite this clear procedure, the act of screwing it tight imposes small deviations in the exact position of the PCB. Rescrewing the PCB multiple times, it was found that the effective MW power varies visibly between different attempts for mounting the MW PCB. Nonetheless, as the ODMR continued to work, the inconsistent quality was accepted for the time-being.

To solve this issue in the future, a fine stage can be used to make fine adjustments to the PCB position. To get the MW wire completely straight, more precise tools are needed.

6.2. Driving the diamond with microwaves while leaving enough space

As mentioned in chapter 5, two MW antennas were experimented with. One goes inbetween the diamond and the objective, which works very well for driving transitions, but stands in the way when the diamond needs to be moved closer. This is needed to measure the back surface of the diamond (the surface farthest from the objective). An alternative ring-shaped antenna that goes around the objective was tested, but did not give good performance. Hence, a good antenna that does not stand in the way and yet drives transitions efficiently, is still welcome, as it would open up opportunities for more measurements.

In order to make this possible, new antenna designs can be made with thinner and/or straighter wires, or alternatively a ring as small as possible that fits tightly around the laser beam. Another option would be to amplify the MW signal even more, for which another RF amplifier will need to be bought and installed.

6.3. Voltage mismatch

As mentioned in section 3.9, the SPAD chip works at 1.1V, while the FPGA's lowest possible GPIO logic level is 1.5V.

From 1.5V to 1.1V, this was mitigated by using voltage dividers. This resulted in some distortion in high-frequency signals, thus causing a need to reduce the clock speed, which was lowered to 1 MHz.

From 1.1V to 1.5V, it was determined that the FPGA interprets everything above 975mV as a logical high, and thus should still be able to read 1.1V signals. The downside is that this leaves very little margin for noise, such that consistent readout cannot be guaranteed. One can argue that missing a few counts is not a critical flaw in the function of the FPGA design. Still, for optimal results, this should be avoided.

In order to solve this issue in the future, a proper level-shifter needs to be designed for this system. It needs to be able to handle at least 50 MHz signals in order to fully make use of the FPGA's clock speed. Alternatively, a faster FPGA can be bought with 1 GHz clock speed, such that the SPAD chip's rated frequency can be fully utilized. Of course, doing this only makes sense if a new fixed version of the SPAD chip is fabricated, which is elaborated on in section 6.5. In order to go to higher frequencies, the PCB and its connections will need to be upgraded, which is addressed in the next section.

6.4. Upgrading the PCB for higher frequencies

More effort was put into making the system functional, rather than optimizing its components and parameters. For MHz range frequencies used in the course of this thesis, the PCB design was sufficient to not cause noticeable distortion in the signals. However, to fully utilize the SPADs maximum count rate of 3 Mcounts/s, the system should operate at a frequency of 1 GHz. This requires a more complex PCB design that treats wires and copper tracks as transmission lines. Furthermore, the 40-pin GPIO cable might give problems at GHz frequency due to the wires being parallel and close to each other for tens of centimeters, which can cause signal distortions and cross talk. Hence, for high frequencies a new PCB needs to be designed that has calculated transmission lines, matching track lengths, and uses differential pairs to handle fast communication over long wires. Most FPGAs support differential pairs and have suitable connectors for GHz communication.

6.5. Defectiveness of the SPAD chip

As mentioned in chapter 5, the SPAD chip was defective in two ways: Firstly, the analog SPADs do not quench properly. The analog SPADs rely on external quenching, which was done passively with a resistor. Different resistor values and operating voltages were tried without success. This indicates that the cause of the problem lies within the SPAD itself. Secondly, the chip did not drive its electrical outputs. No response at all could be measured from the digital communication circuitry. This is a strange problem, because lack of signal is expected to be a hard logical zero. Instead, the voltage became undefined (a floating value between logical high and logical low). This indicates a potential problem in the circuit that drives the outputs. Not being able to perform tests on the SPADs broke the main goal of the thesis, and experiments were continued with alternative photo-sensors.

In order to solve this issue for a future tape-out, a close look has to be taken into the design files of the chip. This is in the domain of electrical engineering, and was considered outside the scope of this thesis.

6.6. Amplification of NXP's SPAD

As mentioned in chapter 5, the connection between the SPAD and the PCB was too noisy. The connection has to be split from SMA to banana, and the banana part picks up too much noise as well as imposing a significant voltage drop of the SPAD pulses, making the amplitude of external interference higher than the peaks of the SPAD. Due to this problem, the SPAD signal could not be amplified. A possible way to solve this is by using the time tagger to count the pulses. The time tagger has configurable threshold voltage between -1.5V - +1.5V, and accepts SMA input, thus preserving the signal integrity optimally. That does mean that the FPGA and Arduino are not needed anymore for testing with NXP's SPAD. Thus, it would serve the purpose of demonstrating that the optics work, but not be a complete test for the FPGA design.

6.7. Using the imager for widefield ODMR

During the preliminary test in the confocal setup, the imager was shown to detect an ODMR dip. Only having up to 15 counts, the graph was very discretized and no Lorentzian dip can be fitted. However, this has potential. The test was done with 0.2 seconds dwell time, which is the same used in the other ODMR experiments of this thesis. When longer photon collection times are used, the imager can gather more counts and thus get a better-resolved graph. Comparing the imager and the SPCM both with 0.2 seconds dwell time is unfair. The imager has about 3 million pixels, so it can do ODMR on many pixels simultaneously. A 2D scan using confocal ODMR takes a long time, as each pixel has to be swept one by one. For a high resolution 2D ODMR experiment, the imager has margin to use much longer photon collection times and still be done earlier than the confocal setup. Possibly with $\sim$ hours of sweeping, the imager can already have a good 2D ODMR measurement,

7

Conclusion and outlook

The goals were to improve ODMR software, design a test setup for CMOS SPADs, and to characterize and test the CMOS SPADs.

7.1. Improvements of ODMR software

Many software upgrades have been done on the ODMR software. The bug that prevents the AWG from initializing has been patched. The problem of the piezo stack getting stuck has been resolved, resulting in no more fatal interruptions during long measurements. This way, measurements of up to a whole weekend have been achieved without the system halting prematurely. Software-side tilt correction has been implemented, which is able to handle an arbitrary amount of tilt, only limited by the diamond hitting the objective. This is roughly 100 micron in the z-direction, and translates to maximum tilts roughly between 100 micron/mm for 1x1mm^2 diamonds to 10 micron/mm for 10x10mm^2 diamonds. Multiple measurement routines were implemented along with a post-processing script that can process their data. An improved fitting routine has been developed that has a successful fit in typically about 95.4% of the pixels in 2D scans. This fitter supports GPU acceleration and was able to process $\sim$ 3 million pixels in 2.5 hours.

7.2. Development of a widefield test setup

The widefield setup was fully constructed. As the CMOS SPADs did not work, instead the imager was used to confirm that the view through the optics is as expected. Using a temporary modification of the confocal setup, the imager was shown to be sensitive enough to detect an ODMR dip above 10% contrast, even though it had just 16 - 25 counts, indicating that it could be used for widefield ODMR if longer photon collection times are used. When mounted in the widefield setup, the imager saw a lot of specks with the green laser, which is not desirable for widefield imaging. This might potentially be resolved by making sure all optics that handle the green laser are clean. Another problem of the new setup was that the microwave antenna does not fit well between the objective and the piezo stage. The edges of the piezo stage protruded farther than the diamond itself. This can easily be avoided in the future by mounting the diamond on a plate with a protrusion, like the plates that we use for confocal ODMR. Such plates have to be made custom, but are feasible to do so, as the same has already be done for the confocal setup. Another option would be to use the new ring-shaped microwave antenna that needs less space, but this comes at a huge trade-off for microwave power, giving only about 1 - 2% contrast where it otherwise would give 8% contrast.

Besides the optical part, the PCB, FPGA, Arduino and computer were fully setup with necessary software and connections. This electrical part of the system could not be tested with the real CMOS chip as it was not working, but using an artificial signal it was shown that the chain from the FPGA towards the computer functions correctly.

7.3. Characterization and testing of SPADs

The analog SPADs available on the CMOS chip have been tested. By using the oscilloscope, it was discovered that they do not produce pulses. Nonetheless, their IV curves looked like typical diode curves with breakdown voltages in the expected range, being generally between 15 - 17 V. The CV measurements failed for most SPADs, although some show a clean curve within the expected range, showing capacitance values in the hundreds of fF. Due to the SPADs not showing avalanche behavior, the optical characteristics could not be measured, which are photon detection probability and dark count rate. Assuming that the SPAD design contains one or more errors that can be resolved, its characteristics will remain unknown until the next tape-out is done and the SPADs are delivered for testing.

The digital SPAD arrays were also tested, but it was discovered that the chip does not drive its outputs despite being powered and connected as in the specifications. This indicates that there might be a design error in the digital communication part of the CMOS design.

7.4. Bio-sensing

As the main goal is to build an integrated quantum bio-sensor, an attempt was made to do a bio measurement using the confocal setup. Despite having good signal and applying an external magnetic field, no HEK cells were detected. The cause of this issue is unclear and needs further investigation.

7.5. Outlook

This thesis contributes to the grand goal of developing an integrated quantum bio-sensor. The CMOS SPADs are an important component in the concept. However, this chip likely contains one or more design errors that need to be checked. Are the SPADs suitable to be used in the integrated bio-sensor? Without being able to properly characterize them, the question remains undecided.

In the future, the SPAD designs can be checked and a new tape-out can be done. This way, they could be characterized. This could potentially be done simultaneously with the waveguide part of the integrated quantum bio-sensor. As scientists, we practice the patience to work on the same project for years, even when things are not working at first. The integrated quantum bio-sensor certainly has potential, but one thing is clear: There is still work to be done.

Acknowledgments

Even though not all experiments went well, it has been an awesome and fun project to work on. I want to thank my supervisors Yannis Varveris and Ryoichi Ishihara for making this project possible and providing me with everything that I needed. Besides supervisors, there are many other people who have helped me in one or more ways, which will be listed below.

People from the research group

From the people of the lab, I want to thank Salahuddin Nur for always being there for me, and providing interesting diamond samples from his bonding experiments. I would also like to thank Nikolaj Nitzsche for his great teamwork in the lab. Also thank you to Rohan Acharya and Zdravko Zdravkov, who were willing to be flexible with me while we both needed the same setup close to our deadlines, and being good teamplayers. Furthermore, thank you to Anty Guo for making a great atmosphere in the lab, keeping me mentally sane while working, and also for the delicious food. Thank you Erbing Hua for helping me use the probe station for IV and CV measurements on the chip. Finally, thanks to everyone else from the lab! All small favors help, even if they are too many to be listed here.

DEMO

I want to thank Yannick van der Linden and Lars Leenheer for soldering multiple microwave antennas for me, even as they got increasingly more complex. I would also like to thank Hans van der Does and Giel Hermans for soldering my PCBs and giving FPGA advice.

QuTech employees

Thank you Jason Mensingh and Erik van der Wiel for being great instructors for using the QuTech wirebonder and 3D printer, and always being ready to help. Thank you Tim Hiep for making beautiful aluminium pieces for me. And thank you Srijit Goswami for being my second examiner.

Misc

Thank you Carel Heerkens for helping me with laser dicing. Thank you Qiangrui Dong and Daan Brinks for preparing the biosamples. Also thanks to Sadik Ilik and Robert van Veldhoven for support with electronics surrounding SPADs. And last but not least, thanks to my family and friends! Your unwavering emotional support and love is what kept my mind from collapsing!

A

Overview of developed code

During the course of this thesis, much code has been developed both for driving the confocal ODMR setup and for the widefield testing setup. This was built on top of already existing code from TNO as well as other students Alfie, Filip and Harsh. I would like to thank them, as part of my code consists of copied pieces from them that were put together. The code is available on my Github with through the following link: `https://github.com/Dylan-Aliberti/MEP_SPAD_quantum_bio_sensing/tree/main`. The Github page also contains elaborate instructions for future operators of the confocal setup that want to use it.

A small overview of all the code is given below. Elaborate explanations can be found in the main content of this thesis, and the overview below only serves as a compact list of all features. For future operators of the setup that want to know how to use the scripts, please go to the above-mentioned Github page and read the README for detailed instructions on usage.

A.1. Confocal setup

This was the setup that already existed at the start of the thesis. Some code from TNO works, but not all has been adapted. Two scripts have been made to handle 2D scans and processing. These started out as copied code from previous students put together in a compact script, but grew with lots of features added. These scripts are ODMR_2D.py and ODMR_2D_process.py, which handle the scanning process and the post processing respectively. See below their features listed.

File: ODMR_2D.py

This file contains all measurement routines and some other code that helps and/or stabilizes the measurement process:

- Measurement routines
  - 2D photo-luminescence (PL) scan
  - 2D ODMR scan
  - z-scan
  - 3D PL scan
- Utilities
  - Saving settings in JSON format
  - Time estimation
  - Optionally triangular scanning area
- Piezo stack handling
  - Serpentine scanning pattern
  - Periodic autozero

File: ODMR_2D_process.py
This file handles post processing of the measurements. This file reads the JSON settings file along with the main data file and is able to see which measurement type was done and draw all relevant plots. It can be called from the terminal and takes a filepath as argument.

File: double_dip_fitter.py
This file has the code of the new fitter, and is a dependency of ODMR_2D_process.py. The fitter has improved success rate and is GPU-accelerated.

File: AWG_init.py
Small script that initializes the arbitrary waveform generator (AWG). This was made as a quick simple patch to the initialization problem described in chapter 4. The 2D measurement scanning scripts do not need this. When using scripts from TNO's RTCS software, this needs to be called. The command to do that has been implemented in the "rtcs esr-scan" command in order not to have to type it every time.

A.2. Widefield test setup
FPGA design
The CMOS chip with SPADs was designed to communicate with an FPGA that handles deserialization, counting and communication with other devices. This was started by copying code from the previous bachelor thesis [4], and further developped to handle the SPAD arrays of the new chip.

- Serial counter module
- Buffer sender module (uses the third party SPI master module)
- Controller module
- Top level entity
- test bench with virtual chip and artificial signal

Arduino
Filename: readout_SPI.ino. The Arduino was used to communicate with the FPGA through SPI and pass the values to the computer through USB. Both are easily done with builtin library, making this a trivial program.

Processing on the computer
Filename: SPAD_array_reader.py. The data received from the Arduino needs to be processed, which is handled by a Python script running on the computer. It detects the start-of-frame signal, interprets all received bytes to full 32-bit integers, and puts them in a 16x16 Numpy array. A plot is made to verify correctness.

B

Dicing the CMOS chips

During this thesis, an attempt was made to dice the CMOS chips using QuantWare's laser dicer. Despite this not being a critical component of the planning of this thesis, useful information was obtained that may benefit future dicing needs.

In order to use the laser dicer, one has to make a layout file that specifies the path of the laser. It has to be a closed loop. The machine has options whether the laser spot should be centered on the path or should stay completely on the inside or outside, hugging the path.

Two of the old chip designs from the previous bachelor thesis were given [4] as first test, along with one of the new designs meant for this thesis if the dicing turns out well. Some bonding pads would serve as markers. As this chip needs to measure light, dirt cannot be allowed to gather on top of it. Laser dicing produces lots of small particles that fly around. In such a case, the chip can be covered in PMMA during the dicing process, which can be removed with acetone later. For this test, PMMA was not applied, but should be for future dicing of similar chips that need to measure light.

Figure B.1 shows the result of the first dicing. Dicing has partially succeeded, but some points of attention came up that must be improved before dicing important expensive chips. They are the following:

- **Markers**
  Bigger markers are needed. Ideally circles with a diameter of 400 micron, which would allow automatic alignment.
- **Glue**
  The chips need to be glued tighter. During the test dicing, one of the three chips got displaced and the other got sucked up by the ventilation system. Wafer dicing foil is recommended, which glues tightly but can be loosened by UV light.
- **Height measurement**
  The machine needs to be able to measure the height of the chip, for which contact is needed. This can potentially damage the chip. If this becomes a problem (no area on the chip can be sacrificed), a possible solution is to make a plate with a pocket that matches the chip's height and size. This way, the machine can measure the height of the plate besides the chip rather than the chip itself, and it will be the same.

Thanks to Carel Heerkens for helping with the laser dicing and giving useful tips!

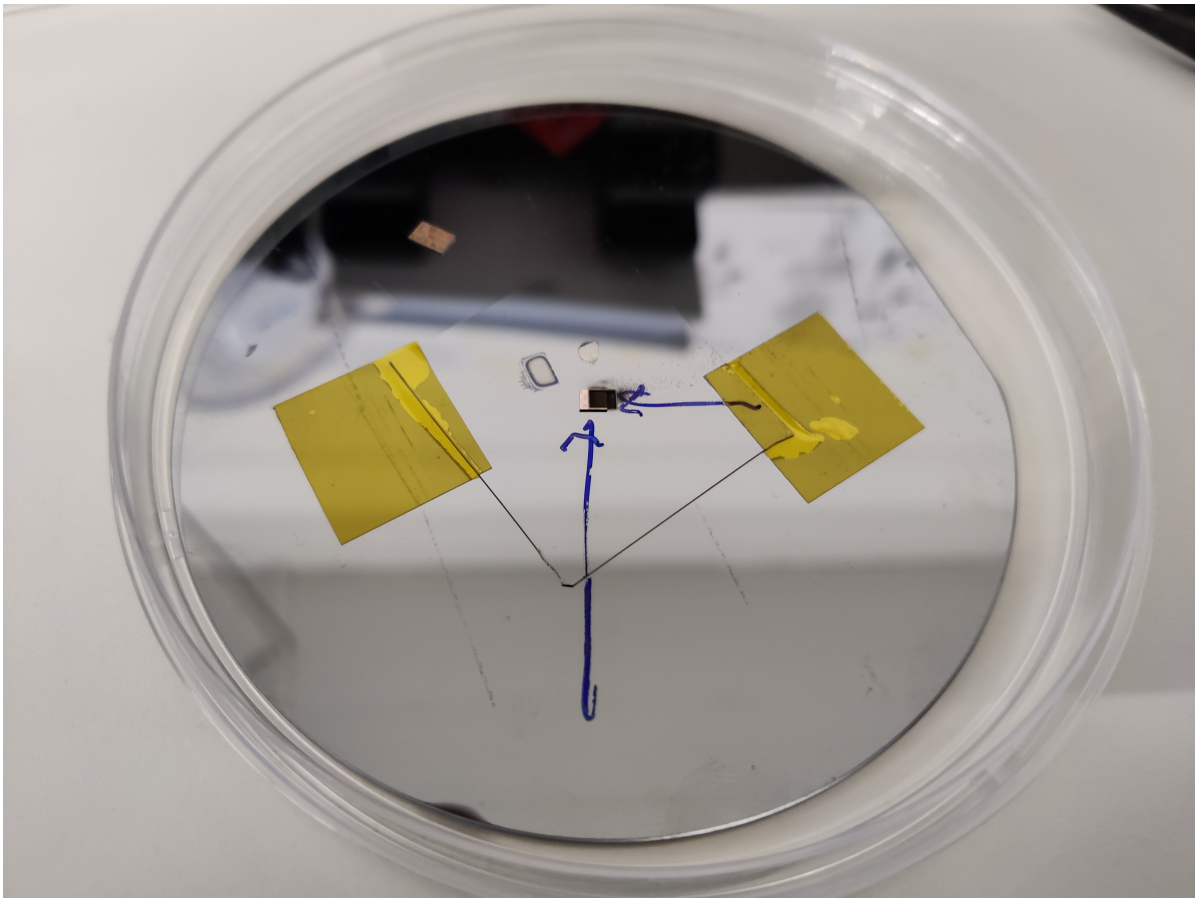


Figure B.1: The chips that were sent for dicing. The 1 inch silicon plate with the chips has been taped to a bigger plate before entering the machine. The middle chip (with the arrows pointing) has been diced. Debris can be seen. The two other spots where the other two chips were glued can be seen. One of them got displaced and can be seen at the top of the image, while the other one ended up in the ventilation system of the laser dicer.

C

All bio measurements

In chapter 5 a bio measurement was shown which was scanned on the surface of the diamond with external magnetic field. In this chapter, all bio measurements are shown, which are either with or without magnetic field and at different depths.

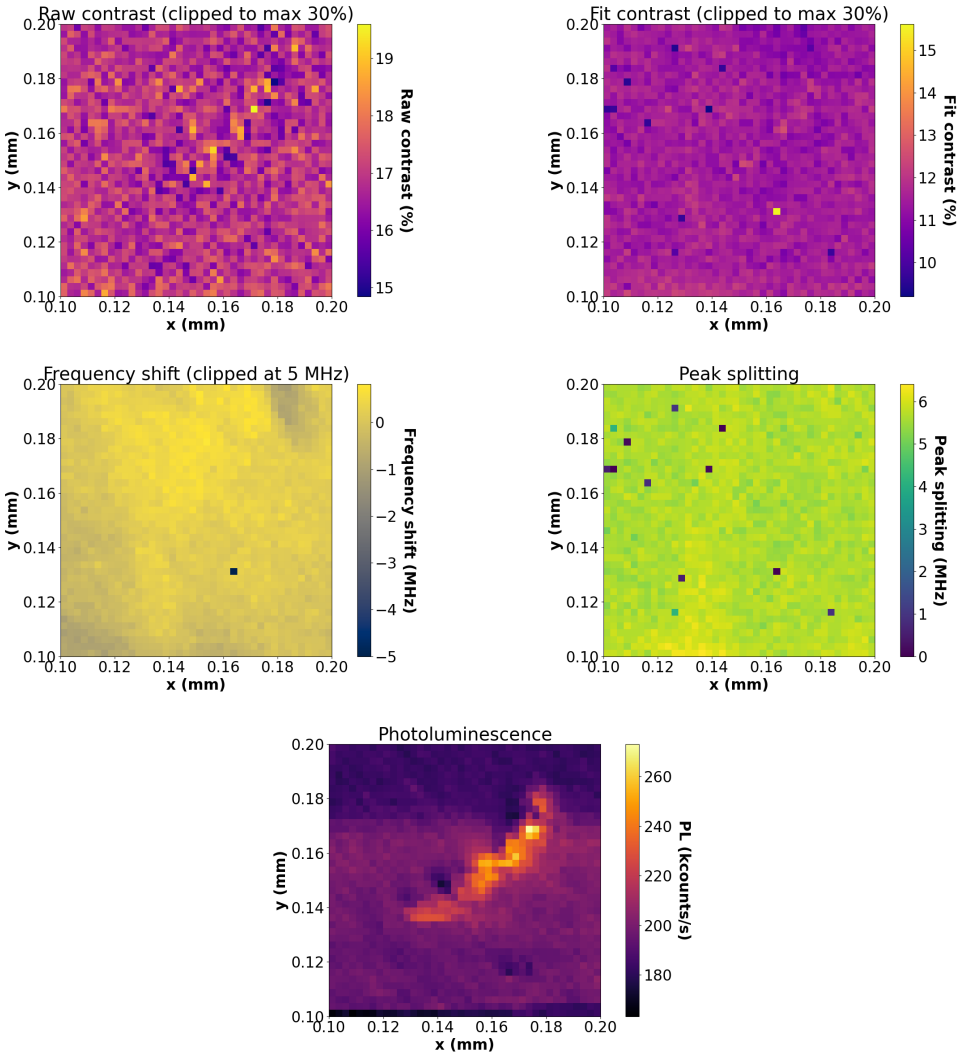


Figure C.1: Without magnet, surface of diamond

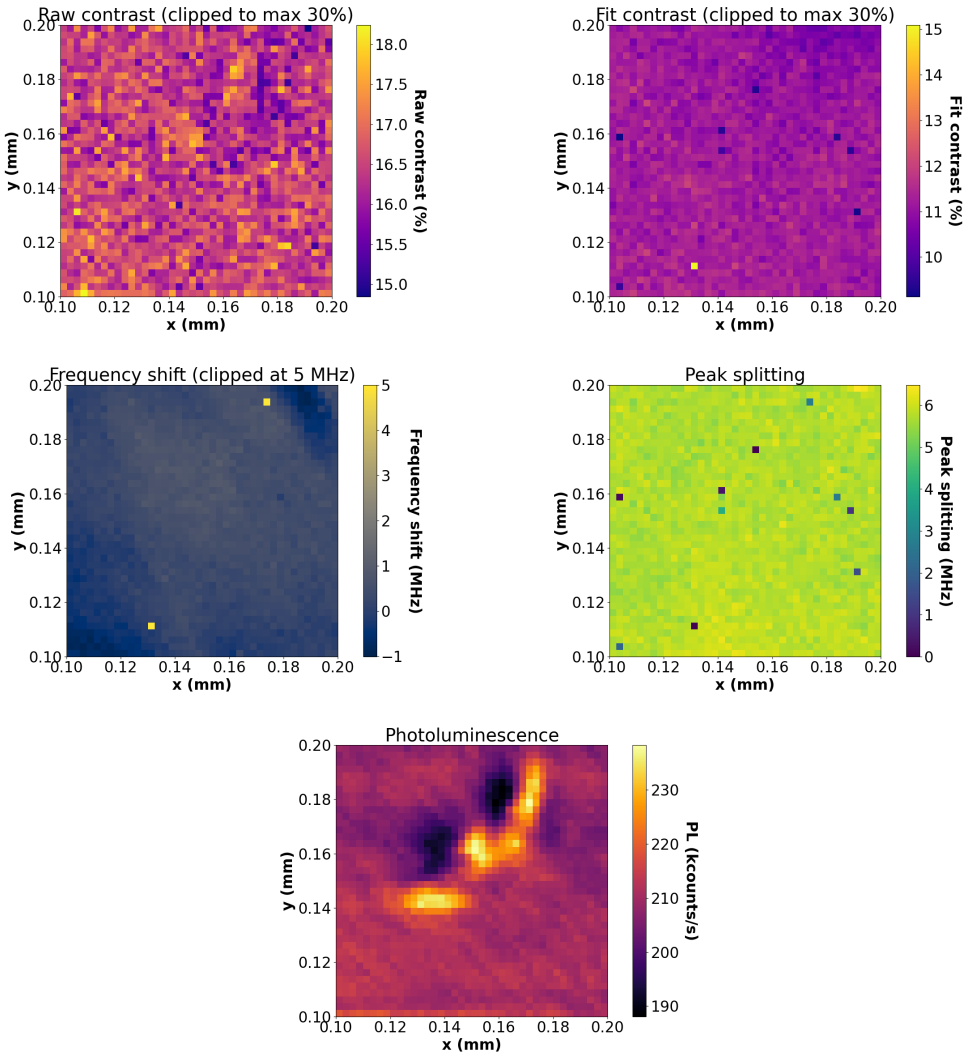


Figure C.2: Without magnet, 10 micron inside the diamond

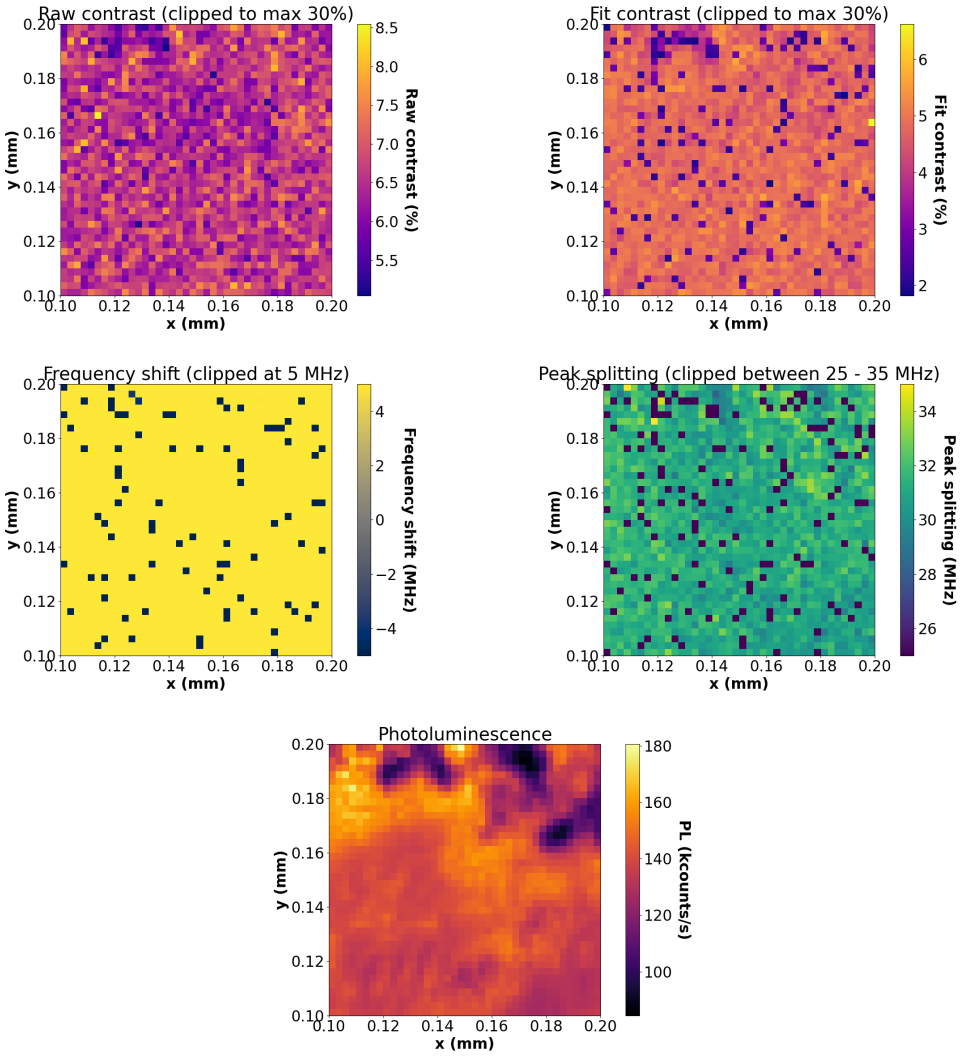


Figure C.3: With magnet, surface of diamond
Note: This is the measurement shown in chapter 5.

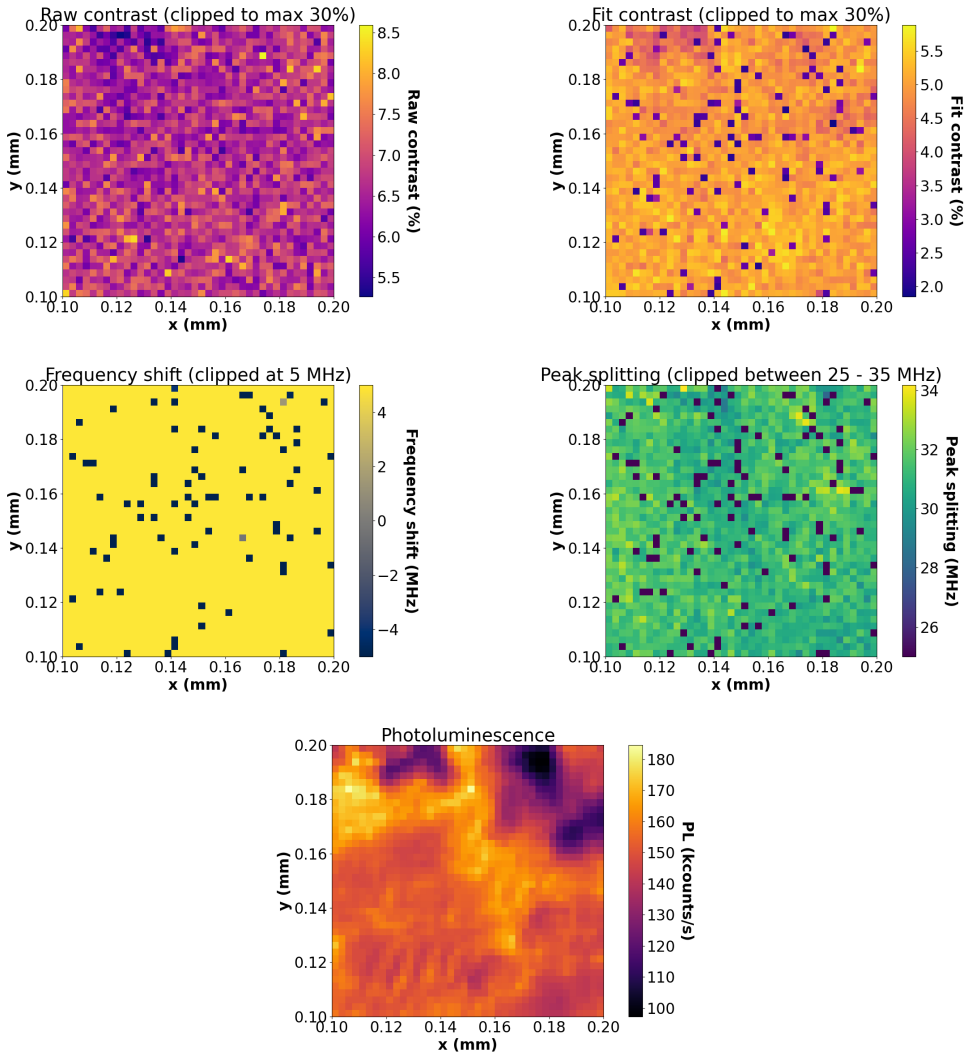


Figure C.4: With magnet, 10 micron inside the diamond

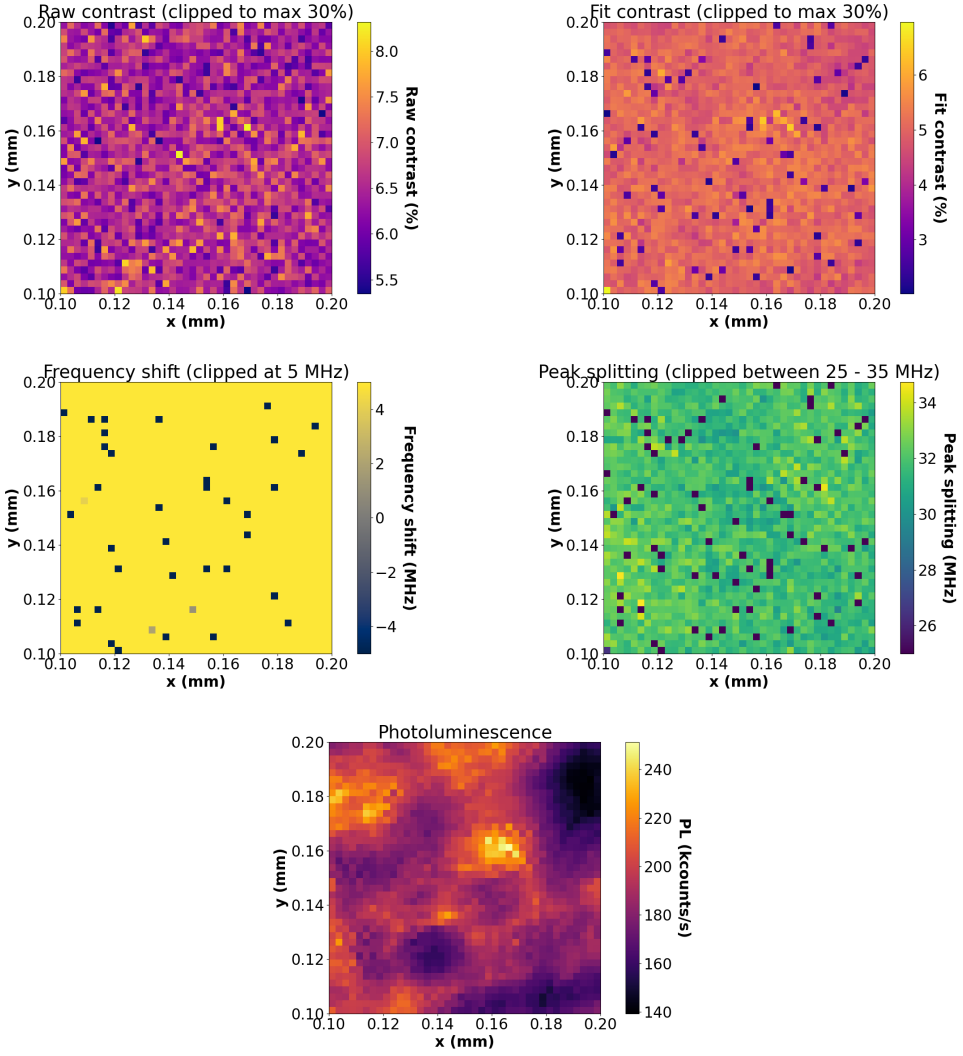


Figure C.5: With magnet, 50 micron inside the diamond

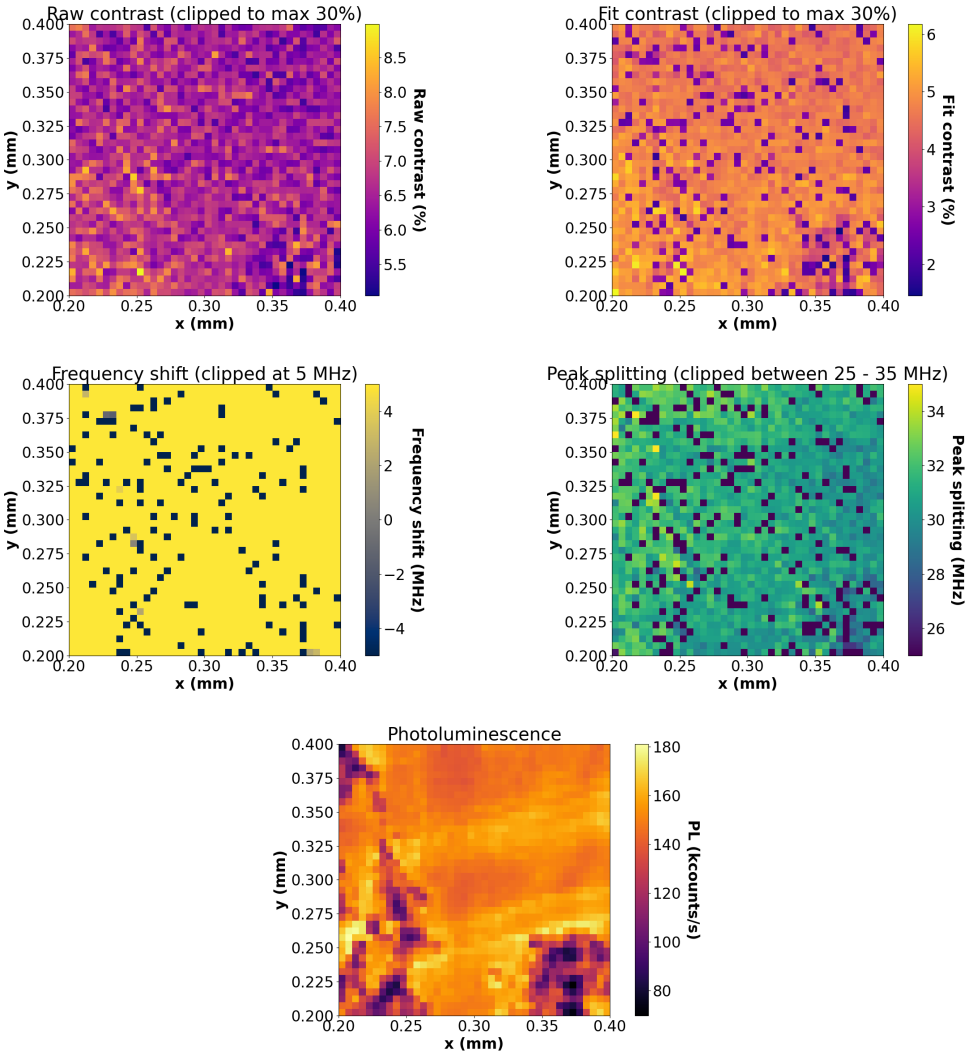


Figure C.6: With magnet, on the diamond surface, on a different and bigger 200x200 micron area.

D

Full list of CV measurements

For all twelve SPADs, CV measurements were taken at 5 different frequencies. In chapter 5 only three representative graphs were shown to illustrate the typical shapes. In this appendix chapter the full list of CV measurements is given. Pay attention to the unit! The scale can vary wildly. Also keep in mind that a current limit was set of 100 μA, which may distort the graphs at very low and very high voltages.

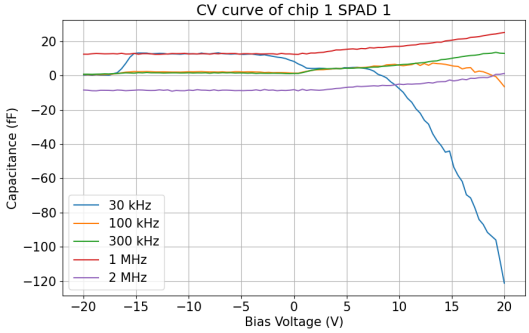


Figure D.1: CV curve of chip 1 SPAD 1

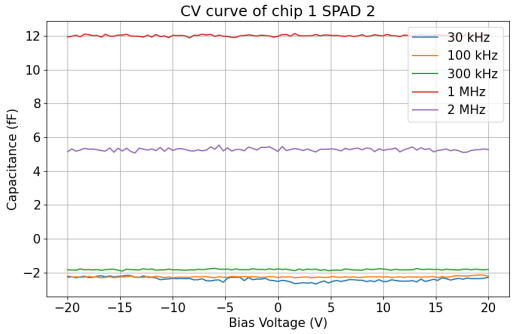


Figure D.2: CV curve of chip 1 SPAD 2

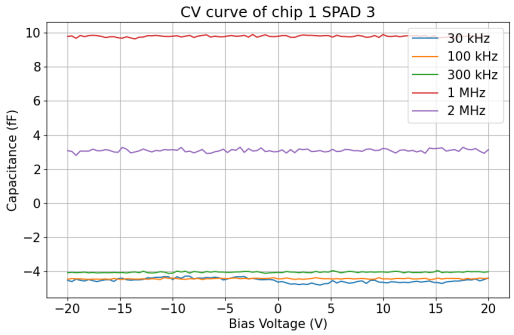


Figure D.3: CV curve of chip 1 SPAD 3

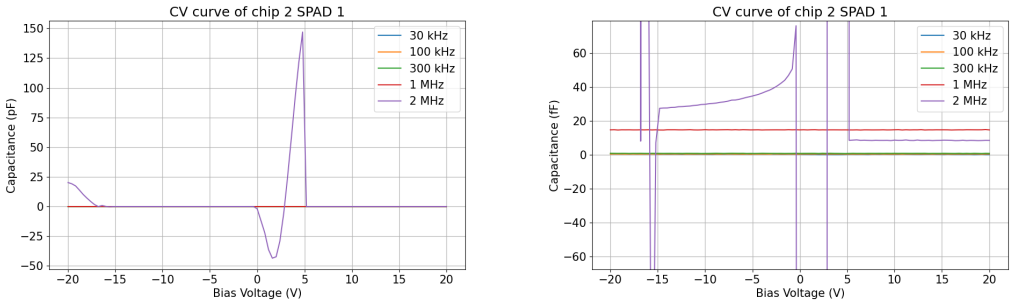


Figure D.4: CV curve of chip 2 SPAD 1. The right graph is a zoomed in version of the left graph.

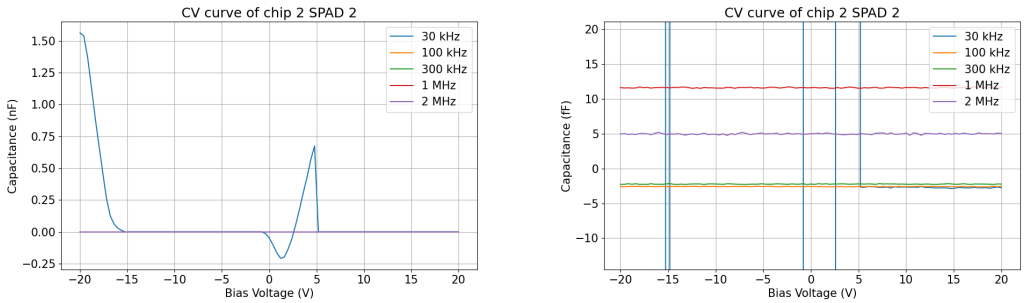


Figure D.5: CV curve of chip 2 SPAD 2. The right graph is a zoomed in version of the left graph.

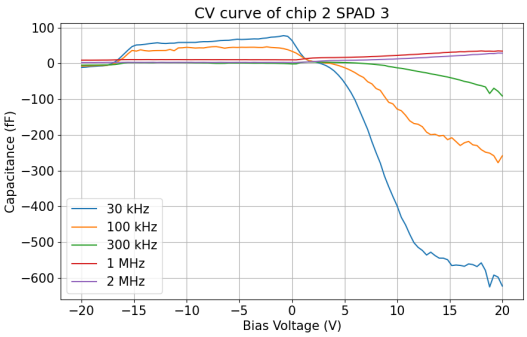


Figure D.6: CV curve of chip 2 SPAD 3

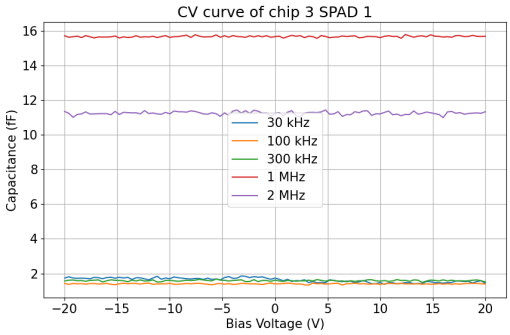


Figure D.7: CV curve of chip 3 SPAD 1

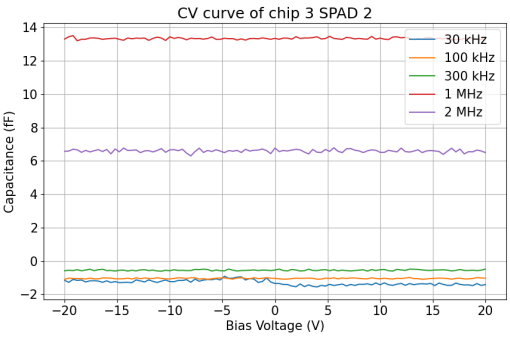


Figure D.8: CV curve of chip 3 SPAD 2

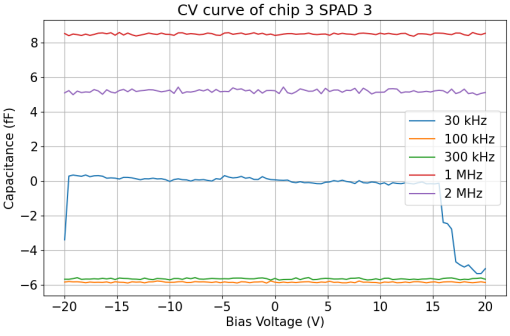


Figure D.9: CV curve of chip 3 SPAD 3

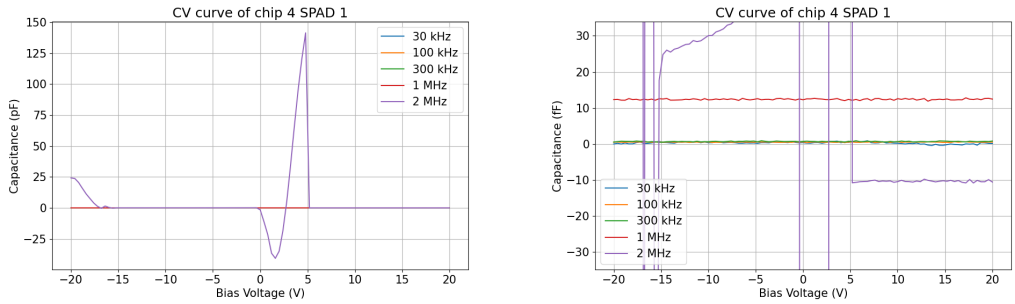


Figure D.10: CV curve of chip 4 SPAD 1. The right graph is a zoomed in version of the left graph.

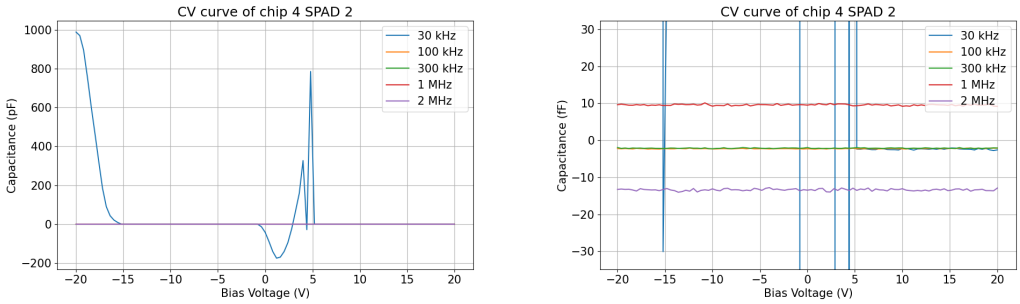


Figure D.11: CV curve of chip 4 SPAD 2. The right graph is a zoomed in version of the left graph.

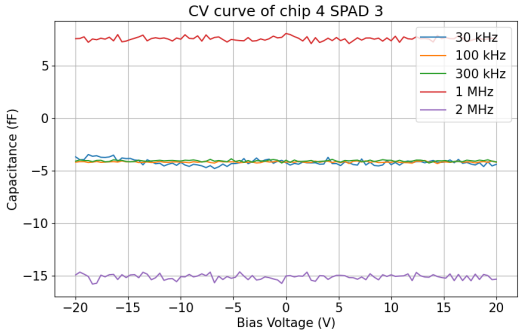


Figure D.12: CV curve of chip 4 SPAD 3

E

Extended results of the custom fitter

This appendix shows extended results on the new custom double-dip fitter that was made for processing 2D ODMR measurements. First of all, a copy of figure 4.3 is copied here for convenience, which is figure E.1. In that figure, a regular case is presented with two dips that are distinguishable by the initial guesser. However, when the dips are more closely together, the initial guesser might mistake them for one big dip. This is the case in figure E.2. In this case, the initial guesser tries to place both peaks in the middle with a splitting of 3 MHz, in the hope that the Adam optimizer will take care of the rest. And in fact, the Adam optimizer is consistently able to properly finalize the fit. Unfortunately, the fitter has one bad weakness. If the dips are asymmetrical, in the sense that one dip is shorter than the other, then the initial guesser consistently fails to find two dips. This is presumed to be the case for the failed fits during the bio-measurements.

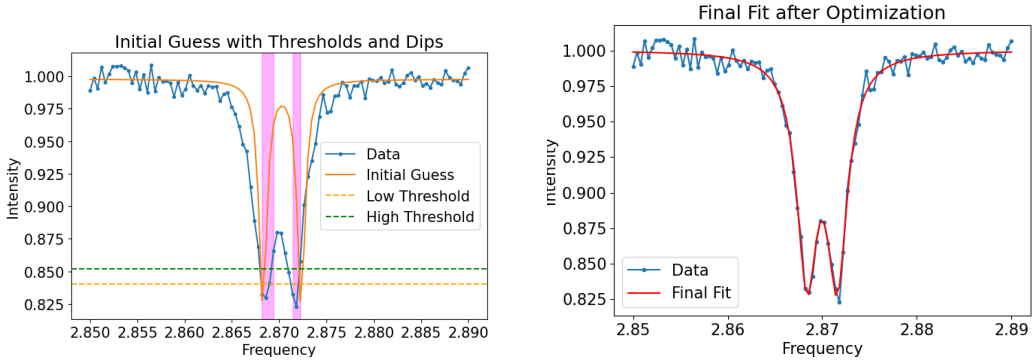


Figure E.1: Demonstration of the new fitter with simulated data.
Left The initial guess. A low and high threshold for dip detection are dynamically determined based on the standard deviation of the left and right tails of the signal. The thresholds are set at the lowest data point plus predetermined factors times the standard deviation. A dip is defined as a signal that passes both thresholds at the sides of the dip. This is indicated by the magenta rectangles. An initial guess for all parameters of the double Lorentzian dip are determined using the found regions. For more detailed information, refer to the main text above.
Right The final fit after going through the Adam optimizer. The fit was a success.

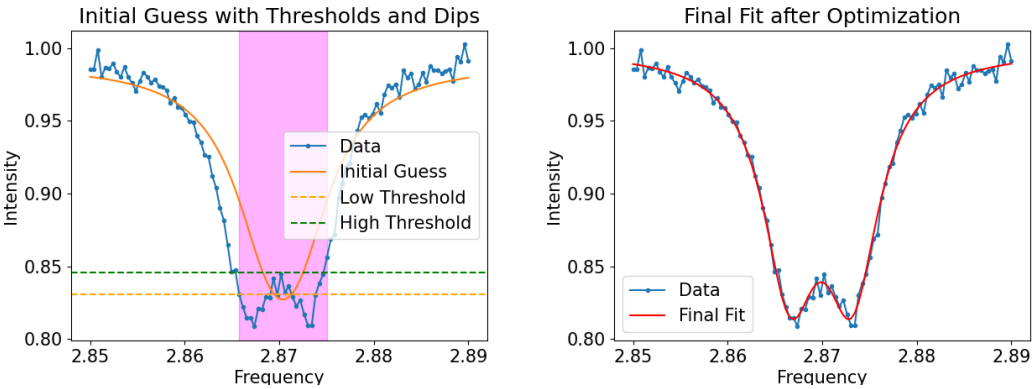


Figure E.2: The fitter at work with two very overlapping dips.
Left The initial guess. Only one dip region is found, and the initial guesser tries to guess something that the Adam optimizer can do something with.
Right The Adam optimizer was able to properly tune the parameters such that the very overlapping dips are well-fitted.

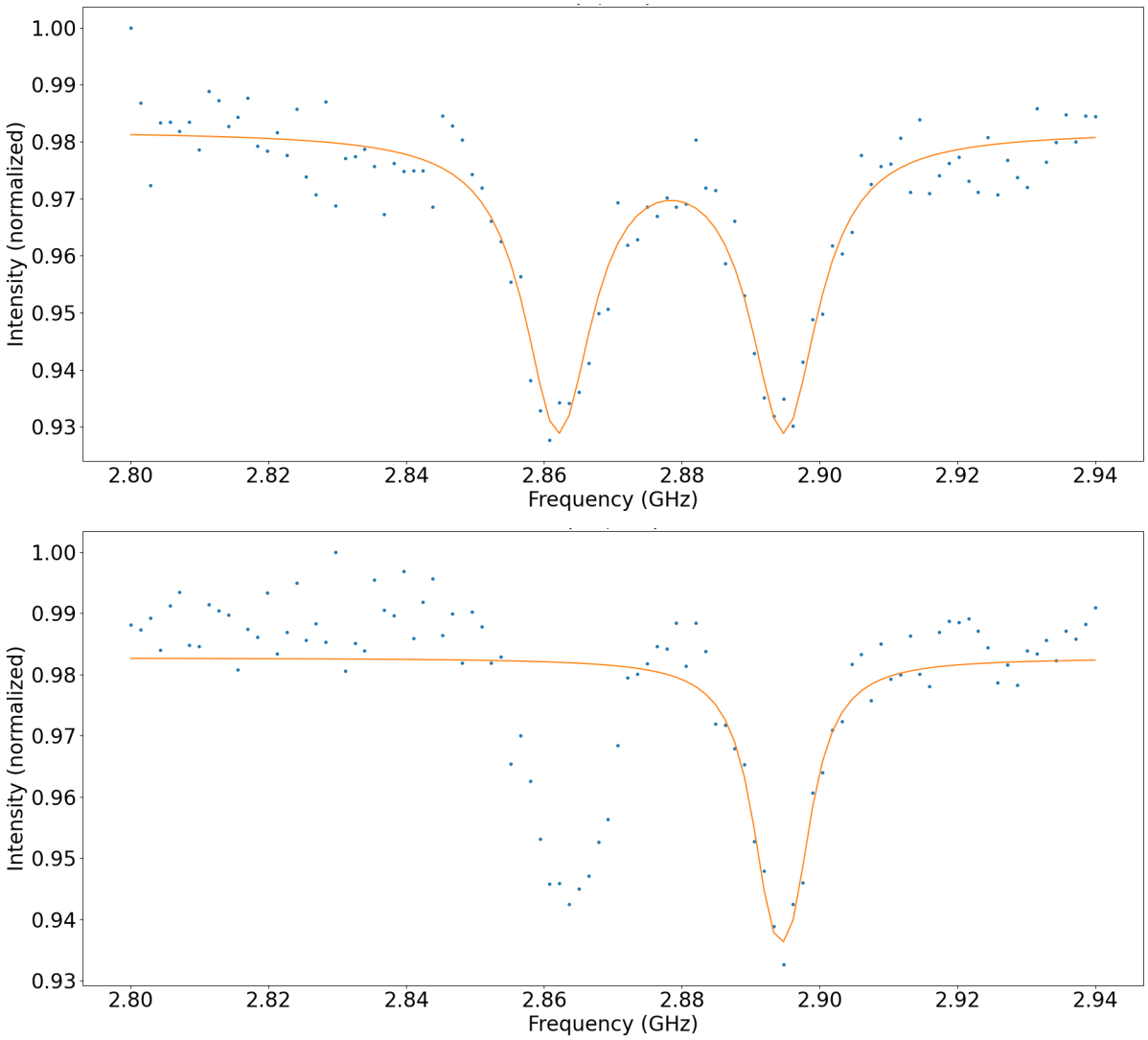


Figure E.3: The new fitter with two arbitrarily chosen pixels from the bio measurements, one with a successful fit and the other without success.
Top The fitter has a successful fit on the double dip.
Bottom The fitter struggles because the right dip has a very low pixel. Because of that, the initial guesser fails to recognize the left dip, and instead focuses both dips of the function in the right real dip. Because of this, the reported frequency shift will be high and the peak splitting near zero, which differ significantly from their real values. This happened in about 4.6% of the pixels during the bio-measurements.

F

Changes made after defending

Before uploading this thesis to the TU Delft repository, some corrections were made on some typos and/or missing and broken information and links. The lists below contain all the exact changes that were made.

F.1. Missing information

- Results: Changed two instances of ”sample [x]” to ”the blank biosample”.
- Acknowledgements: Added Sadik Ilik and Robert van Veldhoven from NXP for their support with electronics surrounding SPADs.
- Titlepage: Added Salahuddin Nur from the examiner committee (he was officially added after thesis was already handed in).

F.2. Spelling mistakes

- Acknowledgements: Fixed capital L in ”And Last but not least”
- Results: Fixed ”the a” in the caption of figure 5.13 (d).
- Results: Removed ”properly” from ”preventing them from getting quenched properly” to bypass an issue with image positioning breaking text.

F.3. Image edits

- Figure 3.2: Changed ”SPAD” to ”SPCM” to better represent the fact that it is the standalone single photon counting module instead of a general single photon avalanche diode.
- Figure 3.2: Added missing long pass filter and changed caption accordingly.
- Figure 3.7: Arrow indicating 40x GPIO connection is now double sided, representing the fact that signals between the FPGA and the PCB go both ways.

F.4. Fixed formerly broken cross references

- Figures 3.15a, 3.15b, 5.18, 5.5.
- Chapters 3, 4, A.

F.5. Additional remarks

The fitting routine has been slightly more finetuned to achieve better succesful fit rate. This includes a feature in the peak detection where peaks are assumed to not be within a specified amount of MHz from eachother. This greatly reduces the number of peaks that are misinterpreted as a double peak. The new version of the code is available on the Github (see appendix chapter A).

Bibliography

[1] URL: https://qdm.io/.

[2] John F. Barry et al. “Sensitivity optimization for NV-diamond magnetometry”. In: *Rev. Mod. Phys.* 92 (1 Mar. 2020), p. 015004. DOI: 10.1103/RevModPhys.92.015004. URL: https://link.aps.org/doi/10.1103/RevModPhys.92.015004.

[3] El-Mostafa Bourim and Jeong In Han. “Electrical Characterization and Thermal Admittance Spectroscopy Analysis of InGaN/GaN MQW Blue LED Structure”. In: *Electronic Materials Letters* 11 (Oct. 2015), p. 11. DOI: 10.1007/s13391-015-5180-0.

[4] Sander Britton, Martijn Janssen, and Eva Nootenboom. *Single-Photon Avalanche Diodes (SPAD) for Quantum Sensors*. Bsc thesis. 2023.

[5] Jakub Cabal. *spi-fpga*. 2021. URL: https://github.com/jakubcabal/spi-fpga.

[6] Wikipedia contributors. *Single Photon Avalanche Diode*. URL: https://en.wikipedia.org/wiki/Single-photon_avalanche_diode.

[7] M. W. Doherty et al. “Theory of the ground-state spin of the NV”. In: *Physical Review B* 85.20 (May 2012). ISSN: 1550-235X. DOI: 10.1103/physrevb.85.205203. URL: http://dx.doi.org/10.1103/PhysRevB.85.205203.

[8] A. Dréau et al. “Avoiding power broadening in optically detected magnetic resonance of single NV defects for enhanced dc magnetic field sensitivity”. In: *Physical Review B* 84.19 (Nov. 2011). ISSN: 1550-235X. DOI: 10.1103/physrevb.84.195204. URL: http://dx.doi.org/10.1103/PhysRevB.84.195204.

[9] Alireza Erfanian et al. “Reduction of afterpulse and dark count effects on SPAD detectors using processing methods”. In: *Results in Optics* 15 (2024), p. 100617. ISSN: 2666-9501. DOI: https://doi.org/10.1016/j.rio.2024.100617. URL: https://www.sciencedirect.com/science/article/pii/S2666950124000142.

[10] Matthew W. Fishburn. “Fundamentals of CMOS Single-Photon Avalanche Diodes”. PhD thesis. TU Delft, 2012.

[11] “Fundamentals of magnetic field measurement with NV centers in diamond”. In: *Qnami* (2020). URL: https://qnami.ch/wp-content/uploads/2020/12/2020-12-07-Qnami-TN1-The-NV-center-1.pdf.

[12] David R. Glenn et al. “Single-cell magnetic imaging using a quantum diamond microscope”. In: *Nature Methods* 12.8 (Aug. 2015), pp. 736–738. ISSN: 1548-7105. DOI: 10.1038/nmeth.3449. URL: https://doi.org/10.1038/nmeth.3449.

[13] Alexander D. Griffiths et al. “Temporal Encoding to Reject Background Signals in a Low Complexity, Photon Counting Communication Link”. In: *Materials* 11.9 (2018). ISSN: 1996-1944. DOI: 10.3390/ma11091671. URL: https://www.mdpi.com/1996-1944/11/9/1671.

[14] L. T. Hall et al. “High spatial and temporal resolution wide-field imaging of neuron activity using quantum NV-diamond”. In: *Scientific Reports* 2.1 (May 2012), p. 401. ISSN: 2045-2322. DOI: 10.1038/srep00401. URL: https://doi.org/10.1038/srep00401.

[15] Shenjie Huang et al. “Optimal Photon Counting Receiver for Sub-Dead-Time Signal Transmission”. In: *Journal of Lightwave Technology* 38.18 (2020), pp. 5225–5235. DOI: 10.1109/JLT.2020.3000723.

[16] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG]. URL: https://arxiv.org/abs/1412.6980.

[17] S. S. Kohneh Poushi et al. “Comprehensive Modeling of Photon Detection Probability in CMOS-based SPADs”. In: *2020 IEEE SENSORS*. 2020, pp. 1–4. DOI: 10.1109/SENSORS47125.2020.9278771.

[18] Edlyn Levine et al. “Principles and Techniques of the Quantum Diamond Microscope”. In: (Sept. 2019).

[19] S. Pellegrini et al. “Industrialised SPAD in 40 nm technology”. In: *2017 IEEE International Electron Devices Meeting (IEDM)*. 2017, pp. 16.5.1–16.5.4. DOI: 10.1109/IEDM.2017.8268404.

[20] Matteo Perenzoni, Lucio Pancheri, and David Stoppa. “Compact SPAD-Based Pixel Architectures for Time-Resolved Image Sensors”. In: *Sensors* 16.5 (2016). ISSN: 1424-8220. DOI: 10.3390/s16050745. URL: https://www.mdpi.com/1424-8220/16/5/745.

[21] Jens Pogorzelski et al. “Compact and Fully Integrated LED Quantum Sensor Based on NV Centers in Diamond”. In: *Sensors* 24.3 (2024). ISSN: 1424-8220. DOI: 10.3390/s24030743. URL: https://www.mdpi.com/1424-8220/24/3/743.

[22] Samantha van Rijs, Lars Visser, and Dagmar Westenbrink. *On-chip SPAD array with IO designed for quantum sensing*. Bsc thesis. 2023.

[23] Alfred Schuurmans. *Strain Mapping Using Continuous Wave Optically Detected Magnetic Resonance on an NV-centre Ensemble*. Bsc thesis. 2024.

[24] Mario Stipčević, Hrvoje Skenderovic, and Davor Gracin. “Characterization of A Novel Avalanche Photodiode for Single Photon Detection in VIS-NIR Range”. In: *Optics express* 18 (Aug. 2010), pp. 17448–59. DOI: 10.1364/OE.18.017448.

[25] Hao Wei et al. “Superparamagnetic Iron Oxide Nanoparticles: Cytotoxicity, Metabolism, and Cellular Behavior in Biomedicine Applications”. en. In: *Int J Nanomedicine* 16 (Aug. 2021), pp. 6097–6113.

[26] Yux Xu, Ping Xiang, and Xiaopeng Xie. “Comprehensive understanding of dark count mechanisms of single-photon avalanche diodes fabricated in deep sub-micron CMOS technologies”. In: *Solid-State Electronics* 129 (2017), pp. 168–174. ISSN: 0038-1101. DOI: https://doi.org/10.1016/j.sse.2016.11.009. URL: https://www.sciencedirect.com/science/article/pii/S0038110116302763.

[27] Jiyuan Zheng et al. “Dynamic-quenching of a single-photon avalanche photodetector using an adaptive resistive switch”. In: *Nature Communications* 13 (Mar. 2022), p. 1517. DOI: 10.1038/s41467-022-29195-7.