POLITECNICO DI MILANO

Scuola di Ingegneria Industriale e dell'Informazione

Corso di Laurea Magistrale in Ingegneria Elettronica

Development of a New Generation of Portable Optical Reader for Biomolecular Analysis

Relatore: Prof. Roberto OSELLAME
Correlatore: Ing. Davide CUCCHI

Tesi di Laurea Magistrale di
Giorgia CALISTI
Matr. N. 798581

Anno Accademico 2016-2017

Contents

List of Figures

List of Tables

Abstract

The real-time Polymerase Chain Reaction (PCR) technique is the most common DNA amplification method up to date, and it finds applications in a number of different fields such as DNA sequencing, diagnosis of infectious diseases, forensics sciences, food safety, and more. Recent trends in healthcare needs and disease prevention have seen the development of Point of Care (POC) techniques, which aim to bring the equivalent of laboratory analyses out of medical facilities. This has generated an increasing need for easy-to-use, portable devices able to perform biomedical diagnostic testing such as real-time PCR.

In this thesis I describe my contribution to the characterization and evolution of a real-time PCR POC handheld device developed by STMicroelectronics. The activities related to the project were carried out over a year-long period working with Marco Bianchessi's R&D group at STMicroelectronics laboratories, in Agrate Brianza, and Roberto Osellame's research group at the Department of Physics of Politecnico di Milano.

The device is composed of two subsystems: one executes thermal protocols and the other is deputed to optical behavior control, namely the selection of excitation light and the detection of the emissions of fluorescent dyes during the PCR reaction. My work has been focused on the characterization of the second subsystem. Its main components are two sets of filters, chosen to select excitation and emission wavelength for four operating windows, four couples of LEDs providing excitation light and four CMOS camera modules for fluorescence detection.

The first part of my work was focused on the device characterization. This was performed by means of two automated acquisition setups designed and assembled for this purpose. The first features a motorized optical switch that allowed a thorough spectral analysis of the light on the PCR reaction area. The second allows the characterization of the transmission filters as a function of the angle of incidence of impinging light.

The need to further expand the real-time PCR device capabilities also led to the design of a new system, which integrates the optimization results derived from my previous work while adding new features. In particular, the optical section has been rethought to work with a single, larger HDR sensor and to extend the detection to a fifth channel in the near-IR region. Since the old system had sensors and lenses integrated in a single camera

module, the introduction of a new standalone sensor brought with it the need to design custom optics.

The last part of my work was therefore dedicated to the design and fabrication of a refractive plano-convex lens that could be suitable for the new imaging system. The lens was fabricated by means of two-photon polymerization (2PP) prototyping technique at the DiamANT laboratory at the Department of Physics of Politecnico di Milano. 2PP is a powerful technique for three-dimensional micro- and nano-fabrications, which allows direct laser writing of complex structures inside a photosensitive resin. The fabrication has been preceded by an objective characterization process. Then, starting from a 3D model of the lens, various design strategies were devised and implemented to find the best approach for the lens fabrication.

So try and get better and don't ever accept less.
Take a plain black marker and write this on your chest:
Draw a line underneath all of this unhappiness,
Come on now, let's fix this mess.

We can get better,
Because we're not dead yet.

Frank Turner

Chapter 1

Real-time PCR in molecular biology

In this chapter I will introduce the essential concepts surrounding DNA amplification, starting from the basic knowledge of DNA structure and developing a detailed description of the amplification procedures and their evolution in the last decades. The discussion is focused around the Polymerase Chain Reaction (PCR) technique, which is the most common DNA amplification method up to date.

The continuously growing number of applications based on real-time PCR and the need for easier and cheaper access to this technology are pushing industry and research towards more portable and simpler solutions than the traditional laboratory machines. Therefore, the end of the chapter is spent explaining this concept and introducing the real-time PCR device developed at ST Microelectronics, which has been the subject of my work.

1.1 The Polymerase Chain Reaction (PCR)

Molecular biology is a branch of science that investigates interactions between the molecular components that carry out the various biological processes in living cells. It links the knowledge coming from biochemistry (that covers the vital processes occurring in live organisms) with the research in the field of genetics. Therefore, modern molecular biology emerged with the uncovering of the double-helix structure of DNA in the 1960s.

The DNA amplification process is a fundamental technique used in clinical and research laboratories for a wide range of applications in molecular biology, in areas as diverse as sequencing, diagnosis and monitoring of genetic diseases, identifications of genetic fingerprints in the field of forensics, food science, ecological field studies and diagnostic medicine. The Polymerase Chain Reaction (PCR) is the in-vitro DNA (deoxyribonucleic

acid) amplification technique that revolutionized molecular biology since its introduction in the early '80s. It was devised in 1983 by Kary Mullis at Cetus Corporation (a biotechnology firm in Emeryville, CA). Ten years later he was also awarded the Nobel Prize in Chemistry for his contributions, although the successful application of the technique as it is widely known today needed considerable additional efforts from Mullis and his colleagues at Cetus. Notable milestones were the discovery and cloning of a stable DNA polymerase and the introduction of thermocycling machines [4], which shaped the modern PCR.

The goal of any amplification technique is to provide enough genetic material of a specific kind, after it has been properly selected and isolated, to allow the performance of reliable study and analysis of a given sample. Prior to the development of PCR the primary method used for DNA replication, known as *DNA cloning*, involved insertion of the sequence of interest into living bacterial cells, which would replicate it along with their own DNA as they reproduced. This process was relatively time-consuming and required intensive labor. The introduction of PCR brought the typical duration of the DNA amplification process from days to a matter of hours. Nowadays, it can take about 30 minutes to complete a full PCR.

1.1.1 DNA basics

The basic PCR process is driven by a thermostable polymerase enzyme, which will synthesize a complementary sequence of bases to any single strand of DNA providing it has a double stranded starting point. To understand this, one has to take into consideration the double-helix structure of DNA and its characteristics.

The deoxyribonucleic acid (DNA) is a nucleic acid that is responsible for storing and transmitting genetic information. A DNA molecule consists of two long polynucleotide chains composed of four types of simpler monomer subunits called *nucleotides* (Fig. 1.1). Each of these chains is known as a *DNA chain*, or a *DNA single strand* [5]. A nucleotide is composed of three parts: a pentose (five-carbon) sugar called deoxyribose (hence the name deoxyribonucleic acid), a phosphate group and one of four nitrogen-containing bases (either *cytosine (C)*, *guanine (G)*, *adenine (A)*, or *thymine (T)*) (Fig. 1.2). The nucleotides are linked together in each single strand by covalent bonds that bridge the sugar of one nucleotide to the phosphate of the next, thus forming a strong backbone. The double-helix structure is then formed by weaker hydrogen bonds between the base portions of the nucleotides, according to the base pairing rules (A with T, and C with G). This implies that the two strands carry the same biological information, which can be duplicated when the strands separate during DNA replication.

As can be noticed from Fig. 1.2a, the carbon atoms of deoxyribose are numbered:

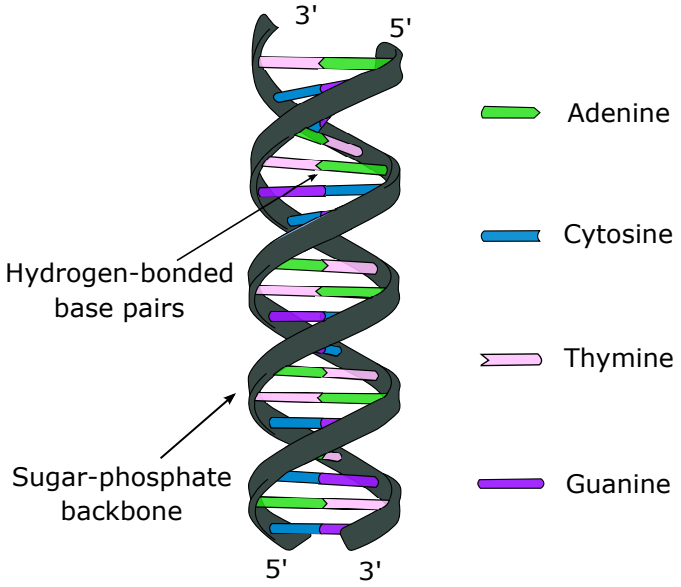


Figure 1.1: DNA and its building blocks.

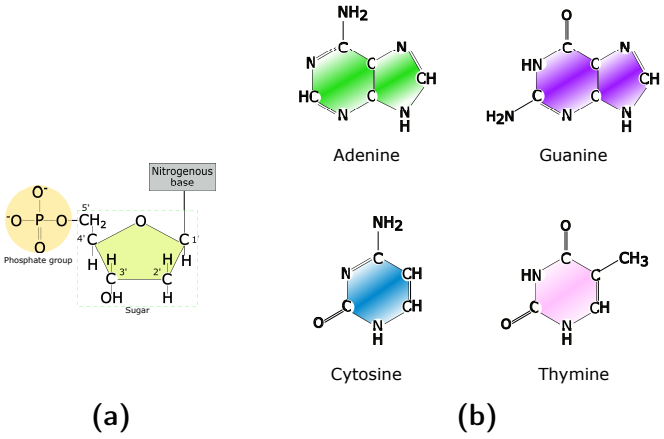


Figure 1.2: Deoxyribonucleotide (a) and bases (b) structures. Note how the carbon atoms are numbered along the sugar chain.

the bond with the phosphate group always happens between the third and fifth carbon atoms, and this asymmetry gives DNA a direction. Nucleotides in the double helix have opposite directions, so that the strands are anti-parallel and thus are said to have a directionality of *five prime* (5') and *three prime* (3'), with the 5' end having a negatively charged terminal phosphate group and the 3' end a terminal hydroxyl group (Fig. 1.3). The two complementary sequences are referred to as *sense* and *anti-sense*.

Living organisms need to perform cell division to grow, and in doing so the DNA in their genome must be replicated to perpetuate the genetic information. A key role here is played by an enzyme called *DNA polymerase*, which can build the complementary DNA strand of one of the two parts of an helix. The enzyme chooses the correct base to pair to those composing the chain and bonds it onto the original strand, while moving in a $5'$ to

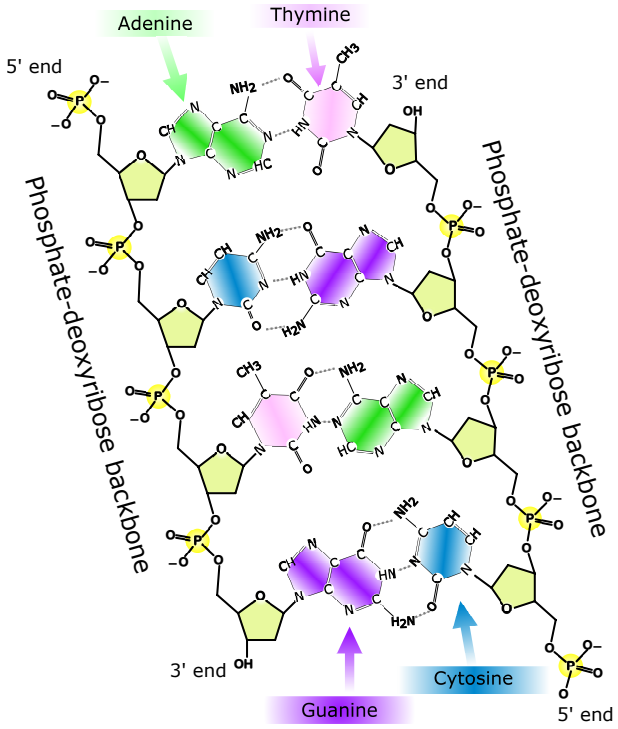


Figure 1.3: Each strand of DNA that composes an helix has a terminal phosphate on the 5' carbon atom and a terminal hydroxyl group on the 3'.

3′ direction. Fig. 1.4 schematically describes this process. Note that the bases on the old strand dictate which ones will appear in the new one, therefore the polymerase produces (in principle at least) a perfect copy of the original DNA.

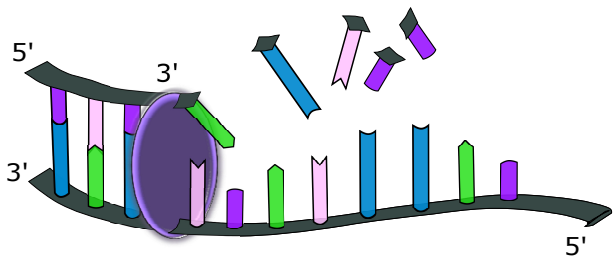


Figure 1.4: The DNA polymerase can rebuild the complementary DNA strand to a single chain, adding the correct bases in a 5' to 3' direction.

1.1.2 Working principle and reaction components

The polymerase chain reaction exploits the DNA replication mechanism to synthesize a great number of copies of a target sequence from a relatively small quantity of biologic material. The DNA first undergoes the process of *denaturation*, which separates the two strands, then, assuming that the target sequence is present in the sample, some small

pieces of DNA called *primers* bind to the longer chains (*annealing* process). The primers are particular sequences chosen to be complementary to the 3' ends of the target strands (both sense and anti-sense) to uniquely selected them for replication. This provides the double strand starting point that the polymerase needs to bind and begin the copying process, called *extension*. This cycle is repeated multiple times, and since the new strands are separated again at the beginning of each subsequent cycle to be used as templates, a chain reaction ensues that exponentially amplifies the original DNA.

For the PCR process to work several key chemical components need to be available:

- target DNA;
- polymerase;
- forward primer;
- reverse primer;
- deoxynucleotide triphosphates (dNTPs, the bases);
- magnesium chloride ($MgCl_2$);
- PCR buffer.

Furthermore, the process needs water as the liquid environment for the reaction to take place: it is the matrix in which the components interact.

The primary purpose of the PCR buffer is to provide an optimal pH ($\sim 8.3 - 8.8$ pH units at room temperature, ~ 7.2 pH units during extension phase) and monovalent salt environment for the final reaction volume. Sometimes it already contains the magnesium chloride, which supplies free Mg^{++} divalent cations required for the polymerase binding activity.

Back in the earliest days of PCR, the amplifications were carried out in laboratories with manual transfers between hot water baths, and the best DNA polymerase available was extracted from the bacterium *Escherichia coli*. This enzyme has the great disadvantage that it becomes irreversibly deactivated at the temperatures of the DNA denaturation step (around 95 °C), therefore fresh aliquots of it had to be added by hand at the start of each cycle. The solution to this problem was initially found when a thermophilic bacterium called *Thermus aquaticus* was isolated by Thomas D. Brock and Hudson Freeze in 1969 near the Lower Geyser Basin in Yellowstone National Park, where it survived and proliferated at extremely high temperatures. A thermophilic organism lives all or part of its life at temperatures between 41 °C and 122 °C, therefore the DNA polymerase extracted from this bacterium is perfectly stable in the required range. This was achieved by Chien

et al. in 1976 [6], when the Thermus aquaticus was actually purified and characterized for the first time.

Following the nomenclature rule for enzymes derived from microorganisms, the DNA polymerase from ***T**hermus **aq**uaticus* is called Taq. It has become the standard choice for PCR, although other thermal-stable DNA polymerases have been isolated from other thermophilic species (e.g. Tth polymerase, Pfu polymerase) and can be preferred depending on the application at hand. The three main aspects that should be considered in this regard are *processivity*, *fidelity* and *persistence*. Processivity refers to the rate at which an enzyme makes the complementary copy of a template, fidelity addresses the accuracy of the copy and persistence describes the stability of the enzyme at high temperatures. Taq polymerase has a processivity of $50 - 60$ nucleotides per second at 72 °C, which is roughly double of most of the other alternatives, it has one of the worse error rates at $285 \cdot 10^{-6}$ errors per template nucleotide and a half-life of about one hour and a half when exposed to a constant 95 °C. For the vast majority of PCR applications the average size of the PCR target sequence (which is referred to as *amplicon*) is less than 500 base pairs, which makes the Taq polymerase error negligible and therefore preferable for its ideal processivity. For applications such as long PCR, with much wider amplicons, fidelity can become more relevant than processivity, thus leading to the choice of other enzymes.

The introduction of Taq polymerase allowed the performance of complete PCR amplifications without opening the reaction chambers, which led to the design and engineering of automated thermocycling machines [7].

If all the reaction components discussed above can be assumed to be optimal, the success of a PCR will ultimately depend in large part upon the design of the specific primers. Since their purpose is to select a unique address in the target DNA, the first aspect to consider is the length of the primers, usually expressed in units of *base pairs* (*bp*). The length can be estimated taking into account the nucleotides probability of occurrence and the size of the genome of interest: the result leads however to general considerations. Let us assume, for the sake of simplicity, that the four nucleotides have the same probability of occurrence at a specified position, that the nucleotide in each subsequent position is independent of the nucleotides in all previous position, and that we are considering a target DNA in the human genome ($L_H = 3.3 \cdot 10^9$ *bp*). Therefore $p(A) = p(T) = p(C) = p(G) = 0.25$ and, since the events are independent by assumption, the probability of any sequence of length n (called *oligomer* or, more specifically, *oligonucleotide*) is

$$p_n = (0.25)^n. \tag{1.1}$$

The number of occurrences of a given oligomer by chance alone is thus

$$n_{oc} = p_n \cdot L_H, \tag{1.2}$$

by requiring that this result equals less than one we can estimate a suitable primer length. This procedure can be adapted to any DNA source specifying the size of the genome

n	$(0.25)^n$	Random occurrences
1	0.25	$825 \cdot 10^6$
2	$62.5 \cdot 10^{-3}$	$206.25 \cdot 10^6$
3	$15.63 \cdot 10^{-3}$	$51.48 \cdot 10^6$
	...	
10	$953.67 \cdot 10^{-9}$	$3.135 \cdot 10^3$
11	$238.42 \cdot 10^{-9}$	792
	...	
14	$3.73 \cdot 10^{-9}$	12
15	$931.31 \cdot 10^{-12}$	3
16	$232.83 \cdot 10^{-12}$	0.77
17	$58.21 \cdot 10^{-12}$	0.19
20	$909.5 \cdot 10^{-15}$	$3 \cdot 10^{-3}$
25	$888.18 \cdot 10^{-18}$	$3 \cdot 10^{-6}$

Table 1.1: Simplified random nucleotide sequence probability and occurrences for the human genome as a function of the number of bases.

of interest and the relative proportions of the bases occurrences, but it can be noticed looking at Table 1.1 that a primer with at least 20 nucleotides has a good chance of being unique. Keeping in mind that a PCR amplification requires two such primers (one in each direction) that lie a specified distance - in terms of base pairs - apart, it is clear that the combination of length and proximity can guarantee a unique address in the target DNA.

The choice of the actual sequence for the primers is a somewhat more complex topic, that requires addressing stability issues from a chemical point of view and taking into account the likelihood of self- and cross-complementarity of the primers themselves.

1.1.3 The basic PCR cycle

As described in Par. 1.1.2, PCR consists of the subsequent repetition of cycles of amplification that can be divided in three phases, each characterized by a specific temperature and a time duration.

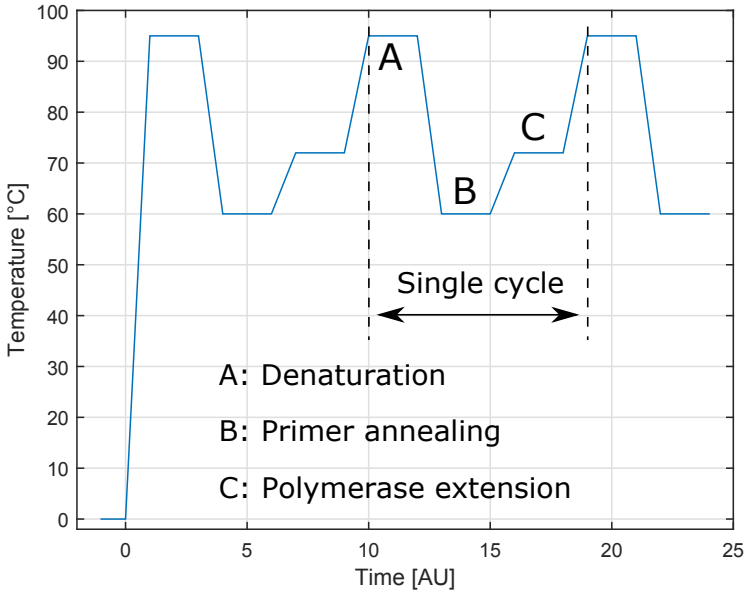


Figure 1.5: PCR temperature curve. Each cycle is composed of three phases carried out at different temperatures.

During the *denaturation step*, labeled in Fig. 1.5 with the letter A, the reaction chamber is heated to 94 °C - 95 °C for 15 s to 120 s. The hydrogen link between the two intertwined strands of the target DNA will at this point break (or "melt"), effectively separating them while keeping the rest of the structure intact. The precise temperature for the denaturation to happen depends partly on the proportion of guanine-cytosine pairs. In fact, G+C pairs are linked by three hydrogen bonds versus two for the A+T couple, and therefore their denaturation requires higher amounts of energy. The target length also plays a relevant role, since longer chains need also an higher temperature to completely separate. If the chosen value is too low or if the time is too short, A+T rich regions of the template DNA will be denatured while others will not, and when the temperature is reduced later in the PCR cycle the template DNA will reanneal into fully native condition. The amount of strand separation is measured exploiting an effect called *hyperchromicity*, which is the increase of absorbance of a material. In their native state, nucleic acids absorb light at $\lambda_{abs} = 260\text{nm}$ because of the electronic structure in their bases. Duplex DNA absorbs less ultraviolet light by about 40% than do single DNA chains, because the close proximity of two bases quenches the single-strand absorbance. If we plot the absorbance of DNA as a function of temperature, we observe that the increase in absorption occurs abruptly over a relatively narrow (about 5 °C) temperature range (Fig. 1.6). The midpoint of this transition is the *melting temperature* T_m, which is defined as the temperature at which 50% of the molecules of the target DNA sequence are hybridized into a double strand and

50% are present as a single strand.

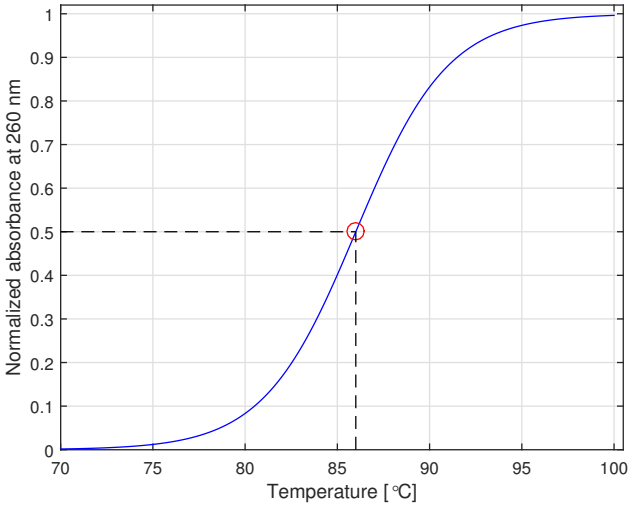


Figure 1.6: Typical thermal denaturation (melting) curve for DNA, showing the variation in optical absorbance at 260 nm with increasing temperature. T_m is defined as the temperature corresponding to half the overall change in absorbance.

It is worth noting that in the first cycle the heat is setting up convection currents in the reaction mix that will start all of the molecules in Brownian motion, which will persist with highs and lows during all of the subsequent temperature changes. The reaction components will thus remain in a constant state of mixing.

When the denaturation process terminates, the temperature in the reaction chamber will be lowered for 15 s - 60 s toward the *annealing* temperature T_a, which is set 3 °C - 5 °C below T_m of the primers. The choice of T_a, which can be of the order of 40 °C - 60 °C, is critical, because it must be low enough to allow for hybridization of the primers to the strand but high enough for the hybridization to be specific. As temperature lowers and settles at T_a, the primers, which are present in concentration much higher than the target DNA, will try and discard random duplexes on the available single strands until they find their perfect complement and begin to anneal, that is, form stable hydrogen bonds. Determining the annealing temperature is critical: if it is too low, non specific annealing of primers may occur, resulting in the amplification of unwanted segments of DNA; if too high, they might not bind at all.

Extension of oligonucleotide primers is carried out at or near the optimal temperature for DNA synthesis catalyzed by the thermostable polymerase, which in the case of Taq polymerase is 72 °C - 80 °C [6]. In this phase the DNA polymerase, exploiting the starting point offered by the freshly hybridized primers, synthesizes new strands by adding the complementary bases to the target sequence in 5' to 3' direction. The next cycle begins

with a return to 94 °C for denaturation. As a rule of thumb, extension is carried out for one minute for every 1000 *bp* of amplified product.

In the first cycle, extension from one primer proceeds beyond the sequence complementary to the binding site of the other primer, thus producing a long chain of indeterminate length from each template strand. In the next cycle, the templates produce long chains again, while the ones that were created in the previous step produce the first sequences whose length is exactly equal to the segment of DNA delimited by the binding sites of the primers (that is, the target sequence). From the third cycle onwards, this segment of DNA is amplified geometrically, whereas longer amplification products accumulate arithmetically (see Fig. 1.7). Assuming to have just one starting DNA helix, after N cycles the final product will contain the two original chains (the template single strands), $2N$ long chains and $2^{N+1} - 2(N+1)$ short segments replicating one side of the exact target sequence, that is, $2^N - (N+1)$ duplexes with both ends defined by a primer[1]. The latter is the

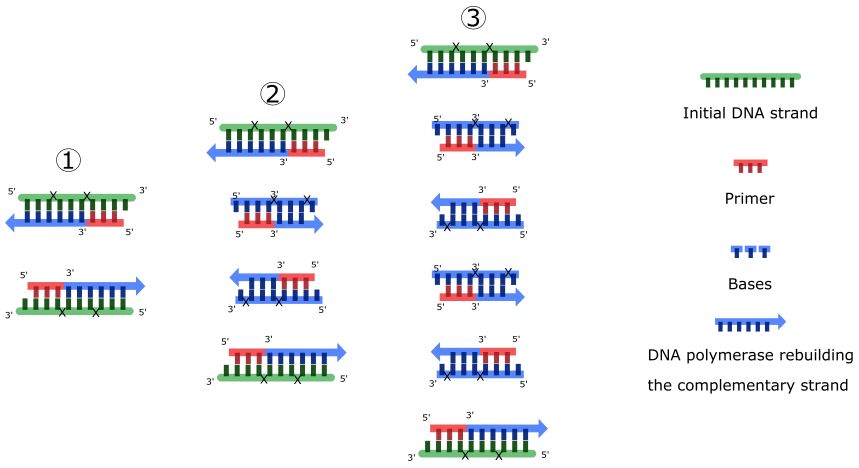


Figure 1.7: Products of the first three cycles of a PCR. Black crosses mark the target sequence that is being amplified.

desired product. So each PCR cycle theoretically almost doubles the amount of amplicon in the reaction: ten cycles multiply it by a factor of about one thousand, twenty cycles by a factor of more than a million. If the initial target concentration is X_0, the total concentration of desired duplexes after N cycles can be expressed as

$$X_n = X_0[2^N - (2N+2)] \simeq X_0 \cdot 2^N \tag{1.3}$$

This theoretical amplification value is never achieved in real applications, because of

[1]To be precise, at the end of the reaction it will be much more probable to find the template and long fragments hybridized with the short sequences than with one another, therefore an extra $-(N+1)$ factor should be added to the final equation to account for this. The total number of desired duplexes becomes $2^N - 2(N+1)$.

several factors that occur in practice. Those include for example the competition between new complementary strands and primers for annealing, where no amplification results from two already existing daughter segments. Also, while the concentration of the polymerase enzyme does not change, its performance tends to worsen as a result of the repeated temperature changes. Meanwhile, more primers and bases are incorporated in the new DNA during each cycle, therefore their availability decreases over time. Because of this, once established in the geometric phase, the reaction proceeds until one of its components becomes limiting. This is generally the case after roughly 40 cycles in PCRs containing about one hundred copies of the target sequence. To account for the consumption of the reaction components and the other factors, the final DNA concentration can be expressed more accurately introducing in Eq. 1.3 the amplification rate r

$$X_n = X_0 \cdot r^N \tag{1.4}$$

with $1 < r < 2$.

1.1.4 Detection of PCR amplification

The overall growth of the amplified DNA concentration follows a sigmoidal shape (see Par. 1.2), that changes the initial exponential trend into a linear function which then saturates to a constant value, referred to as the reaction *plateau*. The plateau phase is where quantitative results are drawn in a traditional PCR. This is known as *end-point detection*. *Agarose gel electrophoresis* is the most common method of analysis, its purpose is that of separating the DNA fragments contained in the sample by length, thus only giving a positive or negative result regarding the presence of the target sequence. Since the analysis is carried out only after saturation of the PCR curve, no meaningful quantitative conclusions about the initial concentration of the target DNA can be drawn.

This detection technique enhances the sensitivity of the traditional electrophoresis procedure to better discern between sequences of similar length. To separate charged molecules according to their size with electrophoresis, they are suspended in a fluid where a spatially uniform electric field $\mathcal{E}$ is applied by means of two submerged electrodes. Each particle will then start to migrate towards the electrode of the opposite sign, and after a brief transient they will reach a steady state where the electrokinetic force $F_{ek} = q\mathcal{E}$ exerted by the electric field is balanced by the drag force $F_{drag} = 6\pi\eta r v$ due to the viscosity of the fluid, that is

$$F_{ek} + F_{drag} = 0. \tag{1.5}$$

Here, η is the viscosity of the fluid, r is the radius of the particle and v is the constant

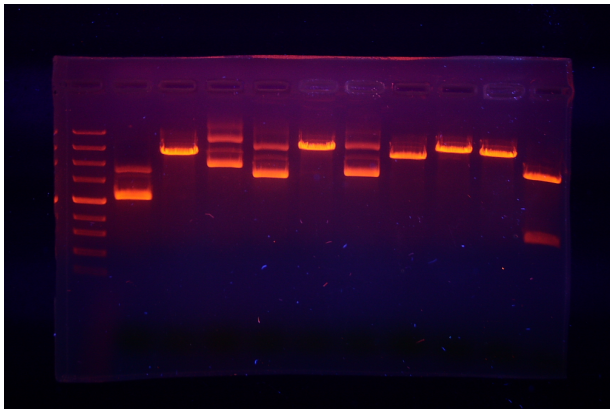


Figure 1.8: DNA fragments in agarose gel stained with ethidium bromide. DNA and ethidium bromide form a complex that emits orange light if placed under UV. (c) Rainis Venta. Used under the Creative Commons Attribution-Share Alike 3.0 Unported license.

migration velocity due to the force balance. Solving for the latter variable, we obtain

$$v = \frac{q}{6\pi r\eta}\mathcal{E} = \mu\mathcal{E}. \tag{1.6}$$

The term $\mu = \frac{q}{6\pi r\eta}$ is called *electrophoretic mobility*, and it shows that particles with smaller sizes will travel faster than the larger ones. Therefore, by employing an optical detection system at the end of the run, different particles can be classified by their arrival time.

Because of their sugar-phosphate backbone (see Par. 1.1.1), nucleic acids carry negative charge that induces them to migrate towards the positive electrode in an electrophoretic run. The separation of the DNA fragments resulting from a PCR is therefore accomplished by exploiting the same principle just described for particles suspended in a fluid. In this case, though, the PCR product is set on a viscous medium, the agarose gel, which is a three-dimensional matrix composed of helical molecules that form channels and pores which are designed to slow down the DNA strands. Their migration behavior through the pores of the gel is linked to their size, which affects its mobility in the same way as described by Eq. 1.6 for generic particles, therefore smaller fragments end up closer to the positive end of the gel slab than longer ones in a given amount if time. The size of the gel pores can also be controlled by simply adjusting the agarose concentration, to prepare gels appropriate for the separation of a wide range of different-size nucleic acid molecules.

The DNA fragments of different lengths are visualized adding specific fluorescent intercalating dyes (Fig. 1.8), such as ethidium bromide, which is strongly mutagenic, or less-toxic proprietary dyes such as GelRed, GelGreen, and SYBR Safe. Intercalators are molecules that fit themselves between DNA bases, therefore when after electrophoresis the

gel is illuminated with UV light fluorescent bands point out the presence of DNA. Ideally, electrophoresis should yield a single strong band of the correct size, which can be checked against reference size markers (*DNA ladders*) run on the same gel.

1.1.5 Drawbacks of traditional PCR

While the introduction of PCR in molecular biology has had a major impact in a large number of applications, it is also true that some of its inherent disadvantages pushed towards the development of new forms of this technique that could better respond to specific scientific and technical needs.

One of the most relevant issues is the lack of quantitative information about the starting concentration of the sample, which simply cannot be obtained because of the end-point detection approach, as described in Par. 1.1.4. Since the result of the analysis is read from the last step of the reaction, when saturation has been reached, all the information on the initial conditions has been lost. Closely related to this matter there is the practical need for the post PCR processing, meaning that the product of the amplification still needs to be manipulated to perform the agarose gel run, which prevents the complete automation of the traditional PCR analysis.

It is also true that, even if the agarose gel results can in principle give information on the yield of the reaction by evaluating the size of the stained bands at the end of the run, its resolution is very poor. The limit is about a tenfold change between different end-point concentrations; this issue has also been resolved with real-time PCR, which can detect as little as a twofold change.

The key element of traditional PCR, the thermostable DNA polymerase that basically unearthed its potential, is also a limiting factor, because it restricts the technique to the analysis of DNA samples. In fact, polymerases need a DNA template to work, as described in Par. 1.1.3, while numerous cases exist in which amplification of RNA would be preferred. To address this need, a procedure called *Reverse Transcription PCR* (*RT-PCR*) has been devised, which adds an initial, and really sensitive, transcription step to the normal procedure to produce a DNA copy of the RNA template.

It is important to highlight that, while the extreme sensitivity of the PCR process is one of its main qualities, it also makes it very vulnerable to contamination. In fields like forensics analysis, for example, PCR sensitivity can be critical, because it has the ability to amplify DNA which may only be present in trace amounts in a sample; but this also means that the same limited sample can be compromised beyond use.

Many other aspects of the PCR process, relevant for specific applications, have been improved over the years. For example, if the size of the amplicon exceeds 5k*bp*, amplification

efficiency will decrease substantially, because the Taq polymerase is unable to correctly handle long chains without introducing significant quantities of truncated product. *Long PCR* allows the processing of larger sequences by adding a proofreading polymerase that raises the yield to normal levels. Other examples are *High-Fidelity PCR*, which minimizes the number of mutations introduced during the reaction for sensitive applications, or *Hot-Start PCR*, which reduces nonspecific amplification due to reactions that happen at room temperature.

1.2 The Real-time Polymerase Chain Reaction

In traditional PCR, detection and quantification of the amplified sequence are performed after the last cycle of the reaction, and involve post-PCR analysis such as gel electrophoresis. In *real-time PCR*, which has the main purpose of making PCR a truly quantitative and automated procedure, the amount of PCR product is measured after each cycle via fluorescent dyes that emit an increasingly fluorescent signal directly proportional to the number of amplicons generated. By monitoring the reaction during the exponential amplification phase, the initial quantity of target concentration can be determined with great precision.

1.2.1 Fluorescence as a detection system

Fluorescence is the process of absorption and subsequent emission of light that is characteristic of determinate molecules, proteins or quantum dots called *fluorophores*. It can be described by means of the Jablonski diagram, an example of which is reported in Fig. 1.9, that represents the electronic states of a molecule and the transitions between them. When a photon impinging on a molecule has enough energy $E_{ph_1} = h\nu_{exc}$ to excite

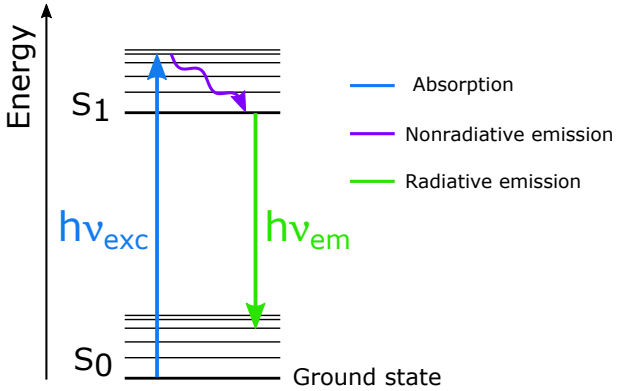


Figure 1.9: Jablonski energy diagram representing the processes of absorption, nonradiative and radiative emission in the case of a single photon excitation.

an orbital electron from its ground state S_0 to a singlet state S_1, it is said that the photon has been absorbed:

$$S_0 + h\nu_{exc} \rightarrow S_1. \tag{1.7}$$

For a given energy gap between the two states, either a single photon with equal or higher energy or two photons (with enough total combined energy) need to be impinging on the molecule to have absorption. Fluorescence occurs when an excited electron relaxes to its ground state by emitting a photon (radiative emission) of energy $E_{ph_2} = h\nu_{em}$, usually accompanied by heat (nonradiative emission in the form of vibrational energy):

$$S_1 \rightarrow S_0 + h\nu_{em} + heat. \tag{1.8}$$

Since electronic states in molecules are composed of bands of available closely-spaced vibrational states, absorption and emission are actually allowed in continuous energy intervals called *absorption* and *emission spectra*.

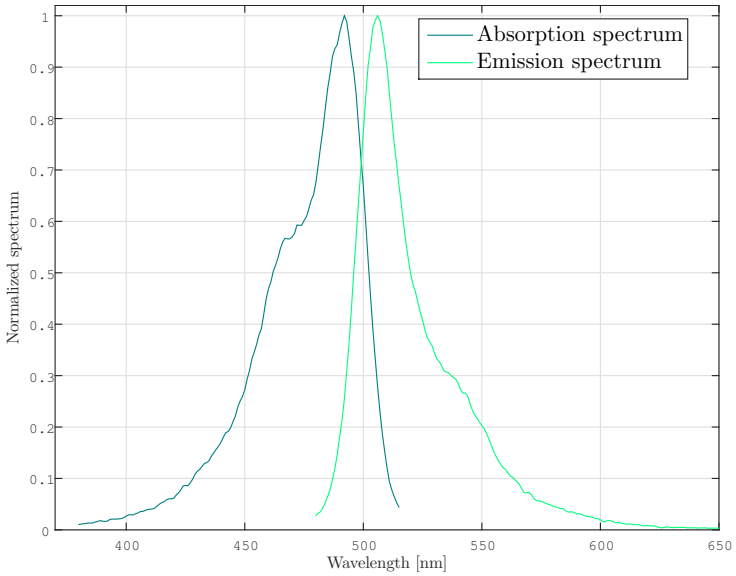


Figure 1.10: Absorption and emission spectra of Zombie Green fluorophore with 14nm Stokes shift.

The property of light absorption and emission in separate bands is what makes fluorescence one of the most useful methods of analysis of biological molecules in life sciences, because it is a non-destructive way of observing, tracking and discerning entities in a sample. Albeit some proteins and cells are naturally fluorescent, the most common way to exploit fluorescence is by *labeling* the molecule of interest with a fluorophore, which can also be referred to as a *marker*. Then, the position and concentration of the marked molecules can be detected by measuring the luminous response of the fluorophore in its emission band.

The absorption and emission spectra, that in many cases are one the mirrored image of the other, are characterized by the separation in terms of wavelength between their respective maxima. This distance is called *Stokes shift* (Fig. 1.10), and is an important parameter to take into consideration especially when choosing more than a fluorophore for a specific application. Using different fluorophores can in fact extend the potentialities of a single PCR analysis, allowing multiple different tests in the same run. It is clear, though, that an higher number of coexisting channels requires an increasingly careful design of the detection strategy, because each emission spectrum has to be clearly and univocally distinguishable from the others to avoid uncertainty in the result interpretation.

1.2.2 Real-time fluorescent PCR chemistries

When designing a real-time PCR experiment, the choice of the detection chemistry that generates fluorescence during the reaction plays a most relevant role. There are two major options, one which is specific to the DNA target sequence (probes) and one that is not (dyes).

The latter is the simplest solution, which consists in the use of a binding dye (*SYBR Green I* is the most common) that attaches to the DNA double helix during amplification. In normal conditions the dye is mildly fluorescent, but it produces a much stronger signal when bound to DNA, as shown in Fig. 1.11. When the dye is added to a sample, it

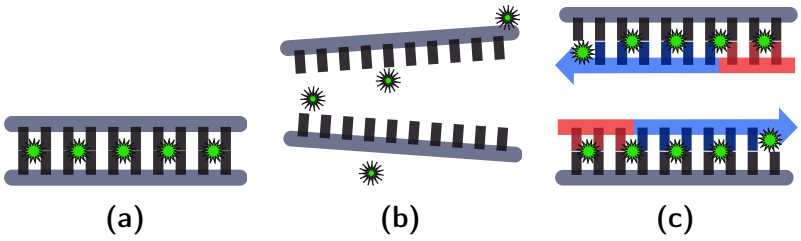


Figure 1.11: Binding dyes can either be intercalators or minor-groove binders, SYBR Green pertains to the first group. At any step of the reaction, there will be some SYBR Green dye bound to double-stranded DNA (a). When it is denaturated, the dye is released and fluorescence is drastically reduced (b). During extension SYBR Green binds again to the new product, emitting a stronger signal as the copies grow (c).

binds to all the double-stranded DNA available. During the PCR it also binds to each new copy, and as more amplicons are created the result is an increase in fluorescence intensity proportionate to the amount of product amplified. Binding dyes can be used to monitor the amplification of any double-stranded DNA sequence, without requiring particular design steps, which reduces assay setup time consumption and running costs. The sensitivity is also remarkable, because multiple dyes can bind to a single amplified

molecule. The downside of this is that also undesired sequences (*primer dimers*) emit fluorescence as the intended ones, resulting in the possibility of false positives. Primer dimers are the product of non-specific annealing and primer elongation events, and as their formation takes place during each PCR cycle, the only way to distinguish them from the specific amplicon is to perform an additional analysis at the end of the reaction. This is done with a *melting* (or *dissociation*) *curve*, that charts the change in fluorescence observed when double-stranded DNA with incorporated dye molecules dissociates into single-stranded DNA as temperature is raised. When the melting point temperature of any sequence is reached, the attached dye is released and the fluorescence signal decreases proportionally. For a very specific PCR reaction, the derivative of the fluorescence vs. temperature curve shows only a narrow peak around the T_m value of the target, while the presence of primer dimer formations is revealed by multiple peaks at lower temperatures (Fig. 1.12).

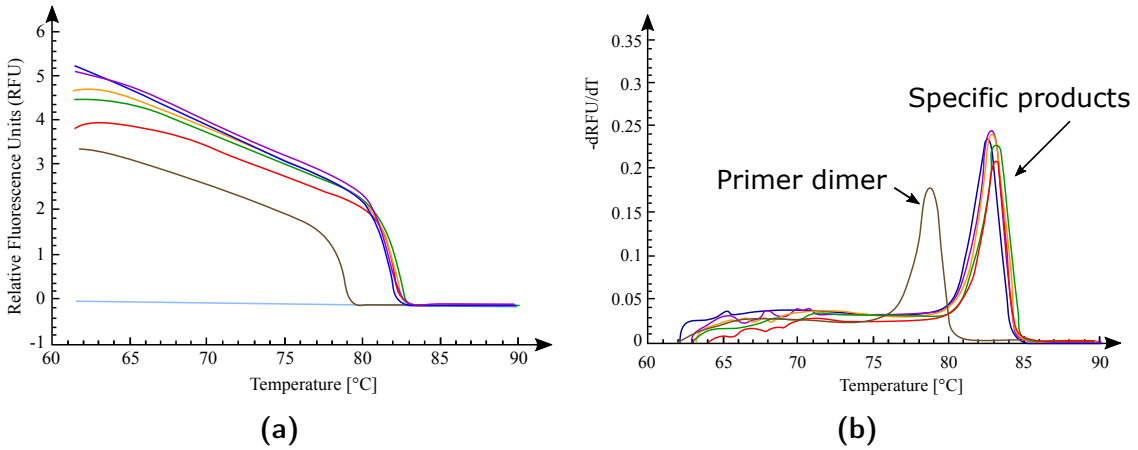


Figure 1.12: Typical dissociation curve from a solution of DNA and SYBR Green I. The change of fluorescence as temperature rises may show unsharp transitions, offsets or drifts in the signal (evident below 80 °C) (a). In this case the derivative curve is plotted (b), and the location of its peak value gives T_m more clearly.

The second detection alternative was made possible with the introduction of labeled oligonucleotide probes, which enabled the development of a real-time method for detecting only specific amplification products. The most widely used are TaqMan probes, which were named after the '80s video game Pac-Man [8]. Each probe is designed to bind downstream of one of the primers that identify the target sequence, and is labeled on its 5' end with a fluorescent reporter while on the 3' end there is a quencher molecule. The operating principle of this detection strategy relies on a property of the Taq DNA polymerase called *5' exonuclease activity* and on an energy-transfer phenomenon called FRET. The former denotes the ability of the Taq polymerase to degrade any DNA which it finds bound to the

template while performing DNA synthesis. FRET (*fluorescence resonance energy transfer* or *Förster resonance energy transfer*) is a mechanism that allows nonradiative energy transfer between two light-sensitive molecules (*chromophores*), donor and acceptor, that are close enough to one another. As shown in Fig. 1.13, if the donor is in an excited state it may transfer energy to the acceptor through nonradiative dipole-dipole coupling. The efficiency of this process is inversely proportional to the sixth power of the distance between the chromophores, which makes FRET very sensitive to relative position variations. While

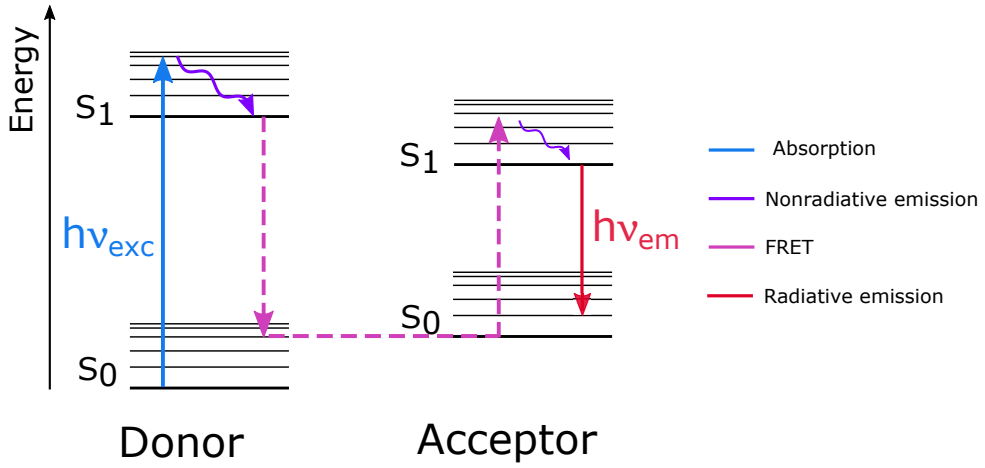


Figure 1.13: Jablonski diagram of FRET. The process involves the energy transfer between a donor and an acceptor chromophore.

the probe is intact, the proximity of the quencher to the reporter ensures good efficiency in the FRET process, so that most of the energy transferred is re-emitted as heat and the overall level of fluorescent signal is low. If a specific probe is added to a sample containing the target sequence it is designed after, it will anneal downstream from one of the primer sites and will eventually be cleaved by the Taq polymerase. The cleavage separates the reporter from the quencher, thus increasing the fluorescence signal, while allowing primer extension to continue as normal (Fig. 1.14). With each cycle additional reporter dye

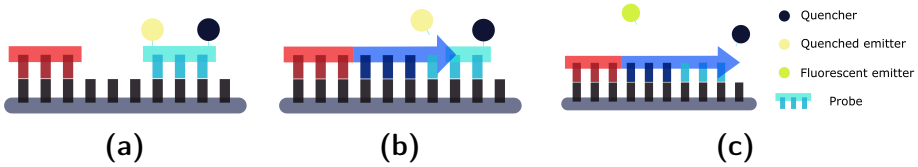


Figure 1.14: TaqMan probes are designed to bind to specific points on the target DNA sequence. While a probe is intact, the reporter emission dye is quenched (a). During the extension cycle, the Taq polymerase cleaves the reporter from the probe (b) separating it from the quencher and therefore starting fluorescence (c).

molecules are cleaved from their respective probes, so that the increase in fluorescence

intensity is proportional to accumulation of PCR product. The choice of this kind of chemistry gives the advantage of adding specificity to the analysis, and this not only removes the risk of false positives but also allows the amplification of distinct sequences inside the same reaction chamber. The price is paid in terms of design: a different probe has to be synthesized for each unique target sequence, which adds cost and time issues.

1.2.3 Real-time PCR results

The instrument that performs a real-time PCR combines the execution of the thermal cycling required for the reaction, performed by a *thermocycler*, with the optical detection of the change in fluorescence intensity. By plotting the measured fluorescence against the relative cycle number, it draws an amplification curve that describes the accumulation of product over the duration of the entire reaction. A typical real-time PCR amplification curve, characterized by four distinct phases, is shown in Fig. 1.15. The *baseline region*

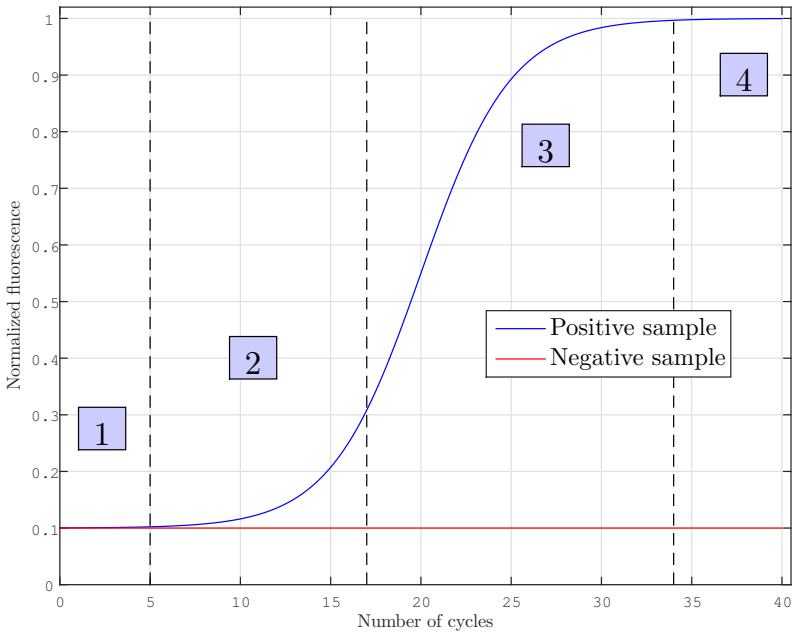


Figure 1.15: Normalized fluorescence as a function of the number of cycles in a typical real-time PCR process. Four characteristic regions (1. Baseline, 2. Exponential, 3. Linear, 4. Plateau) can be observed, each one referring to a specific phase of the reaction.

covers the first cycles of the reaction, when the growth of the target concentration is not relevant enough to make a difference over the background fluorescence signal. In the absence of a negative control, this is the reference value used to define the actual fluorescence variation throughout the process. It is worth noting that a negative sample will give as a result a flat curve, stable to the baseline value until the last cycle. In

presence of a positive sample instead the signal begins to grow, and the cycle at which the amplification fluorescence exceeds a chosen threshold (usually calculated by real-time PCR dedicated software) above the background fluorescence is called the *cycle threshold* or C_t *value*. It marks the early *exponential phase* (see Eq. 1.4). The theoretical doubling of the amplicon can be recognized in this phase, and since the reaction here is very specific and precise the C_t is the most accurate point to estimate the initial target DNA concentration in the sample. Even though there is a quantitative relationship between amount of starting target sample and amount of PCR product at any given cycle number, out of the exponential phase the performance degradation would yield less precise results. Since the cycle threshold essentially shows the number of cycles that are necessary to detect a real signal from the sample, it is clear that C_t values are inversely proportional to the amount of targeted sequences. Typically C_t values below 29 cycles show high amounts of target DNA in a sample, while C_t values around 38-40 indicate minimal quantities or could even be false positives. When the reaction components begin to become limiting the efficiency falls, so that the signal growth slows in the *linear phase*. A *plateau* is then observed when the reaction cannot generate any more fluorescence because of the reagents exhaustion and degradation.

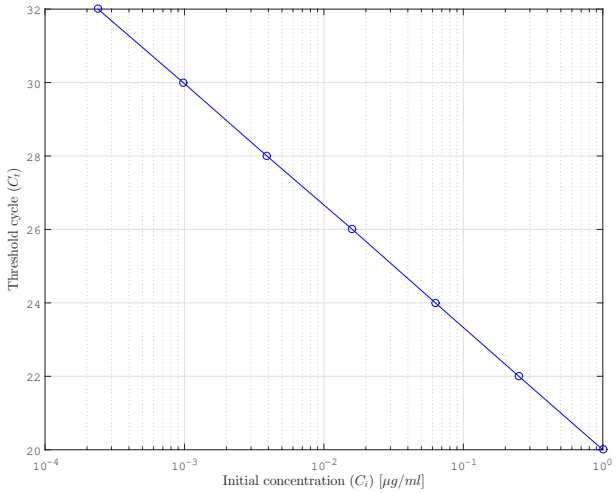


Figure 1.16: Example of a PCR calibration semilog curve. It matches the fluorescence signal at C_t crossing with the concentrations C_i of the initial target.

The quantitative information about the initial concentration of the target is extracted using a standard calibration curve like the one shown in Fig. 1.16[2], that must be generated separately for each sequence of interest. This is done by performing real-time PCRs on a number of known dilutions of the target sequence, and then measuring the relative threshold cycle values. Plotting each C_t value against the starting concentration on a

[2]Data taken from *Guide to Performing Relative Quantitation of Gene Expression Using Real-Time Quantitative PCR* tutorial by Applied Biosystem, `http://www.appliedbiosystems.com/`

semi-logarithmic scale gives a set of points that are interpolated by a straight line, which can then be used as a calibration curve for extrapolating initial concentrations of the sequence of interest in unknown experimental samples. The reliability of this method depends on the validity of the assumption of identical amplification efficiencies for all the known and unknown samples.

1.3 Real-time PCR applications and instruments

DNA amplification and detection are amongst the most critical techniques used in biological research today, and real-time PCR has proven to be the leading technology of choice. Scientists in the most diverse disciplines rely on this method for a wide range of applications, and its increased popularity since the first availability of PCR thermocyclers in the mid-1990s is reflected by the ever-growing number of scientific publications on the subject (Fig. 1.17).

Figure 1.17: Growth of the number of publications describing real-time PCR applications over time. The graph shows the number of papers resulting in PubMed Central database at NCBI (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov) per full year for the key words "real-time PCR".

Quantitation of gene expression is by far the most common application of real-time PCR. Gene expression studies try for example to understand the functioning of body tissues, or the complex interplay of host and microbe responses during infections.

The real-time PCR technology has already had a major impact on life sciences for speeding up the sequencing of genomes from various organisms and, most notably, humans. The *Human Genome Project* (HGP) published the first draft sequence of the three billion base pairs of the human genome on February 2001 [9], which was completed in April 2003, ahead of schedule, thanks to the invaluable support provided by PCR techniques. Albeit virtually every human disease has some genetic basis, until recently this was taken into consideration only in a limited set of instances, when an illness had predictable inheritance patterns because it is caused by a change in a single gene. The information generated from the HGP and subsequent studies is meant to be used for a broad range of biomedical researches, many of which aim to understand the role that multiple genetic factors acting together play in complex diseases such as cancer or diabetes. The near-future endpoint seems to be one where each patient could receive individual treatments tailored on their unique genomic make-up.

The diagnosis of infectious diseases has become another important field for DNA amplification processes since the first reported applications of PCR for the detection of the HIV virus type 1 in 1988 [10]. The PCR technology fitted a growing need for the development of methods that are rapid, specific and sensitive both in the clinical environment and in context that are considered biothreats from either military or civilian (i.e. terrorist) sources.

With the help of PCR, researchers have also been able to clone the DNA from samples as old as 40 000 years, giving birth to new disciplines such as *molecular archaeology* (which studies DNA recovered from human skeletal remains and mummified bodies) and *molecular palaeontology* (which refers to the analysis of DNA from ancient human, animal and plant remains).

Furthermore, PCR technology has rapidly become a well-recognized tool in forensic sciences, by means of the *DNA fingerprinting* technique. It is used to reveal or exclude a link between biological evidence (such as hair, blood and other fluids) and a suspect in a criminal investigation comparing DNA samples. DNA fingerprinting is also used to establish paternity with a strong degree of certainty.

Real-time PCR is now also an accepted analytical tool within the food industry, where it is being employed to confirm the authenticity of foods, detect and quantify genetically modified organisms and, in the food safety field, to identify allergenic species or adulterated[3] foods.

The real-time PCR major impact across so many disciplines of the biological sciences has been a driver to develop and improve existing real-time PCR systems. The first two commercial platforms were introduced in the mid-1990s, and while there is now a

[3] "Adulteration" is a legal term meaning that a food product fails to meet health or safety standards.

continuously growing choice of systems, their fundamental architecture has remained unchanged through time: it consists of a thermal cycler, optics for the fluorescence excitation source and a detector to collect the emission. Additionally, each machine will come with the necessary proprietary computer software for data acquisition and analysis. The main differences between the various platforms concern the technology used to heat and cool the reaction chambers, the strategy for light channeling in and out of them, the excitation method (lasers or filtered broadband sources), the number and type of available optical channels and sample capacity. Also the data processing software can present specific differences, mainly focused on the choice of the adopted quantification strategy. While theoretically only a single copy of a target sequence is needed as a template, typically several thousand copies of the target DNA are seeded into each reaction. This can amount to a starting DNA quantity ranging for example from 10 µg to 1 pg per 50 µL, depending on the template complexity [11]. In relation to this, all the reagents described in Par. 1.1.2 need to be added in correct quantities to optimize the amplification reaction. The total reaction volume will therefore range between 10 µL and 200 µL, and usually any specific instrument will have a single well capacity which is a smaller subset of this range (e.g. 5 µL - 30 µL). Modern standard real-time PCR machines usually host racks of 32, 48, 72, 96, 384 or more reaction wells, sometimes in multiple configurations.

1.4 The Point of Care perspective

Operating a real-time PCR machine typically requires well-trained personnel inside a specialized biological laboratory, because a deep knowledge of the required protocols to setup the system is essential, along with the expertise to read and analyse the results. On top of this, a single machine can easily cost more than $\$40,000$, therefore owning one often remains a prerogative of a relatively limited number of laboratories and institutions. These factors contribute to adding a long time and considerable cost to the performance of a single analysis, besides the clear limitation due to the need of physical access to the laboratory facility. This is in contrast with the potentialities of the wide number of applications that have become synonym with real-time PCR, as discussed in the previous paragraph, that depict a landscape where this technology could easily find place in a variety of environments.

On the other end, recent trends in healthcare needs and disease prevention have been generating a growing demand for *point of care* (*POC*) testing. The driving notion behind POC is that the equivalent of laboratory testing can be conducted at or near the site where the person being tested is located, with the results returned during the same visit and immediately usable for patient care.

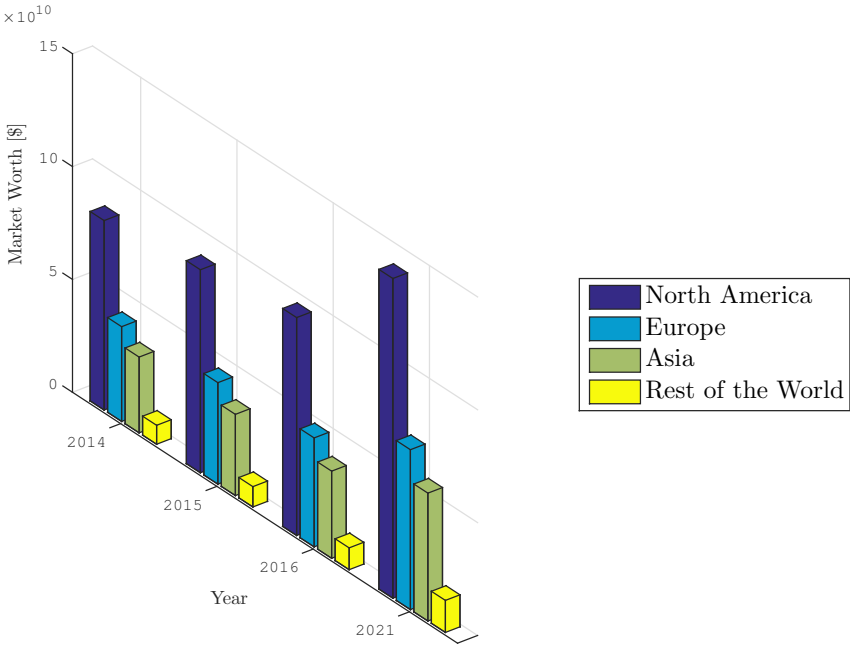


Figure 1.18: Global 2014-2016 POC market by region plus 2016-2021 forecast with an overall compound annual growth rate of 8% [1].

This approach is especially relevant in resources-limited settings like rural or underdeveloped areas, where healthcare infrastructure is weak or absent and access to medical care is a challenge. In developing countries, the struggle to provide minimal healthcare to all people living in adverse geographical and economic conditions is also accompanied by the need to monitor critical infectious disease such as HIV, tuberculosis or malaria. The situation becomes even more critical in epidemic situations, like in the recent Zika fever outbreak, which started in Brazil mid-2014 and has since reached the Americas, Pacific Islands and Southeast Asia.

With the growing potential of POC testing in the developing world in mind, the World Health Organization (WHO) provided guidelines for those designing new devices known as *ASSURED*, which are shown in Table 1.2.

In a detailed report about the availability and use of HIV diagnostics in low and middle income countries [12], for example, the WHO documented how implementation and diffusion of POC testing is rapidly rising. Various laboratory technologies are used for diagnosing HIV infections, between those *cluster differential 4* (*CD4*) testing is primarily used for monitoring HIV patients to determine when antiretroviral therapy (ART) should be started and to establish treatment efficacy. This test is one of those which present the hardest challenges in terms of improving access, because of user charges for laboratory services, malfunctioning machines or a lack of reagents. The WHO reported that *'POC CD4 was found to be associated with an increased likelihood that patients would achieve*

Affordable	for those at risk of infection
Sensitive	minimal false negatives
Specific	minimal false positives
User-friendly	minimal steps to carry out test
Rapid and Robust	short turnaround time, no need for refrigerated storage
Equipment-free	no complex equipment
Delivered	to end users

Table 1.2: The ASSURED guidelines for POC devices as suggested by the World Health Organization.

CD4 testing and eligibility assessment compared to laboratory based testing, and it was found to reduce the time intervals at these critical steps in the care pathway. There was also evidence that use of POC testing could increase the likelihood of HIV-positive individuals being initiated on ART in some settings.' [13], and found that while in 2012 19% of the CD4 platforms in use in the considered countries were POC, the next year the number went already up to 23% [12].

The increase in lifestyle related illnesses such as cardiovascular pathologies, neurological disorders, geriatric diseases, cancer or diabetes also demands for POC solutions in developed countries, where home-based operation of user friendly devices can offer a more personalized and patient-centred care. Here the role required of POC technologies is slightly different because of the specific necessities that need to be fulfilled, but the underlining goal remains providing affordable health monitoring and care in settings where they would be traditionally absent. So, while some of the ASSURED criteria remain relevant, patients in developed countries will focus their needs on a relatively more sophisticated equipment that can provide quantitative results, decision support and connectivity to online information systems, for example. A 2015 IEEE paper summarizing the panel discussion at the *IEEE Engineering in Medicine and Biology Point-of-Care Healthcare Technology Conference* [14] reported that technology innovations in medicine and health, combined with the continuous growth of life expectancy, generate peculiar medical needs and require the development of new solutions. It states that *'to provide quality healthcare at reasonable costs, there has to be a convergence of preventive, personalized, and precision medicine with the help of technology innovations across the entire spectrum of POC'.*

A typical classification of POC technology splits devices into small handheld, truly portable, machines and larger bench-top ones, which resemble those available in a laboratory. The second group is made of devices that are a smaller and less complex version of the original instruments, designed to be operated by non-laboratory trained staff to care for

patients that need periodical checks. A great variety of systems are instead classified as handheld POCs, which comprise simple dipsticks, quantitative and qualitative strips, and cartridge devices. Fairly common examples include systems for glucose monitoring, blood gas analysis, drug screening, infectious agents identification and pregnancy and fertility tests. The most complex devices are operated by loading a sample in a disposable cartridge, which is inserted into a reader for measurement and evaluation. This kind of systems embodies the *Lab on Chip* (*LOC*) concept, that grew from the microelectronics industry two decades ago. Its aim is to integrate into a small instrument, with the aid of miniaturization and microfabrication techniques, the required components to perform all the processes associated with a traditional laboratory analysis, from sample collection to end result. In a LOC a silicon, glass or polymer microchip is equipped with typical microfluidics devices (such as microchannels, microvalves, micropumps), control electronics and various types of sensors, to form a complete system that can also be enhanced with advanced user interfaces (such as touch screens), built-in bar code scanners, connectivity to information systems and more.

1.5 STMicroelectronics device and purpose of this work

While POC devices are already employed in both developing and developed countries, there is a growing number of unmet needs that require design of new devices and improvement of the existing ones. A recent study assessing primary care clinicians' desire for an increased availability of POC tests [15] found that all respondents in the five countries that participated in the survey *'wanted POC tests to help them diagnose acute conditions (infections, acute cardiac disease, pulmonary embolism/deep vein thrombosis), and some chronic conditions (diabetes, anaemia)'.* Reports on the global POC market [1] highlight how the major forces driving the industry go towards providing better patient care by improving the turn-around-time (TAT) that POC testing offers, while at the same time granting easiness of access and use to the widest user base possible. This implies the need to design less complex laboratory tests and more user-friendly devices, easily accessible by less qualified personnel. Another relevant issue concerns the growing volume of data generated by POC instruments, that raises the need for expanded connectivity capabilities and a strong integration with efficient data management systems. In this way the information can be organized and made directly available to physicians, to create individual medical records of patients.

Taking the chance to add a competitive contribution in this wide landscape, STMicroelectronics has started the development of a small handheld real-time PCR POC device, called Q3-Plus, which revolves around the use of disposable LOC silicon testing cartridges.

The well known biocompatibility of silicon dioxide makes it a favorable choice for the reaction substrate of the chambers, which are the core of a disposable chip that comes to the user already preloaded with all the needed chemistry. The instrument has, since its inception, undergone a steady evolution process that has led it to a condition of great functional maturity.

In this thesis I describe my year of work at STMicroelectronics and in the IFN-CNR laboratories at the Department of Physics of Politecnico di Milano, which has been focused on the analysis and development of some features of the Q3-Plus. The first months were dedicated to learning how the optical subsection of the device works in detail, while also performing an actual characterization of its performances. The insight gained from those activities later led to conjectures and experiments that eventually gave life to a redesigned and optimized structure of the entire block. At the same time I have been aiding the research group at STMicroelectronics with the planning of the next evolution of the device, particularly focusing on the extension of its optical capabilities and on the redesign of the detection system. The second part of my internship has then been mostly spent working at the Department of Physics, learning and applying the principles of the Two-Photon Polymerization (*2PP*) technique in order to design and fabricate a suitable lens for the new prototype.

Chapter 2

STMicroelectronics real-time PCR Point of Care device

STMicroelectronics Q3-Plus device is a compact platform that can perform real-time PCR analyses on up to four different wavelengths at the same time. It can be described as the ensemble of three subsystems, namely the disposable Lab on Chip, the temperature control circuitry and the excitation/detection structure. Their design and working principles are illustrated in this chapter, with particular detail on the Optical Control System, which has been the main subject of my work. The instrument is also equipped with an optimized software interface, which guides the user through the test operation and reports the results at the end of the analysis. This is also briefly discussed to complete the system overview.

2.1 STMicroelectronics Q3-Plus real-time PCR platform

With the purpose of creating a low cost handheld device that could bring POC diagnostic capabilities to a wide range of scenarios, from household or patient bedside locations to disadvantaged rural settings, in the past few years STMicroelectronics has devised and developed the Q3-Plus real-time PCR platform, shown in Fig. 2.1. The three *Q*s that name it describe its main features: it performs Qualitative, Quantitative and Quick analyses of DNA and RNA target molecules.

Inside the Q3-Plus device the traditional real-time PCR analysis is performed on disposable Lab on Chip cartridges, where all the amplification, detection and quantification processes are carried out. This has been made possible thanks to advances in silicon

Figure 2.1: STMicroelectronics Q3-Plus device.

micromachining and device integration brought on by MEMS (*Micro Electro-Mechanical Systems*) technologies. The main feature of a MEMS device is the capacity to integrate on the same silicon substrate many types of structures that pertained traditionally to separate domains, such as microfluidic structures (channels, pumps, valves, reservoirs), microsensors (temperature, pressure) and the typical CMOS control and processing electronics. The Lab on Chip model is, in fact, very well transposed into practice through the synthesis of those structures, especially because the small size that descends from such integration process grants a great advantage in terms of electronics performance, reduced reagents consumption, sample requirement and analysis time. Furthermore, the whole system can be realized relying upon well known and mature semiconductor mainstream technologies, which means that most of the fabrication can be done with already available and optimized industrial equipment. The key point here is the opportunity to use silicon as a substrate for the reaction, both in its natural (oxidized) form or after proper functionalization. The latter is a process used to modify characteristics such as wettability or hydrophobicity, link assay-specific molecules to the surface or, in general, change its features. In fact, from a chemical point of view silicon dioxide is biocompatible, meaning that it does not interfere with the chemistry inside the cartridges, and has optimal thermal properties. Silicon relatively high thermal conductivity (149 $\mathrm{W/(m \cdot s)}$) assures accurate temperature homogeneity across the chip, while low specific heat capacity (0.7 $\mathrm{J/(g \cdot °C)}$) guarantees fast ramping rates between heating and cooling steps.

Fig. 2.2 shows a comprehensive view of the Q3-Plus internal structure. The disposable cartridge contains a number of wells (usually six) where the sample and the needed chemistry are mixed when a test is performed. Everything is then sealed with a transparent

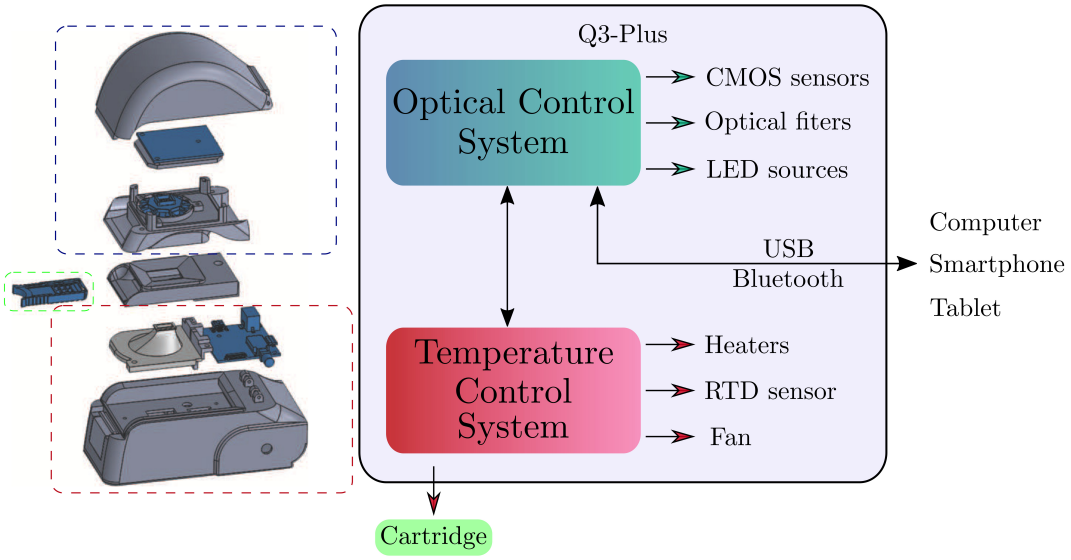


Figure 2.2: Q3-Plus SolidWorks model exploded view, with its main components highlighted.

slider and placed inside the input slot of the device. The *Temperature Control System* (*TCS*) performs the initial calibration and all the subsequent heating/cooling steps required by a specific testing protocol. All the excitation and detection procedures happen on the upper part of the device, where the light sources, filters and detection sensors are located. This is called the *Optical Control System* (*OCS*).

As a whole, the compact hardware module appears as a closed system to the end user, so that its operation can be performed by unskilled personnel at the point of care with the guidance of a dedicated control software. This allows the operator to easily exploit the Q3-Plus full potential without necessarily having to grasp the complex environment behind it.

2.1.1 Disposable chip

As shown in Fig. 2.3, the disposable cartridge is composed of a silicon substrate, a polycarbonate ring glued on it, and a plastic holder. The ring divides the reaction volume in six chambers, while the holder keeps everything together and is the mechanism used to slide the chip inside the Q3-Plus. A transparent slider, used to protect the sample and the chemistry once they are together inside the wells, is usually added to cover the cartridge. The reaction wells can also be pre-filled with all necessary biological reagents, and sealed with paraffin wax for additional protection.

All the components of the cartridge are carefully chosen so that no optical nor chemical interference is introduced in the PCR and fluorescence detection. Silicon would for example

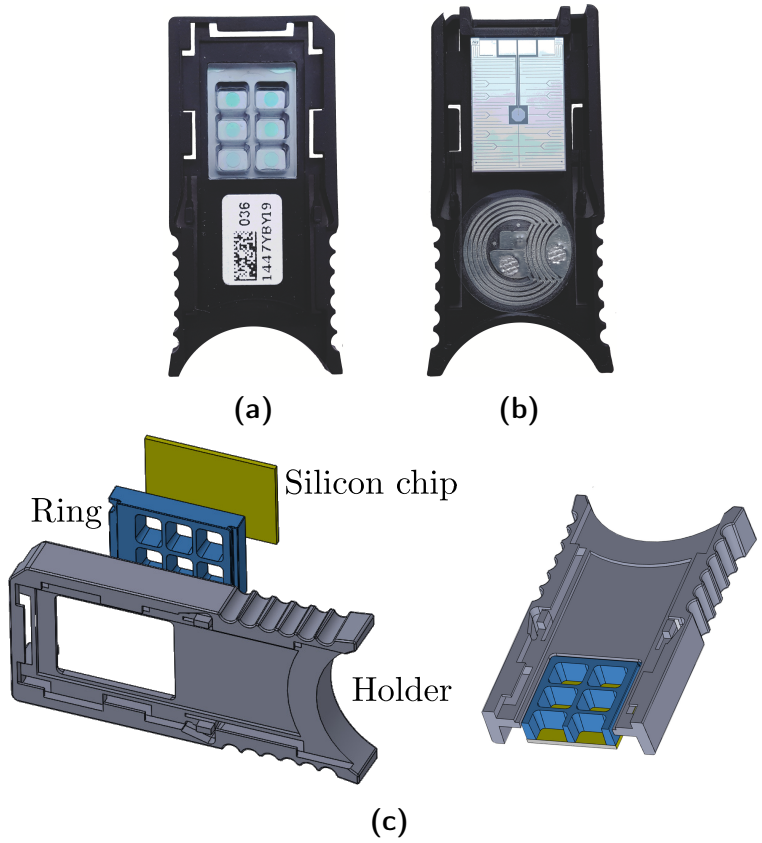
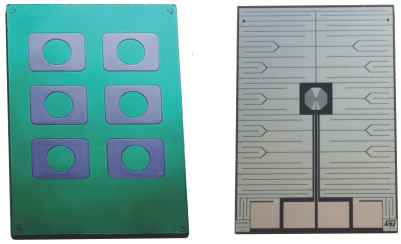


Figure 2.3: A standard real-time PCR cartridge seen from above (a), below (b) and from a CAD design (c). The model cross section shows how ring and chip joined together define the reaction wells. In Figure (b) an RF-ID label is also visible.

inhibit the PCR by absorbing the DNA polymerase, therefore a layer of silicon dioxide is grown on top of it to avoid this phenomenon, while also increasing its hydrophilicity. The ring material needs to be biocompatible, transparent to the operating wavelengths of excitation and emission and must not emit fluorescence in those windows, so it is made of polycarbonate and glued on silicon; for the same reasons the cover slide is printed in polymethylmethacrylate (PMMA), which is used instead of polycarbonate because it is more transparent. Depending on the desired behavior of the substrate, the ring and chip are glued together either with UV-curing adhesive (hydrophilic) or siliconic glue (hydrophobic). Each well is 3 mm deep and has a reaction volume that can range from 5 µL to 20 µL.

Fig. 2.4 shows the two sides of the silicon chip: the top one is the surface where the reactions happen, on the bottom face there are the components deputed to thermal control. Those are four aluminum resistors connected in parallel and a temperature sensor, placed at the center of the chip. On this side there also is a small section with four contact pads, which connect the resistors and the sensor with the TCS through a *flex*



(a) (b)

Figure 2.4: Upper (a) and lower (b) sides of a silicon chip. Figure (a) clearly shows the six spots that correspond to the center of each well; here the substrate is functionalized with a layer of *fluoroctatrichlorosilane* (*FOTS*) wich prevents adhesion of reagents outside the circle.

board. The resistors are arranged across the whole silicon surface to heat it uniformly, and the achieved temperature is measured by the sensor, which is an aluminum *resistance temperature detector* (*RTD*) with a nominal resistance value of 1 kΩ at 25 °C. The resistance versus temperature relationship of an RTD can be well approximated as linear within its working range as expressed by Eq. 2.1, where the α coefficient is material-specific ($\alpha_{Al} \approx 3.9 \cdot 10^{-3}$ °C^{-1}) and R_0 is the nominal RTD resistance at a reference temperature T_0.

$$R_{RTD}(T) = R_0 \cdot [1 + \alpha_{Al} \cdot (T - T_0)] \tag{2.1}$$

The silicon chip has an area of 17.65 mm by 12.15 mm and is 675 µm thick; Fig. 2.5 shows a full 8" uncut wafer of chips.

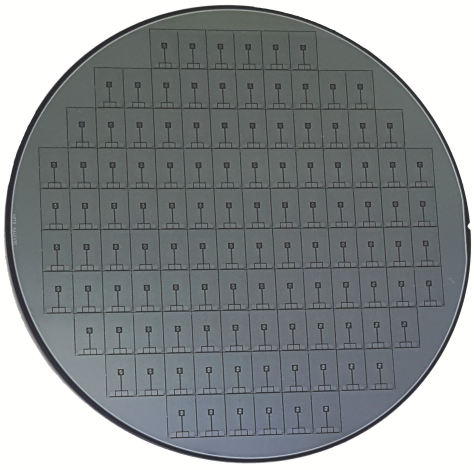


Figure 2.5: 8" wafer of chips, ready for cutting.

2.1.2 Temperature Control System (TCS)

To execute a real-time PCR according to a given protocol, it is necessary to coordinate the required thermal cycling with the sample excitation and the relative fluorescence detection at each step of the sequence. The collected data needs then to be sent out for elaboration and display before each new cycle starts. Some of this activity is overseen by the TCS, which is shown in Fig. 2.6. The TCS PCB measures 5 cm by 4 cm

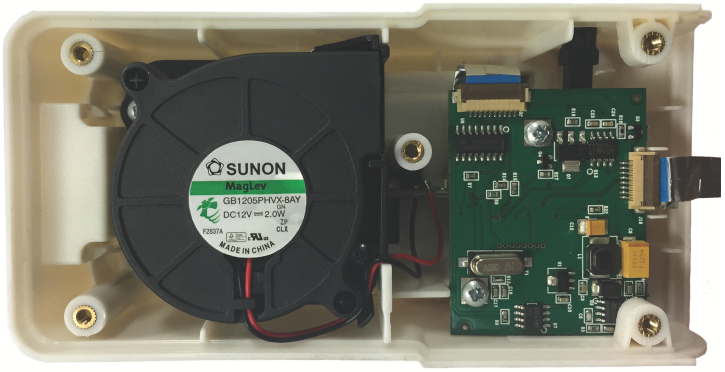


Figure 2.6: Temperature Control System board and cooling fan inside the Q3-Plus case.

and mounts an STMicroelectronics STM32F103 microcontroller, plus all the necessary components to manage heating, cooling (a fan plus the relative circuitry) and temperature detection. The board is connected with the Optical Control System via the *UART* (*Universal Asynchronous Receiver/Transmitter*) protocol, and is equipped with a micro-USB connector that allows communication with a computer. The actual USB control is however located on the OCS board, the TCS just hosts the connector. A conceptual scheme that summarizes the TCS structure is represented in Fig. 2.7.

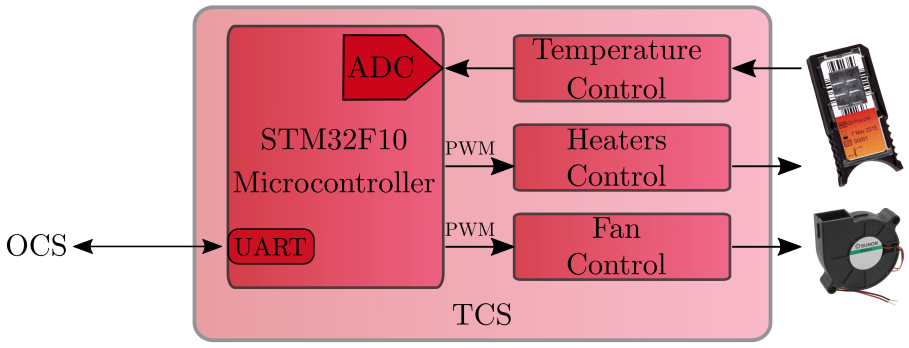


Figure 2.7: Block diagram of the main TCS board functions.

The board is powered at 12 V through a jack connector, and by means of voltage

dividers the required bias values are routed to the different components across the TCS. The microcontroller itself needs 3.3 V to operate. The heating and cooling processes are controlled with the same mechanism: the power each section receives is modulated by a *Pulse Width Modulation* (PWM) signal, provided by the microcontroller, which drives the gate of an nMOS transistor used as a switch (Fig. 2.8). A PWM signal is a square wave with a fixed frequency and variable duty-cycle, in this specific case $f_{PWM} = 600$ Hz and the high and low amplitudes are 0 V and 3.3 V. Since the thermal and mechanical time

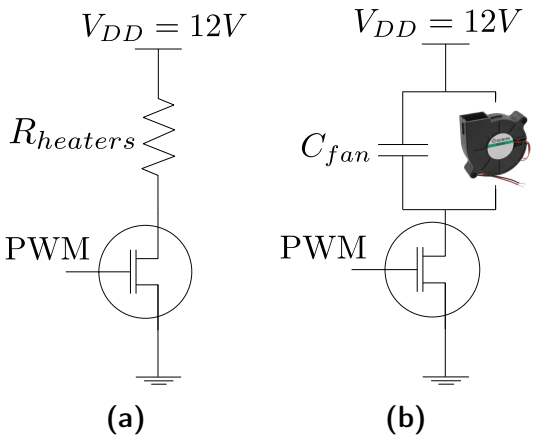


Figure 2.8: Simplified circuit diagrams of the heating (a) and cooling (b) sections.

constants of the resistors and fan are substantially larger than the PWM period, the power actually dissipated on the components is an average value that depends on the selected duty-cycle. The average dissipated power can be calculated as

$$P = (V_H \cdot DC)^2 / R_{eq} \tag{2.2}$$

where V_H is the high value of the PWM signal, DC is its normalized duty-cycle and R_{eq} is the equivalent resistance of the component; in the case of the silicon chip it is just the integrated resistors value, which on the whole amounts to 23 Ω. Given the dependence of power on the square value of the applied voltage, it is more convenient to directly connect the resistors to the available 12 V signal to obtain a faster process. The heating system has also a protection circuit designed to safeguard the microcontroller from overvoltages that could happen when the cartridge is not inserted in the device. This is not necessary in the cooling circuitry, because there are no removable components here.

The fan is the GB1205PHVX-8AY model by Sunon, designed to also operate with a 12 V DC voltage. It has a maximum 6000 RPM, employs a brushless motor and has 2 W of power consumption. The mechanical low-pass filtering of the fan is enhanced with a parallel 100 nF capacitance C_{fan}, as shown in Fig. 2.8b. Just like in the previous case the duty-cycle set by the microcontroller regulates the device behavior, in particular the

rotational speed is proportional to the duty-cycle and so is the air flow which is directed onto the cartridge.

The chip temperature is controlled in a feedback loop with a *Proportional Integral Derivative* (*PID*) algorithm that sets the duty-cycle values of the PWM signals described above. The PID controller continuously calculates the error value $e(t)$ as the difference between a desired setpoint temperature and the actual temperature read from the sensor, and applies a correction based on proportional, integral and derivative terms as shown in Fig. 2.9. The controller attempts to minimize the error over time by adjusting the control

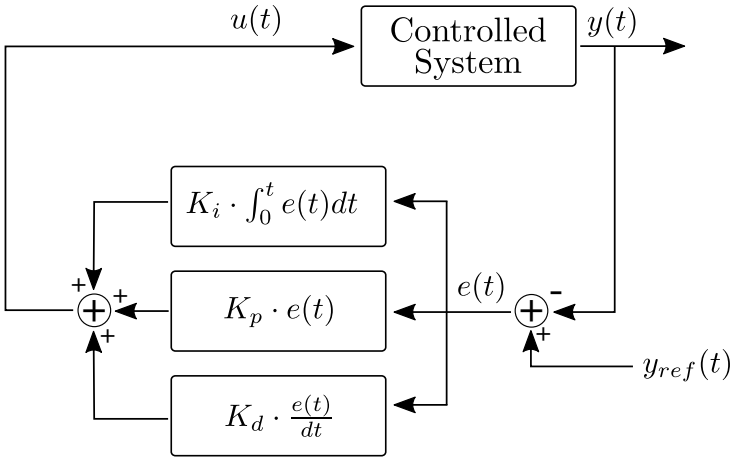


Figure 2.9: Block diagram of a PID controller in a feedback loop. $y(t)$ is the measured temperature value, $y_{ref}(t)$ is the temperature desired setpoint value, $e(t) = y_{ref}(t) - y(t)$ is the error and $u(t)$ is the control variable that acts on the PWM signals generation. K_p, K_i and K_d are, respectively, the proportional, integral and derivative gain used to tune $u(t)$.

variable $u(t)$, which drives the PWM signals, to a value determined by a weighted sum:

$$u(t) = K_p\ e(t) + K_i \int_0^t e(t)dt + K_d \frac{de(t)}{dt}. \tag{2.3}$$

The first term produces an output value that is proportional to the current error value, it needs to be fine-tuned to find a compromise between slow response and instability risk. The integral term uses the accumulated temperature error that should have been corrected in the past to detect whether the corrective action has been too strong or too weak, modifying $u(t)$ not only in relation to $e(t)$ but also to the time for which it has persisted. The derivative term considers the rate of change of error, acting on the prediction of future system behavior and thus improving settling time and stability. An accurate choice of the PID parameters, which are the gains of the three terms that weigh the control variable, is required to guarantee high stability and short transient times to the control loop. The algorithm needs then to be practically implemented in the microcontroller, therefore the

actual control variable is described as a function of a discrete-time variable, k:

$$PWM[k] = K_p\ e[k] + K_i \sum e[k] + K_d\ \{e[k] - e[k-1]\} + PWM[k-1]. \tag{2.4}$$

The temperature is sampled at f_s =20Hz, therefore every 50 ms the RTD output voltage is read, $e(t)$ is calculated and new PWM parameters are sent to control the heaters and the fan. The PID algorithm is used only for temperature fine-tuning, when the difference between actual and desired temperature is less than ±2 °C; if the difference is positive and higher than 2 °C the PWM signal for the heaters is deactivated and only the fan will work, in the opposite condition the heaters will have a PWM with the maximum possible duty-cycle while no cooling action will occur.

Each time the device operation is initiated for a new test, a double calibration (board and chip) is performed, to ensure an accurate and reliable thermal control. Fig. 2.10 shows the circuitry involved in the board calibration. The purpose of this operation is to

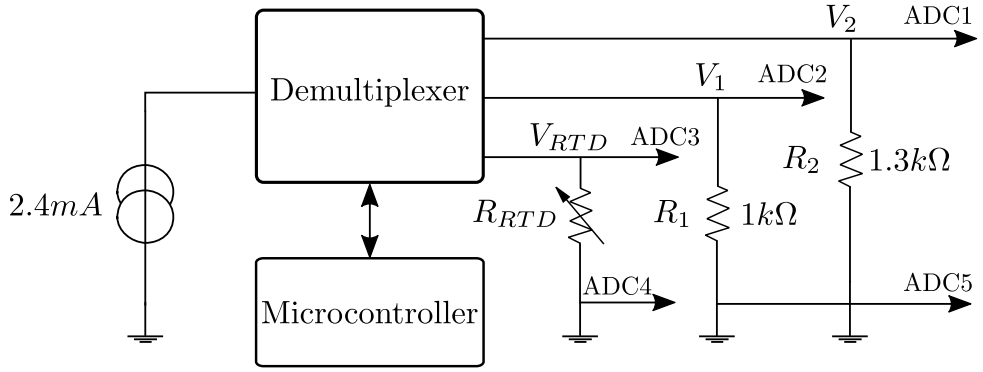


Figure 2.10: TCS board calibration schematic. R_{RTD} is the temperature sensor variable resistance that needs to be read.

calculate the relationship between temperature variation and RTD resistance variation. Since the resistance value is calculated by reading the voltage drop on the resistor when a known current flows through it, the knowledge of the dependence of the current itself on temperature variations will allow a correct reading of R_{RTD} throughout the real-time PCR execution. The calibration system employs a 2.4 mA current generator, a demultiplexer and two high precision resistors of 1 kΩ and 1.3 kΩ respectively. Those two values are chosen as they are the limits of the allowed working range of the RTD. As a first step, to measure the actual current value I_0 at room temperature, the demultiplexer lets the generator feed one resistor at a time while the relative voltage drop is sampled and averaged. This gives a calibration voltage-resistance curve in the range of interest. Each time R_{RTD} needs to be read, the microcontroller first samples the voltage drop on R_1 and evaluates the current flowing on the resistor as $I(T) = I_0\ \beta(T)$, where $\beta(T)$ is a coefficient

accounting for the temperature effect. Then V_{RTD} is also sampled, and by expressing it as $V_{RTD} = R_{RTD}\ I(T) = R_{RTD}\ I_0\ \beta(T)$ the correct value of R_{RTD} can be extracted. The chip temperature can then be evaluated as per Eq. 2.1.

The chip calibration procedure is executed so that an absolute temperature value can be effectively associated to a resistance variation. This is done by means of a Sensirion STS21 digital temperature sensor, which communicates with the microcontroller via the *Inter Integrated Circuit* (I^2C) protocol. The sensor is located on the flex board inside the ventilation channel, so that during calibration, with the fan spinning at maximum RPM, it is in thermal equilibrium with the chip. Measuring temperature with the STS21 and resistance with the RTD and applying a proprietary STMicroelectronics algorithm, the microcontroller is then able to extract the relation between the sampled resistance value and the reference absolute temperature.

2.1.3 Optical Control System (OCS)

The Optical Control System, shown in Fig. 2.11, is located in the upper lid of the Q3-Plus, and its main function is to perform all of the operations connected to fluorescence detection. Its heart is the STMicroelectronics STM32F207VE microcontroller, which

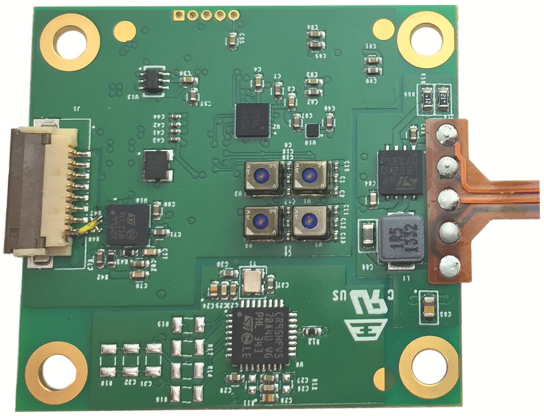


Figure 2.11: Optical Control System board. The four camera modules are visible in the center of the PCB.

drives four couples of LEDs which provide excitation light and four CMOS camera modules that acquire the fluorescence signal emitted from the sample. The cameras communicate with the microcontroller through a *Digital Camera Module Interface* (*DCMI*) bus. As described before, the OCS interacts with the TCS via UART interface and manages USB external communication with a computer, but it also hosts a flash memory module, where acquisition data are stored, and Bluetooth circuitry that allows communication with an Android smartphone/tablet app. The app lets the user control the Q3-Plus device directly

from a smartphone or tablet, so that a test can be started and the results visualized without the need of a computer. A block diagram of the structure described above is shown in Fig. 2.12. Each LED couple (which is composed of two nominally identical LEDs placed

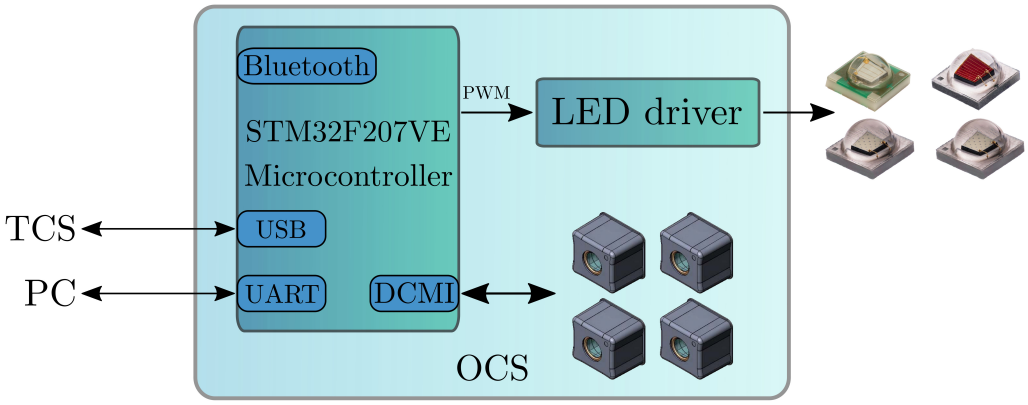


Figure 2.12: Block diagram of the main OCS board functions.

in different spots) is designed to be the source of light for one of four detection channels, and is therefore associated to one of the CMOS cameras that will acquire the response of that channel. The optical system is completed by two sets of excitation and emission filters, which are designed to precisely define the characteristic wavelength intervals of each channel. The theoretical behavior of this section will be addressed in Par. 2.2.

As the real-time PCR analysis proceeds, each CMOS camera module takes pictures of the reaction wells at regular intervals, and the relative fluorescence values are extracted isolating in the acquired data six *regions of interest* (*ROIs*) that correspond to the center of each well. The cameras are four STMicroelectronics VW6558 modules (Fig. 2.13), which have been chosen because of their small size (3.5 mm by 3.03 mm by 2.5 mm). This allows the integration inside the OCS of one of such devices for each channel. The VW6558

Figure 2.13: STMicroelectronics VW6558 camera module. The casing encapsulates the sensor and a stack of lenses above it.

sensor has a 640 x 480 VGA resolution, 60 dB dynamic range, 38 dB signal to noise ratio, and has an on-board 10 bit ADC (grayscale resolution). The lens system integrated into

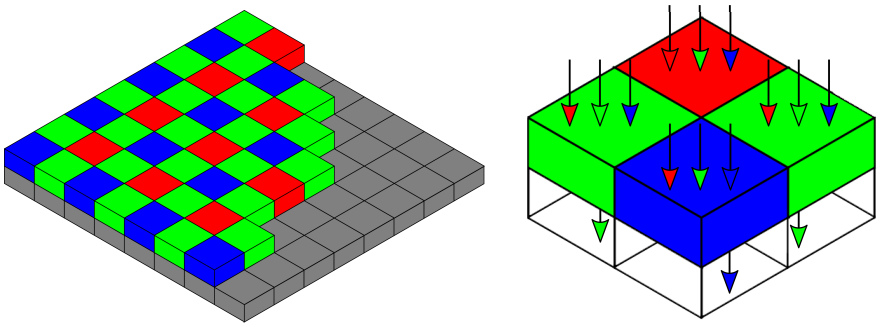


Figure 2.14: Bayer array pattern (a) and single Bayer pixel (b).

the package is designed to focus objects from 15 cm to infinity, has an horizontal field of view (HFOV) of 60°, a focal length f_{sensor} =1.25 mm and an F-number of 2.74. Inside the VW6558 sensor one can identify an analog and a digital block. The first one manages the circuit bias, row and column addressing blocks around the pixel matrix, multiplexing, amplification and conversion of the output voltages. The digital block, instead, takes care of generation and routing of the clock signals across all the components.

The single sensing entities are, as it is usual for a CMOS imaging device, reverse-biased *pn* junctions (photodiodes) which generate current when photons reach the active area and are absorbed. The output voltage of each photodiode increases with the number of photons that hit its active area, and is therefore proportional, to a first approximation, to the incident light intensity. Such a structure is inherently a monochromatic detector, because it does not differentiate between wavelengths. If the photon impinging on the pixel has at least as much energy as the silicon energy gap, it can be absorbed, generate an electron hole-pair and contribute to the diode photocurrent. This is why each individual photodiode of the sensor is covered with a filter which will block all photons except those with the right energy to be defined as either red (R), green (G) or blue (B). The arrangement of such filters, depicted in Fig. 2.14a, is called *Bayer array*, and the reading from a single 2 by 2 matrix of RGB-filtered photodiodes (Fig. 2.14b), appropriately processed with merging algorithms, gives the color of a single pixel of the final image. The green photons are sampled twice in this arrangement because the photopic curve, which is a representation of the human eye sensitivity, has a peak in the green region of the visible spectrum. The sensor outputs the sampled data as a raw Bayer pattern matrix, with the color channels still separated, which is actually more useful for the purposes of fluorescence detection because it helps discriminating between different contributions.

The digital grayscale value obtained at the end of the reading chain does not only depend on the fluorescence emission, but is also a function of two parameters of the sensor, that can be configured by writing appropriate values in the relative control registers. Those

parameters are the *integration time* (T_{INT}) and the *analog gain* (AG) in dB. The first quantity defines the time interval during which the photodiodes are collecting charge between subsequent reads of the matrix, while the second is the gain applied at the pixel output at readout. The analog gain is expressed in hexadecimal units as a value between $x0000$ and $x0010$ (corresponding to 0_{10} and 240_{10}) with an $x0010$ step (16_{10}); calling n the gain value specified into its control register, then Eq. 2.5 gives the analog gain value:

$$AG_n = \frac{256}{256 - n}. \tag{2.5}$$

The equation represents an hyperbole defined between 1 ($n = 0$, AG_{dB} =0 dB) and 16.03 ($n = 240$, AG_{dB} =24.1 dB).

Since this imaging sensor is employed in a real-time PCR device, which is not a typical application, its linearity and dynamic range characteristics are much more relevant than spatial resolution, that is the number of pixels utilized to construct a digital image. In fact, since the goal here is to monitor the fluorescence variation of the sample over time, having a blurred image of the reaction spots does not weigh so much on the final result, as long as the entire droplet where the reaction takes place is always captured by the sensor. On the other hand, a good linear behavior of each photodiode is important because the calculation of fluorescence relies on it. An extended dynamic range, moreover, allows measurement of low intensity emissions, so that also less efficient dyes can be detected and the excitation light power can be reduced. Lowering the light power, in turn, reduces the chances of photochemical alterations (known as *photobleaching* or *fading*) that fluorophores can go through during the PCR reaction, which permanently damages them.

Another relevant issue in fluorescence measurements is the contribution due to *dark current*, which sums a random offset to the value read from each pixel. Dark current consists of charges thermally generated inside the depletion region of *pn* junctions, and since the rate of this process depends on the crystallographic defects of the particular junction, it is not only different for each pixel but also very difficult to estimate. A good solution to reduce the offset caused by such a current employs a group of *dark pixels* added at the end of each active row of the sensor: those pixels are identical to the others, but are covered by a metal layer that prevents photons from impinging on their active area. Therefore, the only current that flows through them is thermally generated, and since they are biased in the same conditions as the normal pixels are, and built close to them, is it fair to assume their dark current is about the same. What is typically done thus is measure and average the dark pixels current, and subtract each value row by row to compensate the offset.

2.1.4 Computer software

A dedicated control software designed to optimize the user experience completes the picture. It can be adjusted to give full protocol flexibility, when aimed to an expert user base, or maximum easiness of use for untrained personnel. In the first case an open software is delivered to end users such as biologists or biotechnologists, and it features fully customizable reaction parameters as well as enhanced control over the result analysis. This is also used for the validation phase of any particular test that can later be transposed as a closed software application, implementing what is called a *sample to answer* procedure. In this second approach the software interface is optimized to automate the entire molecular testing process, with simple on-screen instructions that limit user interaction to the minimum, so that just about sample loading and analysis initialization are required to obtain a result. It is clear that each closed software version is limited to a well defined assay (or group of assays), and therefore lacks the flexibility of a traditional laboratory instrument.

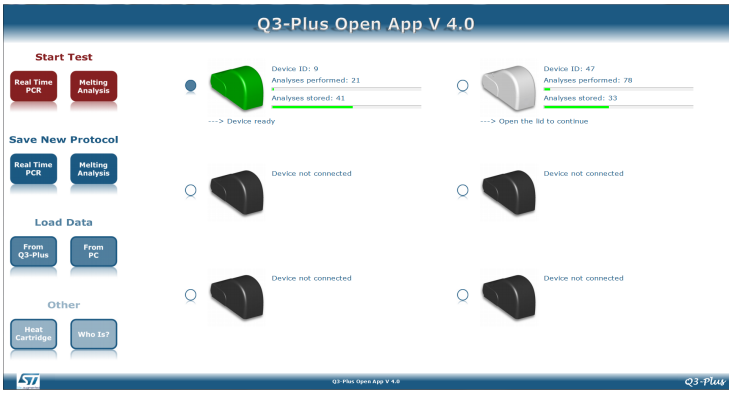


Figure 2.15: Q3-Plus Open App home screen.

Whichever the software version, it can control the performance of standard real-time PCRs as well as melting analyses of the amplified product, the fitting of raw data, C_t calculations and absolute quantifications of the initial concentration. The home screen of the latest open software version is shown in Fig. 2.15. It can manage up to six devices at the same time, which go through board and chip calibration (see Par. 2.1.2) as soon as they are connected and recognized.

From the menu on the left the user can load an old experiment from the embedded flash memory or define a completely new one. Fig. 2.16 shows how each parameter can be tweaked, in particular a channel can be selected or deselected and its operation fully customized. The *Optical Parameters* section in fact gives control over the relevant excitation parameters (such as exposure, analog gain and led power), the number of wells to consider in the analysis and the choice of a phosphor-specific matrix that is used during

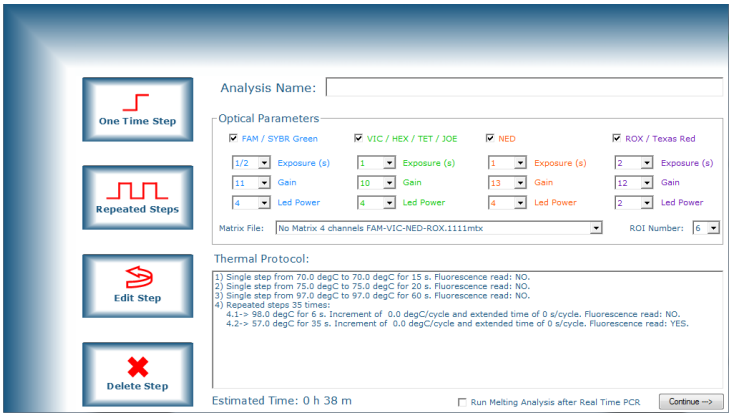


Figure 2.16: Q3-Plus Open App setup screen. All the relevant parameters for the definition of a new protocol can be modified here.

result analysis to take into account the coss-talk effects. The thermal protocol si also fully customizable, since temperature step values and durations need to be individually inserted before the procedure can go on. As a last step the user is led through a review of the chosen parameters channel by channel, with an interface, shown in Fig. 2.17, that also allows a fine tuning of the ROIs (position and size) where the software will perform the actual fluorescence analysis. While the experiment is performed the software shows the

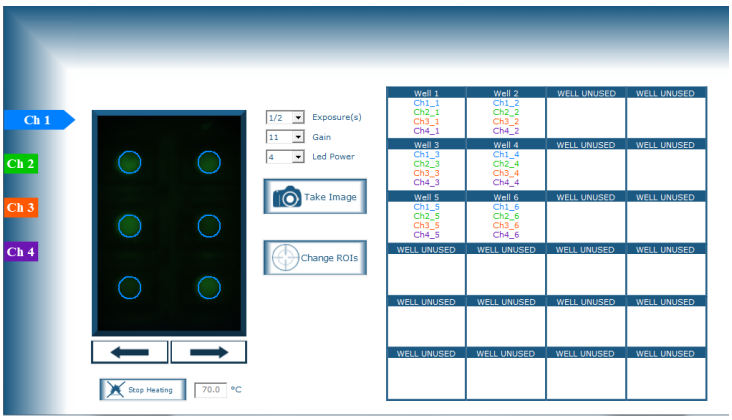


Figure 2.17: Q3-Plus Open App review screen. It sums up the optical parameters chosen for each channel and allows the optimization of the ROIs.

fluorescence reading in a real-time chart, and at the end of it a result screen is loaded as shown in Fig. 2.18. Here the user can work through data taking actions such as defining the fluorescence thresholds to calculate the C_t values, restricting visualization to just some of the channels, inspecting raw data, tuning the ROIs and more. Results of old saved experiments can also be loaded at any time from the main screen and reviewed in this environment.

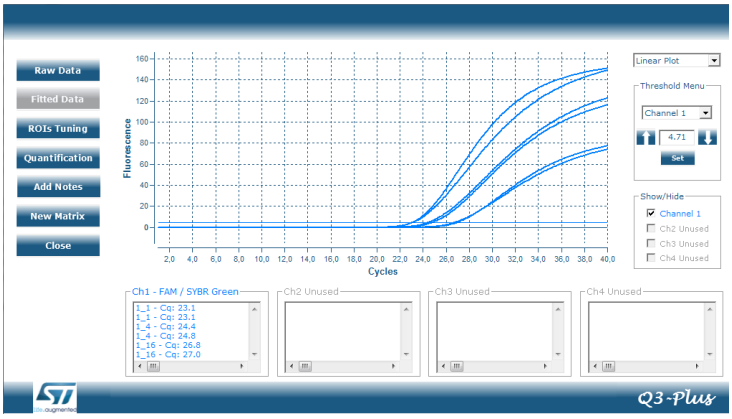


Figure 2.18: Q3-Plus Open App results screen. This real-time PCR was performed on a sample with two serial dilutions, 1:4 and 1:16. As expected, the C_t difference between them is about 2.

2.2 Excitation and detection architecture

The Q3-Plus device features four separate fluorescence reading channels, which means that in a standard 6-wells cartridge up to twenty-four distinct reactions can be performed simultaneously. Each channel is built around excitation/emission wavelength intervals specific to certain widely-used commercial labeling dyes, so the LED source is filtered to match the dyes absorption spectra as much as possible, and fluorescence is collected through a filter as close as possible to the emission peaks. Table 2.1 summarizes the available choices. The most frequently used dyes are, for each channel respectively, FAM, VIC, NED and ROX, therefore the rest of this discussion will refer to those. It is important to highlight that the choice of the NED for the fourth channel follows from a constraint on the available wavelengths range. In fact, since the module used for detection embeds in its lens system an IR filter, the design space is confined in the visible spectrum. Had it been possible to work in the IR, the fourth channel would have been chosen in that range.

Channel	Dyes	Emission range [nm]
1	FAM™/SYBR® Green/EvaGreen®	510-535
2	VIC/HEX™/TET™/JOE™/MAX™	540-560
3	NED™	560-580
4	ROX™/Texas Red®	610-635

Table 2.1: Commercially available labeling dyes compatible with the four Q3-Plus fluorescence channels.

A standard real-time PCR instrument is usually equipped with detectors and optics, specific to various channels, that are mechanically moved in correspondence of the sample

when emission from a specific fluorophore needs to be detected. Such a complicated solution does not find a reason to exist in a compact device like the Q3-Plus, and moreover the small size peculiar to the VW6558 CMOS camera modules makes it reasonable to reserve a separate detector for each of the four channels. Each module is therefore coupled with a fixed filter, designed to specifically detect the emission of a single fluorophore, and no moving parts are involved.

Both excitation and emission filters come from Semrock BrightLine fluorescence filters line, specifically designed for *"research and clinical microscopy, high-throughput screening, real-time PCR, and other fluorescence-based instrumentation"* [16]. Optical thin-film filters typically consist of multiple alternating layers of different refractive index dielectric materials deposited on a glass substrate. Thanks to the wavelength-dependent interference of light reflected or transmitted at the many interfaces between the layers, any desired bandpass behavior can be progressively approximated by playing with the number and thickness of the layers themselves. From a manufacturing point of view, the thin-film layer deposition on Semrock filters is made with the *ion-assisted ion-beam sputtering* (*IBS*) technology, which gives an high degree of control on the process uniformity and is based on hard oxide materials, that provide more durable coatings than traditional soft-coating approaches. Significant performance advantages such as steeper edges and flat stopbands are achieved by optimizing the structure using a combination of two edge filters (high-pass and low-pass) deposited each on one side of the glass substrate. This leads to coatings that are tens of microns thick with well over 100 dielectric layers [17].

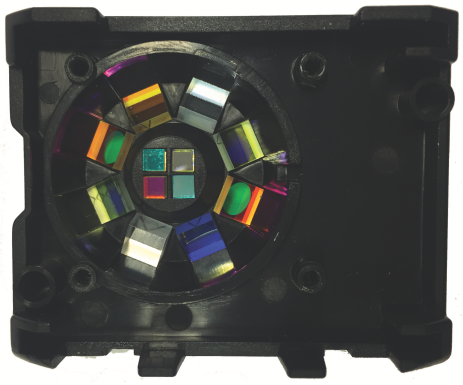


Figure 2.19: Filters sets inside the Q3-Plus lid. The smaller ones at the center of the holder are the emission filters. The others, notably tilted to match the orientation of the LEDs, are the four couples of the excitation set.

Optical Density (*OD*) is a quantity used to describe the transmission of light through an highly blocking optical filter, and is defined as $OD = -log_{10}(T)$, where T is the transmission of the filter. To guarantee a good spectral separation all filters have been selected with OD6 blocking, which means that in the blocking regions outside the pass-band

the transmission has to be 10^{-6} or lower. Fig. 2.19 shows a picture of the actual sets of excitation and emission filters inside their holder.

Fig. 2.20 depicts the parameters used by Semrock to define the spectral characteristics of a filter, which are used to form part of its name for a quick identification. The naming convention follows the template *XXyyy-CW/GMBW*, where XX is the product family, yyy is the revision, CW is the center wavelength and GMBW is the guaranteed minimum 90% bandwidth. The full width half maximum (FWHM) is then defined as $FWHM \approx GMBW + 0.01 \cdot CW$. In the case of the BrightLine filters described above the product family is FF, which stands for Fluorescence Filter.

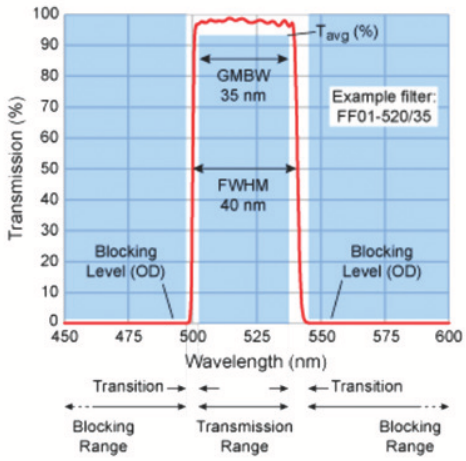


Figure 2.20: Semrock filters specifications diagram.

2.2.1 Excitation

Excitation of the labeling dyes is provided by means of four couples of LEDs, properly filtered to limit light emission in bands that match their absorption spectra. The choice of two LEDs per channel comes as an endeavor to illuminate the reaction wells as uniformly as possible, because the excitation light impinges diagonally on the chip, rather than orthogonally as is typical in commercial PCR machines. This is the result of the architectural trade-offs required to keep the Q3-Plus structure compact, with both excitation and detection happening in the small lid of the device. Fig. 2.21 shows a SolidWorks model of the crown of LEDs, where it can be appreciated how the two LEDs of each channel are placed in opposite positions, symmetrically around the center of the chip.

The LEDs belong to the XLamp XP-E and XLamp XP-E2 series by Cree, and come in a 3.45 mm by 3.45 mm package with a 115° viewing angle. The actual number of different

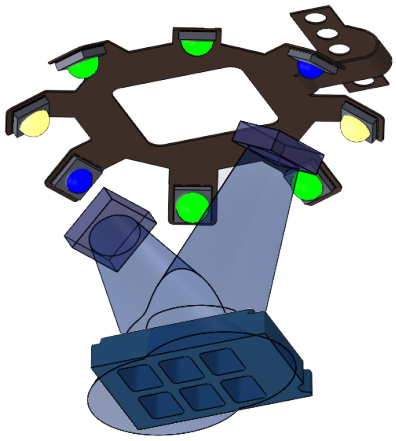


Figure 2.21: SolidWorks model of the LED crown, with the relative position of the chip below it. Each LED is behind a filter that defines the excitation band of a channel. The two gray cones simulate the light path from the filters to the chip.

colors employed is just three, as summarized in Table 2.2; this is because the second and third channel share the same LED but crop its spectrum in different spectral windows. In

Channel	Associated Fluorophore	Color	Dominant Wavelength [nm]
1	FAM	Blue	465-485
2, 3	VIC, NED	Green	520-535
4	ROX	Amber	585-595

Table 2.2: Cree LEDs and associated fluorescence channels.

the same way as the TCS controls heaters and fan (see Par. 2.1.2), the OCS employs an STMicroelectronics STM32F207VE microcontroller to generate a PWM signal deputed to switch the LEDs on and off.

Table 2.3 lists the filters chosen to select the excitation bands of the four channels. Each filter has an area of 6 mm by 6 mm and is 2 mm thick, except for the filter of the third channel which is 3.5 mm thick.

Fig. 2.22 shows the superposition of the LEDs spectra with the four theoretical bands of the emission filters. The darker areas in the picture point out the overlaps, which constitute the portions of spectrum that are actually allowed to reach the reaction wells. The shapes resulting from this superposition are those expected to be measured when sampling light inside a real instrument. The filters spectra are directly available from Semrock website, in the specification page of each product, while the LEDs have been characterized after the purchase because the manufacturer provides plotted emission curves

Channel	Associated Fluorophore	Semrock Filter
1	FAM	FF02-472/30
2	VIC	FF01-530/11
3	NED	FF01-556/20
4	ROX	FF01-550/88

Table 2.3: Q3-Plus excitation filters.

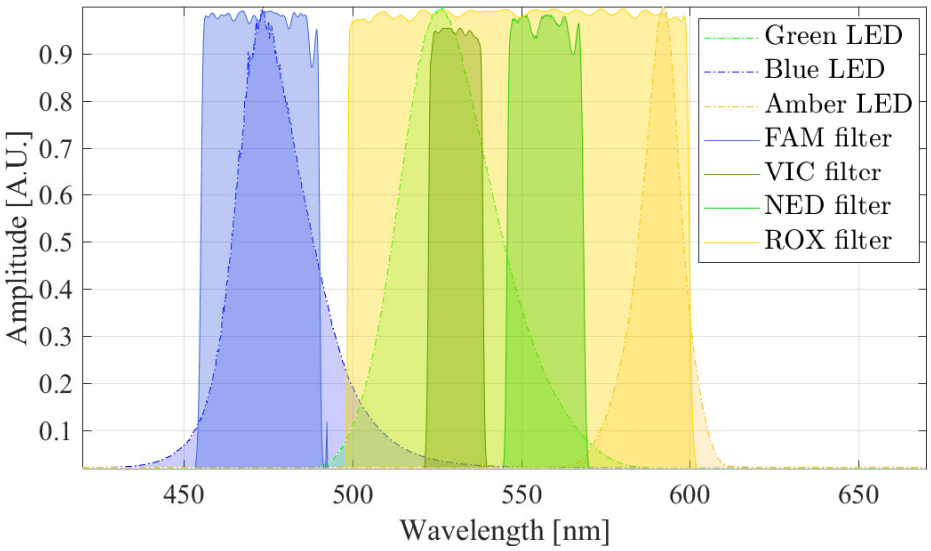


Figure 2.22: Excitation LEDs (dotted lines) and filters spectra (continuous lines).

but no numerical data.

Each fluorophore will be excited by the portion of light that reaches the wells and falls in its absorption spectrum. Fig. 2.23 shows FAM, VIC, NED and ROX absorption spectra as provided by their manufacturers. The spectra of light passing through the filters have been highlighted in gray, while the fluorophores are shown in color. Note that not only the design challenges of keeping four channels together do sometimes prevent the matching of exciting light with the absorption spectrum peak, but from the figure it is clear how some degree of cross-talk is inevitable, due to the fact that the fluorophores have broad spectra. This means that excitation meant for a certain channel will trigger a small response from at least one of the others, and to account for those contributions the collected data will go through a specific correction algorithm before the PCR results are displayed.

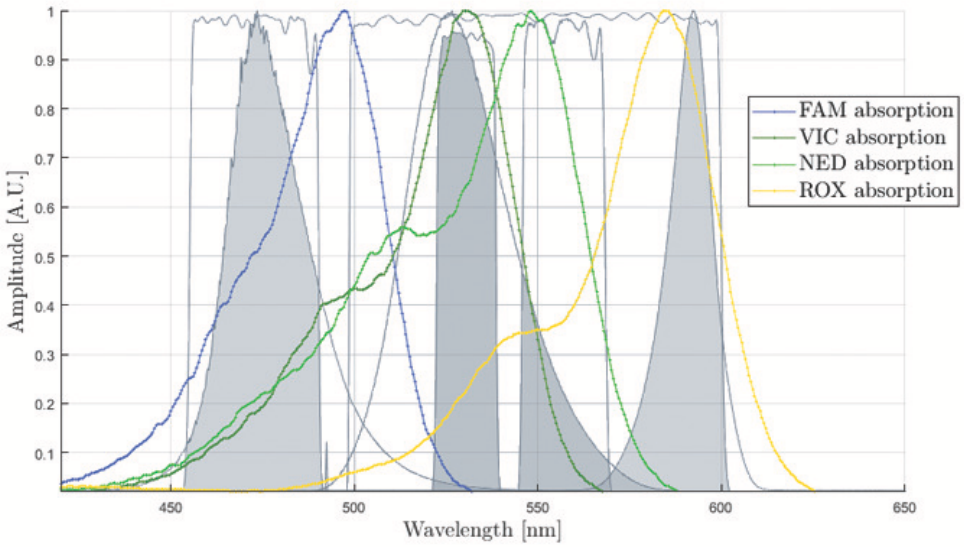


Figure 2.23: Excitation light reaching the wells (in gray) and fluorophores absorption spectra (in color).

2.2.2 Emission

Just like for the excitation section, choice of the emission filters passing bands is nontrivial, because of the trade-off between an higher amount of light collected in a fluorophore emission spectrum, granted by a wider band, and a reduced cross-talk between the channels, which is always present because of the large tails of the emission spectra. Table 2.4 lists the four emission filters, that, measuring just 3 mm by 3 mm for a 2 mm thickness, are smaller than the others because they need to cover just the camera areas. Here the NED filter actually belongs to a different family, called *EdgeBasic*, and the BLP code in the name stands for Basic Long Pass. Fig. 2.24 shows the transmission of the

Channel	Associated Fluorophore	Semrock Filter
1	FAM	FF01-524/24
2	VIC	FF01-556/20
3	NED	BLP01-568R
4	ROX	FF01-650/60

Table 2.4: Q3-Plus emission filters.

emission filters , superimposed with dotted lines are the fluorophores emission spectra. Note how the NED filter has an high-pass behavior, in contrast with all the others which are bandpass filters. This choice is a consequence of the modest performances of the NED dye, which requires a larger allowed bandwidth to compensate for the weak emission signal

it produces. From this representation it is quite clear how shifting the transmission bands closer to or farther from an emission peak can significantly alter the crosstalk contributions.

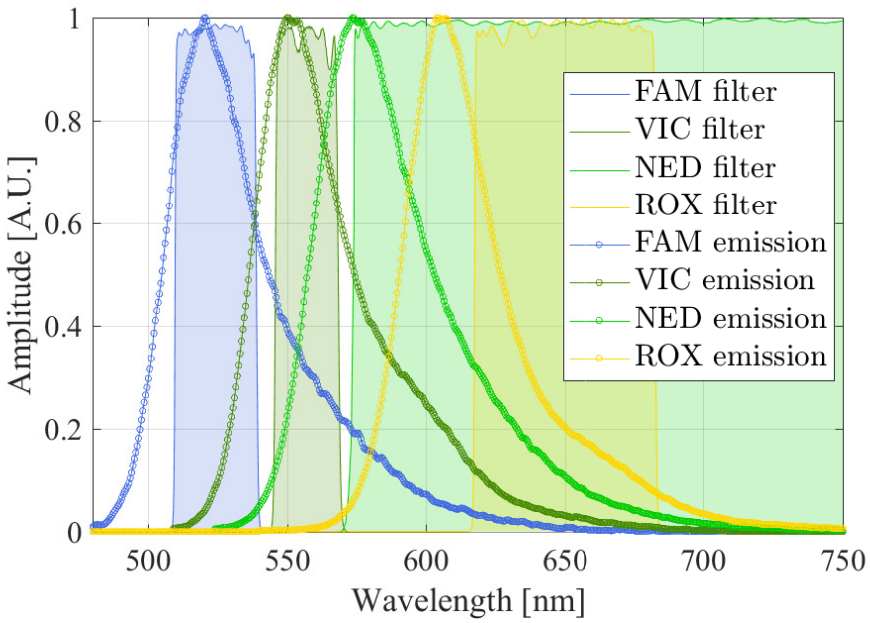


Figure 2.24: Emission filters (continuous lines) and fluorophores (dotted lines) spectra .

Chapter 3

Characterization of the existing device

In this chapter I am going to cover the characterization activity I have conducted on the Q3-Plus optical system. First I have focused my work on the excitation section, designing and 3D-printing a device to collect light inside the Q3-Plus wells. This provided an easy way to characterize both the excitation spectra and the sensitivity of the whole system to the fundamental parameters of a PCR protocol. Subsequently I addressed the study of the filters behavior, so I designed a new measurement setup that allowed the extraction of their transmission spectra as a function of the angle of incidence of light. This led to the proposal and implementation of some structural changes in the device.

3.1 Excitation light measurement setup

For an optimal performance of the Q3-Plus, it is fundamental that the spectral behavior of the light that reaches the reaction wells matches the theoretical design. The best way to verify this is to acquire the excitation signals exactly in the spots where the PCR reactions take place. To this end I developed the testing solution represented in Fig. 3.1. This consists of a casing, called CCC, with a modified holder containing a custom 3D-printed structure. The latter has the same size of the standard Q3 silicon chip and features seven holes in it, with an optic fiber passing through each one of them. The holes are tight and deep enough to hold in place seven branches of optical fiber, which I later fixed with a drop of optical glue as a precaution. Six of the holes are centered inside the wells, where the reaction volumes should be during normal operation of the device, while the seventh is placed on the point that has been chosen as reference during the original system design. This point should, therefore, reflect the ideal behavior of the system.

This structure fits into the cartridge plastic holder as a normal chip would. To carry the fibers out of the device I needed to open a way out that did not change the operating conditions of the system, so to this end I removed the push button which opens the lid of the Q3-Plus, located on the lower side of the front panel. The final result is shown in Fig.

Figure 3.1: CCC with the modified holder next to an unmodified Q3-Plus.

3.2, where it is also visible the patch cord connecting one of the fibers to a spectrometer. The link between fiber and patch cord is obtained with a simple 3D-printed connector I designed when initially approaching the light measurement task. This not only required a manual intervention each time I had to acquire a signal from a different position, but it also introduced some uncertainty in each measurement because of a slightly different signal coupling that necessarily originates from the repeated extraction and insertion of the fibers. This is the main reason that led me to conceive a new and more efficient experimental

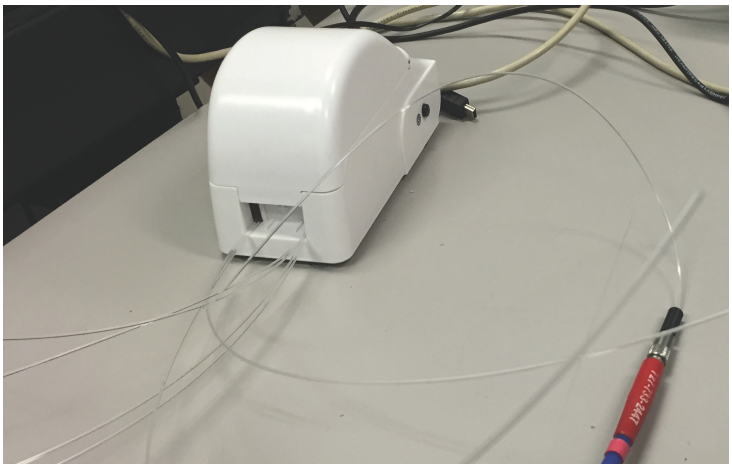


Figure 3.2: Closed Q3-Plus with the seven branches of fiber coming out of the opening button slot.

setup, which is schematically depicted in Fig. 3.3. Here, the presence of an optical switch makes the transition between a measurement channel (i.e. an optical fiber coming out of the Q3-Plus) and another as uniform and repeatable as possible.

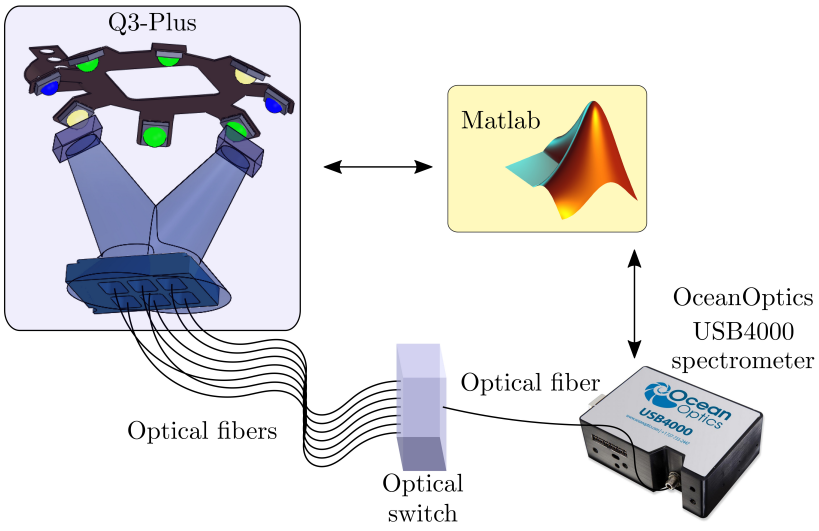


Figure 3.3: Conceptual scheme of the measurement setup I used to characterize the excitation section.

The system is designed around a Matlab script that I wrote to control at the same time both the Q3-Plus and an USB4000 spectrometer by OceanOptics.

When connected to a computer a Q3-Plus appears as a serial device on a COM port, and Matlab features specific functions to read from and write to such devices. At its first connection a Q3-Plus receives a unique COM port number, so that it appears, for example, as "COM3" in the device manager. Once this identifier is known, a serial port object can be created with the *serial* function, i.e. `s = serial(COM3)`. Then binary data can be written to (`fwrite` function) and read from (`fread` function) the device, essentially allowing every Q3-Plus command interface instruction to be executed. This gives, in particular, complete control over the LEDs and CMOS cameras behavior, which I needed during my tests.

OceanOptics USB4000 is a modular spectrometer responsive from 200 nm to 1100 nm that features a 16 bit, 3 MHz ADC and interfaces to a computer via USB 2.0 communication. The detector equipped in the instrument is an high-sensitivity 3648-element CCD array with a 1 MHz readout rate. The spectrometer features an optical resolution of 0.1 nm FWHM and integration times from 3.8 ms to 10 s. Wavelength calibration and linearity correction coefficients are stored in an on-board EEPROM. OceanOptics provides a java application programming interface (API) that allows integration of basic functions in the Matlab environment via COM object interface, so that it is possible to communicate with

and control the spectrometer without using the OceanOptics proprietary software.

My first attempt at the optical switch is shown in Fig. 3.4. It was designed to be

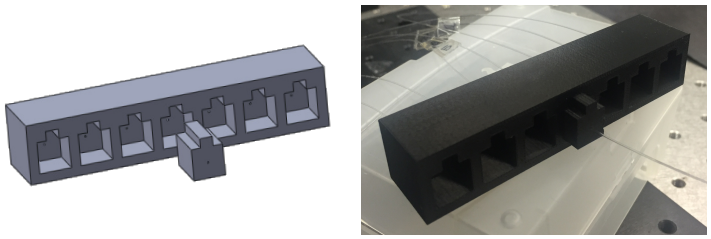


Figure 3.4: SolidWorks model and 3D-print of the automated optical switch.

perpetually connected to eight branches of fiber, seven aligned inside a long rack with holes and one placed inside a single plug. The seven fibers were those coming from the Q3-plus, and the last one went directly to the spectrometer patch cord input. The plug was obtained as the negative image of the hole that was mirrored to make the seven slots, so once all the fibers were properly inserted the alignment between facing branches was greatly enhanced with respect to the previous implementation.

While this solution worked, it still required a manual intervention each time a signal from a different position had to be acquired. To fully automatize the acquisition process then I designed another switch, which is built around a stepper motor and can be controlled with Matlab code in pretty much the same way as the Q3-Plus is. The CAD model and the printed pieces are shown in Fig. 3.5. The switch is composed of three parts: a base, a fixed fiber holder and a rotating cap. The base is designed to hold still the stepper motor

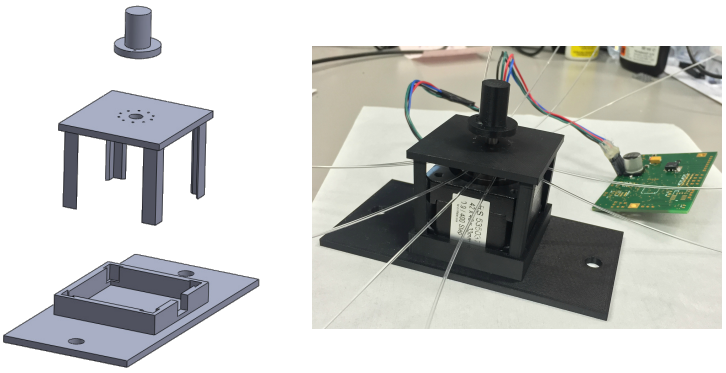


Figure 3.5: SolidWorks model and 3D-print of the manual optical switch.

and to keep its center aligned on an optical table. The latter is a vibration control platform that is used to support components for laser and optics related experiments in which the alignment of optical elements needs to remain stable over time. Its surface is therefore covered with a 25 mm grid of M6[1] holes where optical equipment can be mounted. I

[1] An M6 hole denotes it is designed for an ISO standard metric screw with nominal outer diameter of 6 mm.

designed the base so that its center, which corresponds to the motor shaft, is aligned to one line of holes when the setup is screwed to the optical table. While this is not useful in this setup, it will prove essential to align the components for the filters characterization. The fixed fiber holder has four legs, designed to wedge into the base when the motor is in place, and those sustain a platform. The latter has a central hole, where the motor shaft can go through and rotate without touching the platform, and ten fiber holes equally spaced over an angle of 360°. Lastly there is the rotating cap. It is designed to fit on the top of the motor shaft so it can move with it, and has two fiber holes. When the cap is pulled all the way down on the shaft it gets very close to the fixed platform, so that by controlling the motor a fiber lodged in the cap can be aligned to one of those in the platform, effectively coupling to one another.

The stepper motor is the 42SH33-4AM model from RS, and it is controlled by a L6470 microstepping motor driver by STMicroelectronics. It has a step angle of 0.9°, which means that it covers a full 360° circle in 400 steps, and can achieve a 1/128 step resolution thanks to the L6470 microstepping capabilities.

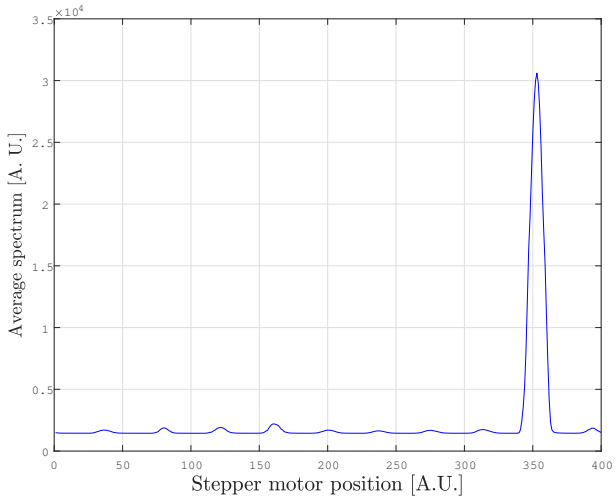


Figure 3.6: Example of result from a calibration of the optical switch. The peak shows there is an optimal position where the fibers are aligned.

Before initiating a new acquisition there is a self-calibration procedure to be run from a separate script, that is needed to set the motor on a meaningful home position. I wrote it as an half-way alternative between simple mechanical positioning solutions, such as reference notches, and more complex systems involving additional electronics. When the procedure starts, one of the fibers which are not inside the Q3-Plus needs to be illuminated with a source of light. Normal ambient illumination proved to be enough, just keeping the free end of the fiber oriented towards the ceiling lights gives a strong enough signal to complete the calibration. Then what the script does is move the motor with the smallest

step possible along the full 360° circle, ending back where it started. At each step the spectrometer acquires the spectrum and saves the data in a matrix. After all the positions have been scanned, a simple mean is taken for the data at each step and the one with the highest value gives the position of the fiber that is receiving light. Then the motor is moved back there, and it is saved as home position, so that a reference is fixed. Fig. 3.6 shows an example of the calibration script output.

3.2 Excitation light measurement results

The setup described in the previous section was used to gather information about the excitation light that can be provided by the optical system of the Q3-plus. Fig. 3.7 shows the labels given to every point, with the color legend which will be kept the same in the remainder of the chapter. For the acquisitions I used a plastic 1 mm multimode step-index

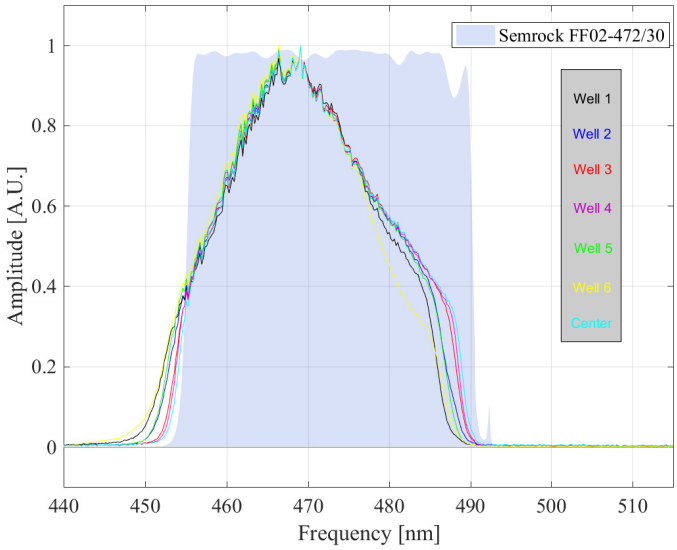
Figure 3.7: Labels for the seven acquisition points on the 3D-printed chip. Each one is associated to a color that is used to plot the acquired spectra in the following figures.

PJR-FB1000 fiber by Toray, with a PMMA core ($n = 1.49$) and a fluorinated polymer cladding ($n = 1.35$). They also feature a numerical aperture of 0.63, an angle of incidence of 78° and less than 0.2 dB · m attenuation at 650 nm. Each acquisition has always been preceded by the measurement and storage of a reference background spectrum, with all of the LEDs turned off, that is subtracted from the acquired data during the subsequent elaboration.

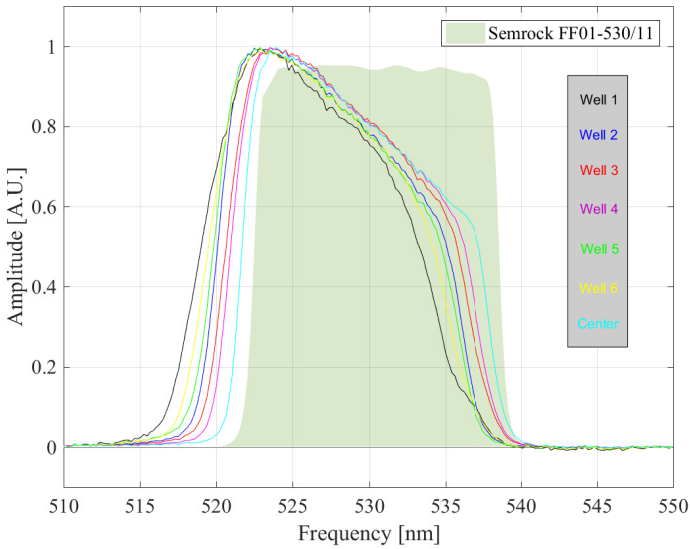
3.2.1 Excitation spectra

Fig. 3.8 and 3.9 show the normalized excitation spectra in the seven positions on the chip, divided by channel. The theoretical filter transmission of each channel is also

plotted with the experimental data, to give a reference of the expected positions of the passband cuts. Each curve is obtained as the average of six consecutive acquisitions, with an integration time of 60 ms, background noise subtraction and the LED power parameter set at the minimum possible. The curves labeled with "center" in each plot represent the



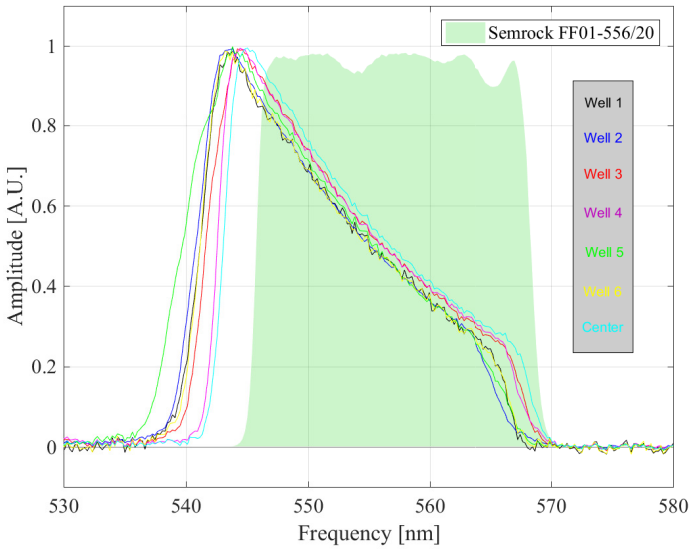
(a) Channel 1



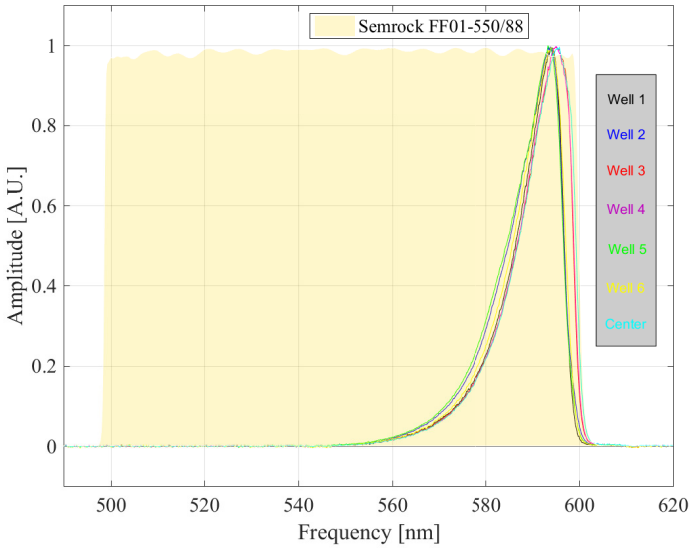
(b) Channel 2

Figure 3.8: Excitation spectra from seven points of the chip for the first and second fluorescence windows.

measurement of what should be the ideal behavior of the system, and it can be appreciated how those actually happen to always be the farthest curves on the right. The central point should be receiving the light that impinged orthogonally on the filter, while the other



(a) Channel 3



(b) Channel 4

Figure 3.9: Excitation spectra from seven points of the chip for the third and fourth fluorescence windows.

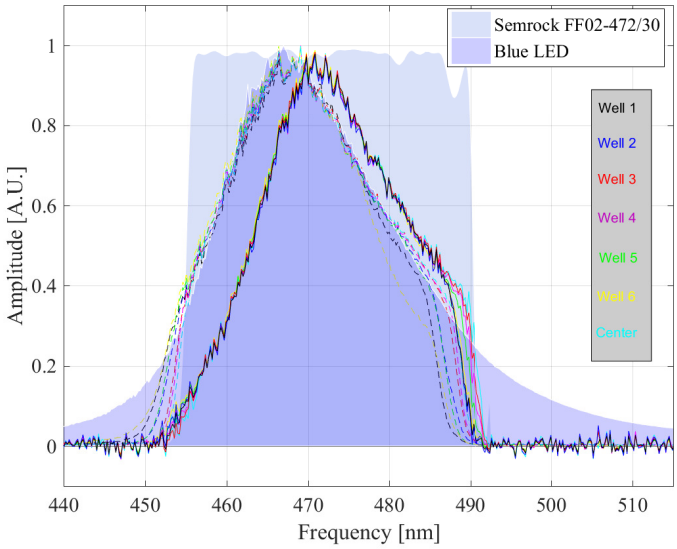
positions would receive rays that were skewed with respect to that ideal path. Since it is known that the bandwidth of a filter shifts towards shorter wavelengths as the angle between incident light and filter surface deviates from 90° (see Par. 3.3 for more details), this is an expected behavior. Moreover, the two positions closer to the center (numbers 3 and 4 in the figures), which receive less skewed light rays, always produce the first two curves after the ideal one. It should also be noted that the maximum spread between the central spectrum and any other curve at an amplitude reference value of 0.2 is no greater

than 4 nm for all the channels, which agrees with the angular deviation of the light that hits the points from 1 to 6 with respect to the center, which goes from 4° to about 8° by design.

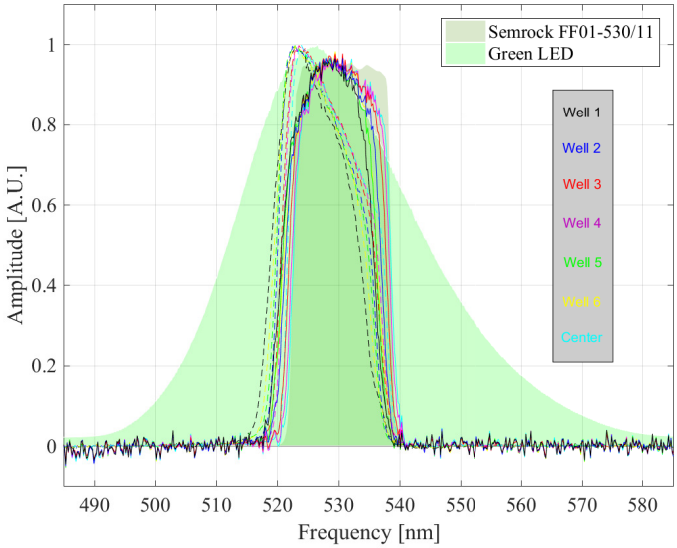
What was not expected, however, is the strong discordance between the acquired spectra and the cutting wavelengths of the filters, with particular reference to the second and third channel where the curves fall as far as 6 nm out of the passing band.

A simple explanation could be a shift in the frequencies measured by the spectrometer, due to aging or wear; this was easily disproved by repeating the tests with another device, which led to similar results. A likely cause for this kind of behavior could also be the presence of out-of-specification tolerances on the filters bandwidth. This hypothesis led to an in-depth analysis of the filters behavior, which will be described in detail in Par. 3.3 and 3.4.

The last possible explanation was an inaccurate manufacturing of the plastic components of the Q3-Plus: this thought originated considering that such a shift of the spectra could also be related to an unwanted tilt between the LEDs and the filters, something that was not present in the design but that could nevertheless have shown up in the final products. The fastest way to check if this conjecture had some basis was to 3D-print the original design and repeat the measurements. The comparisons between the data acquired with the molded pieces (dashed lines) and with the 3D-Printed version (continuous lines) for all the channels is plotted in Fig. 3.10 and 3.11. One can observe non-negligible differences in the passband position and shape. Specifically, the shift ranges from 3 nm to about 11 nm in the worst case, and reflects a difference of the relative inclination of the LED mounting plane with respect to the surface of the filters.

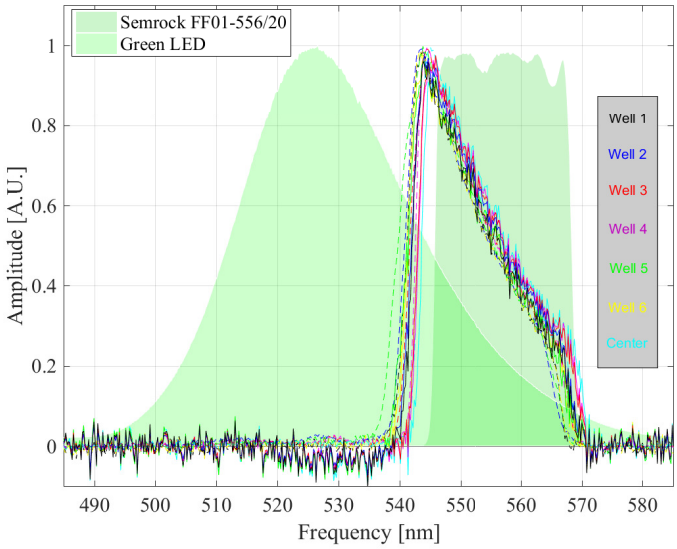


(a) Channel 1

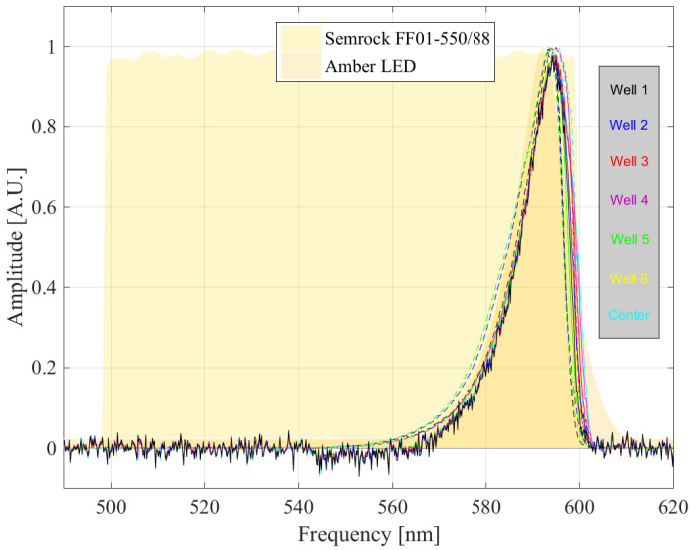


(b) Channel 2

Figure 3.10: Comparison between spectra acquired from manufactured (dashed lines) and 3D-printed (continuous lines) devices for the first and second fluorescence windows.



(a) Channel 3



(b) Channel 4

Figure 3.11: Comparison between spectra acquired from manufactured (dashed lines) and 3D-printed (continuous lines) devices for the third and fourth fluorescence windows.

3.2.2 Sensitivity

Given the automation level reached by the acquisition setup, it was possible to devise a series of experiments to collect a wide amount of data to infer the dependence of the fluorescence value detected on the control parameters that define an analysis protocol. The goal of this tests was to define a sensitivity $\mathcal{S}$ (Eq. 3.1), that, multiplied by the acquired fluorescence[2] $\mathcal{F}$, could give a normalized value $\mathcal{N}$ which is only a function of the fluorescence level and not of the specific parameters, that is:

$$\mathcal{N} = \mathcal{S}(FR,\ LP,\ AG) \cdot \mathcal{F}. \tag{3.1}$$

Those parameters are frame rate (FR), led power (LP) and sensor analog gain (AG), and they can assume 7, 10 and 16 different values, respectively. The hypothesis is that $\mathcal{S}$ has a linear dependence on them, so that the sensitivity to any combination i of the three parameters can be expressed as

$$\mathcal{S}_i = f_{FR}(FR_i) \cdot f_{LP}(LP_i) \cdot f_{AG}(AG_i). \tag{3.2}$$

To simulate the fluorescence emissions that happen in normal operation I cut some notches on the ring surface of an empty chip, and deposited a mix of fluorescent dyes to have a luminous signal in each channel. Fig. 3.12 shows the image of a test chip with various dye concentrations for the first channel. The acquisition process is driven by a

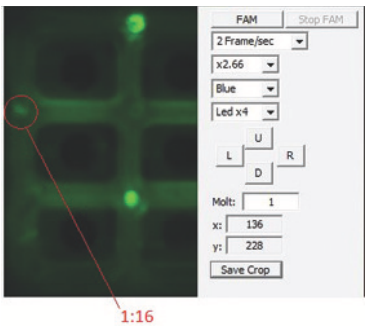


Figure 3.12: Chip with fluorescent dye spots emitting in the first channel. This is an example of a test to choose the best concentration to have a good signal for the acquisitions.

Matlab script I wrote to explore subsystems of the Q3-Plus parameters three-dimensional space. To ensure dependencies from temporal drifts due to the prolonged operation of the device are separated from the linear progression of the parameters values, a randomization algorithm calculates all the possible ($FR,\ LP,\ AG$) triplets and scatters a number of

[2]Here fluorescence is calculated as the mean value of the pixels of an image.

repetitions of each one in a reference matrix. In each step the script programs a different triplet in the device and acquires a fluorescence image. The saving process reverses the randomization, so that the data is stored in a precise location in a multi-dimensional matrix, ordered by channel, parameter triplet and repetition number.

Each experiment was preceded by a one-time analysis run by a simple algorithm, which took and elaborated dark pictures. I did this to map the hot pixels in the CMOS sensors of the particular Q3-Plus that was being tested, in order to later exclude them from elaboration. To initiate the operation the acquisition script needs a starting point, which is a certain triplet that can be different for each channel, and three variation intervals which will define the number of permitted values that the parameters will assume above or below the starting points. The length of a single experiment could go from 16 hours to about two days, even though is important to point out that a good portion of that time is spent transferring and saving data.

The acquired data are then averaged between repetitions of the same sets of parameters, and can be elaborated to extract the dependencies of the result from one of the variables. Fig. 3.13 shows the data after elaboration, subdivided by parameter and channel. Each plot contains the mean contribution of those sets of parameters that did not deliver saturated or dark images, but that could effectively represent operational conditions. Each

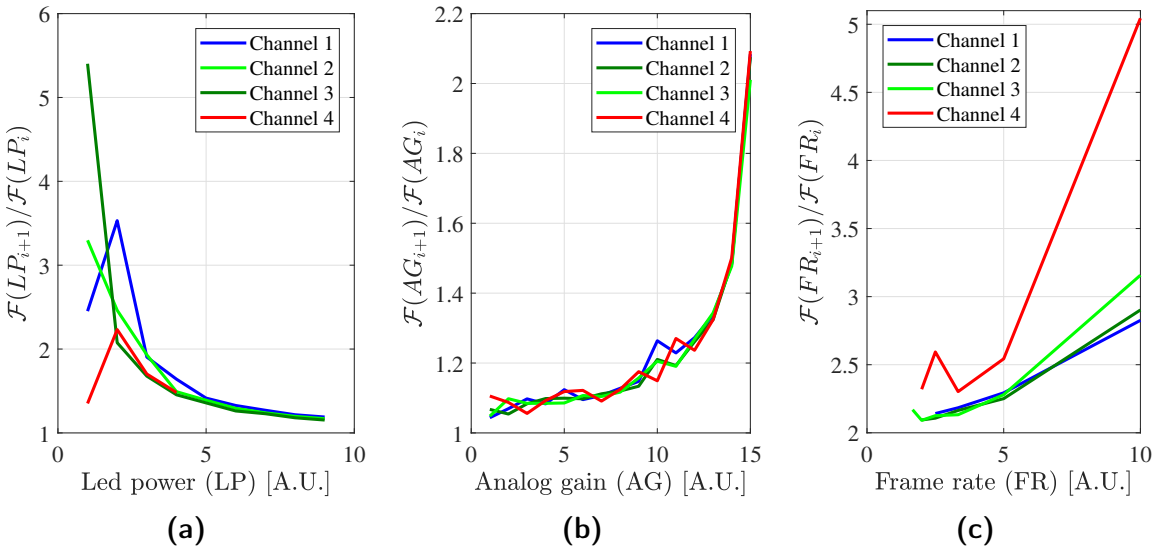


Figure 3.13: Dependence of the fluorescence in the four channels as a function of led power, analog gain and frame rate. Each point represents the ratio of the mean fluorescence for two consecutive values of a parameter, with the other two fixed.

point of a curve represents the ratio of the mean fluorescence for two consecutive values of the parameter that is being considered. Such a representation is meant to highlight the sensitivity of the fluorescence to the examined parameter, i.e. the amount of fluorescence

change corresponding to a change in that parameter. This analysis showed that it is indeed possible to define a sensitivity with three separate dependencies on the acquisition control parameters, and the numerical results have been included in the data elaborations of the real-time PCR analyses. Furthermore, this kind of information is useful to identify equivalent sets of parameters corresponding to a certain fluorescence signal. As a matter of fact, while designing a real-time PCR assay, noise and timing considerations could have different relevance, and being able to obtain the same result with different sets of parameters constitutes a powerful optimization tool.

3.3 Filters characterization setup

To individually characterize the transmission of each filter I composed a setup with a light source, an iris to limit the size of the light beam, three holders to manage the filters different sizes and a spectrometer. All these components need to be carefully aligned on the optical table with post holders and mounts. Fig. 3.14b shows the practical implementation, with a filter inserted in its holder. I used Thorlabs 150 W High-Intensity Halogen OSL1-EC

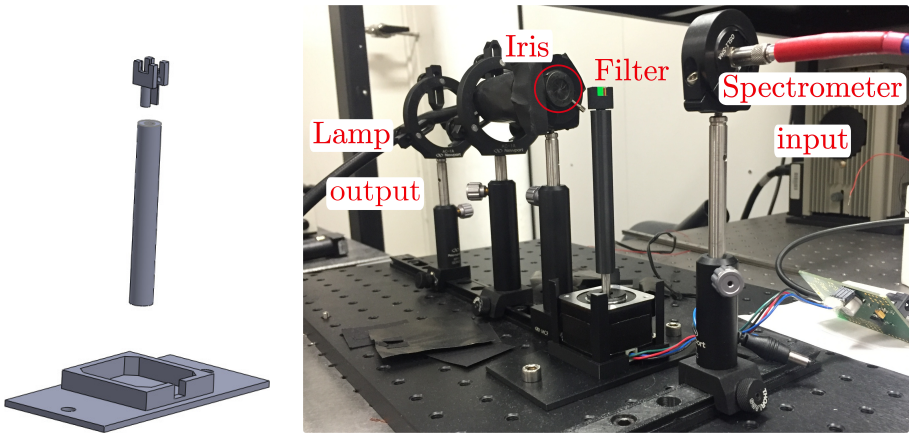


Figure 3.14: CAD model of the filter holder structure (a) and experimental setup for the characterization of filters (b).

Light Source to illuminate the filters through a ∅5 mm fiber bundle. I kept the stepper motor and its base from the setup described in Par. 3.1 and designed a simple system to hold the filters on the rotating shaft: as Fig. 3.14a shows, it is a cylinder that fits on the shaft and has interchangeable casings the size of the various filters. The lamp fiber is aligned with a couple of post holders with Newport AC-1A Universal Fixed Lens Mounts, and to keep its light confined onto the iris surface I encapsulated the end of the fiber in an aluminum cylinder painted in black. Except for the motor, the other components are

mounted on two orthogonal dovetail optical rails. During the acquisitions I also kept a rectangular aluminum shield with a small opening placed on the filter holder, towards the light beam, to limit any sort of extra light reaching the spectrometer input fiber.

The stepper motor allowed me to characterize the filters transmission as a function of the angle of incidence (AOI) of light. The approximative position of the filter holder that would correspond to orthogonal incidence could be found by substituting the filter with a pinhole: by measuring the light passing across a number of angles, I defined the home position as the one with the highest mean signal value. Fig. 3.15 shows an example of the calibration script output. To complete the setup preparation a measurement of the

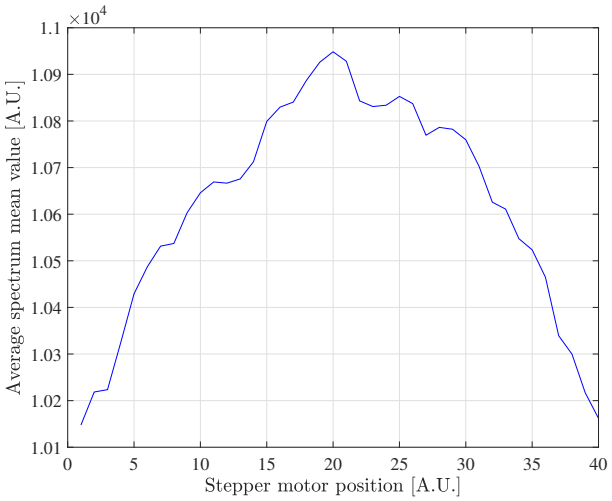


Figure 3.15: Graphical result of the calibration analysis to find the stepper motor position that would best place the surface of a filter orthogonal with respect to the incident light. The script scans 40 0.9° steps and for each one averages the spectrum acquired through a pinhole mounted on the filter holder.

Thorlabs lamp spectrum and one of the background light need to be manually performed and stored, to be used in the subsequent elaborations. After a filter is placed in the holder the characterization can be completely controlled by a Matlab script. The following results were obtained setting the integration time of the spectrometer to 100 ms, and each spectrum was calculated as the average of 10 consecutive acquisitions. A single graph shows a scan of 30 positions, with steps of 0.9°, half for 15 angles before the home position and half for the other side. Therefore the steps covered a range of $\pm 0.9° \cdot 15 = \pm 13.5°$. The transmission has been calculated as in Eq. 3.3, where $S(\lambda)$ is the sampled spectrum and $D(\lambda)$ and $R(\lambda)$ are the dark signal (background light) and lamp reference spectrum initially acquired.

$$T(\lambda) = \frac{S(\lambda) - D(\lambda)}{R(\lambda) - D(\lambda)} \tag{3.3}$$

3.4 Filters characterization results

As the light AOI is changed from normal to the plane of the filters to more oblique angles, two changes are expected in the filter transmission due to the increasingly longer geometrical paths traveled by the light rays. The first and immediately evident one is a shift of the spectrum towards shorter wavelengths. This follows from the fact that there is a reduction in phase difference that occurs at increasingly oblique angles between transmitted and reflected waves that travel through the multiple coating interfaces. As the interferences of those waves define the wavelength regions where the light is blocked, different AOI will produce different cuts in the spectrum. Given that the passband is the result of the union of high-pass and low-pass multidielectric layers, it is also expected to see a different behavior in the cutting bands shift because the two components are separately and differently designed. The central wavelength of the filter should then shift with the AOI as described in Eq. 3.4:

$$\lambda = \lambda_0 \sqrt{1 - \frac{\sin^2(\alpha)}{n_{eff}^2}} \tag{3.4}$$

where λ_0 is the unshifted wavelength at normal incidence, α is the AOI and n_{eff} is the effective refractive index specific of the filter.

The second consequence of the growing angle of incidence of light is that the behavior of the s and p components of the traveling electric field, which are respectively perpendicular and parallel to the plane in which the rays move, will separate. The effective index of a thin film is defined for the s component as $n_s = n\cos(\alpha)$ and for the p component as $n_p = \frac{n}{\cos(\alpha)}$ [18], where n is the refractive index of the material. As the AOI increases the polarizations will diverge, and this will be reflected in the final performance which is given by the average of the two spectra. For this kind of filters [19], the transition from the stopband to the passband should show a step around $T = 0.5$. This behavior becomes relevant for incidence angles greater than 30°, therefore in the next graphs it is expected to produce just a barely appreciable deformation of the cutting bands only in the first and last spectra of any group.

A third effect of the AOI shift would be a decrease of the transmission band value, but also this behavior can be noticed only for strong deviations from normal incidence and is not expected to play a relevant role in those results.

Looking at the results, as the angular position proceeds from −13.5° to 0° and the angle changes sign, there is an expected progression of the spectra from left to right and back. This is highlighted with a gradual color change in the curves of the plots, which go from pink to blue as the acquisition steps proceed. For each group of plots are also reported

Specification	Channel 1	Channel 2	Channel 3	Channel 4
Transmission band	$T_{avg} > 93\%$ $457 - 487$nm	$T_{avg} > 90\%$ $524.5 - 535.5$nm	$T_{avg} > 93\%$ $546 - 566$nm	$T_{avg} > 92\%$ $506 - 594$nm
Center Wavelength [nm]	472	530	556	550
Guaranteed Minimum BW [nm]	30	11	20	88
FWHM BW [nm]	34.6	16.5	25	100

Table 3.1: Excitation filters optical specifications.

the FWHM and 93%, 95%, 90%, 92% band-limit values as a function of the angular step. Those quantities are used in Semrock datasheets to define filters specifications. The FWHM hes been calculated as the distance between the wavelengths where transmission is 50% of the average passband value, evaluated on the plateau region far enough from the transitions. The 93%, 95%, 90%, 92% bandwidth values follow the same logic.

Table 3.1 gathers the excitation filters specifications as given by the manufacturer.

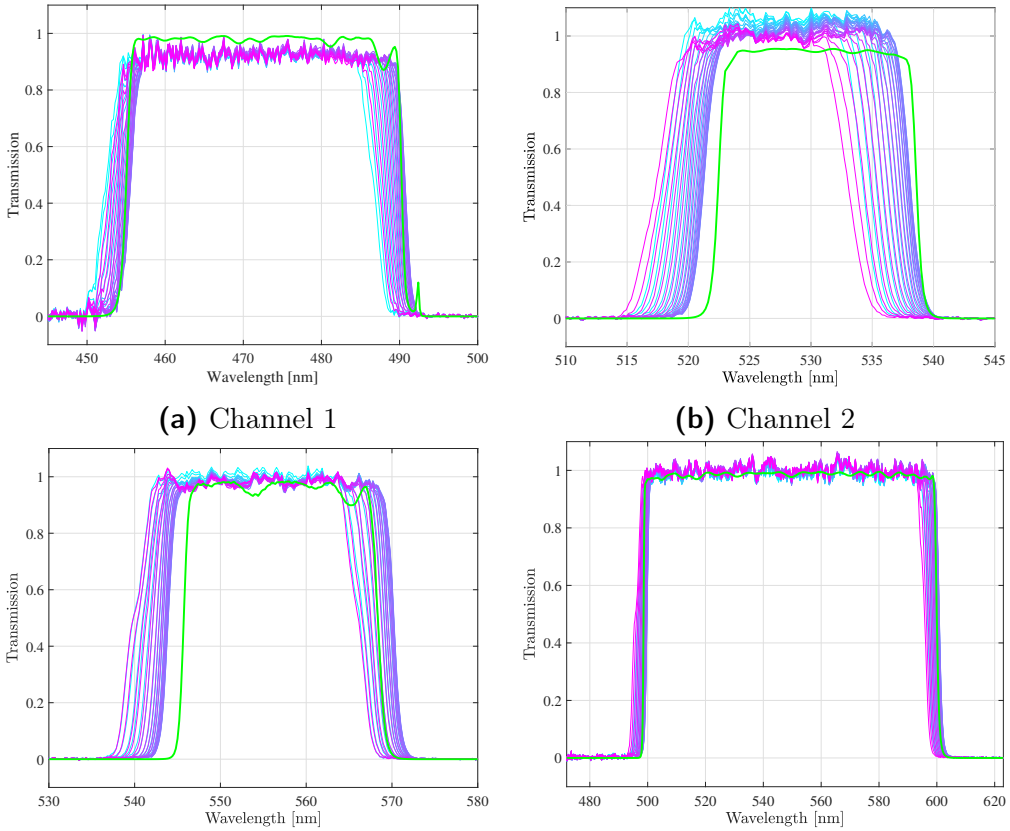
Fig. 3.16 reports the results of the excitation data acquisitions ordered by channel. Note that the farthest spectrum on the right for each plot represents the filter transmission for orthogonally impinging light, therefore if this curve overlaps the theoretical transmission it means that the filter meets the manufacturer's specifications.

Table 3.2 gathers the emission filters specifications as given by the manufacturer. Fig. 3.17 reports the results of the emission data acquisitions ordered by channel.

Specification	Channel 1	Channel 2	Channel 3	Channel 4
Transmission band	$T_{avg} > 93\%$ $510 - 534$nm	$T_{avg} > 93\%$ $546 - 566$nm	$T_{avg} > 93\%$ $584.1 - 900$nm	$T_{avg} > 95\%$ $620 - 680$nm
Center Wavelength [nm]	522	556	579	650
Guaranteed Minimum BW [nm]	24	20	/	60
FWHM BW [nm]	29.6	25	/	66.5

Table 3.2: Emission filters optical specifications.

Elaboration of the numerical data extracted from those acquisitions allowed me to



(a) Channel 1 **(b)** Channel 2

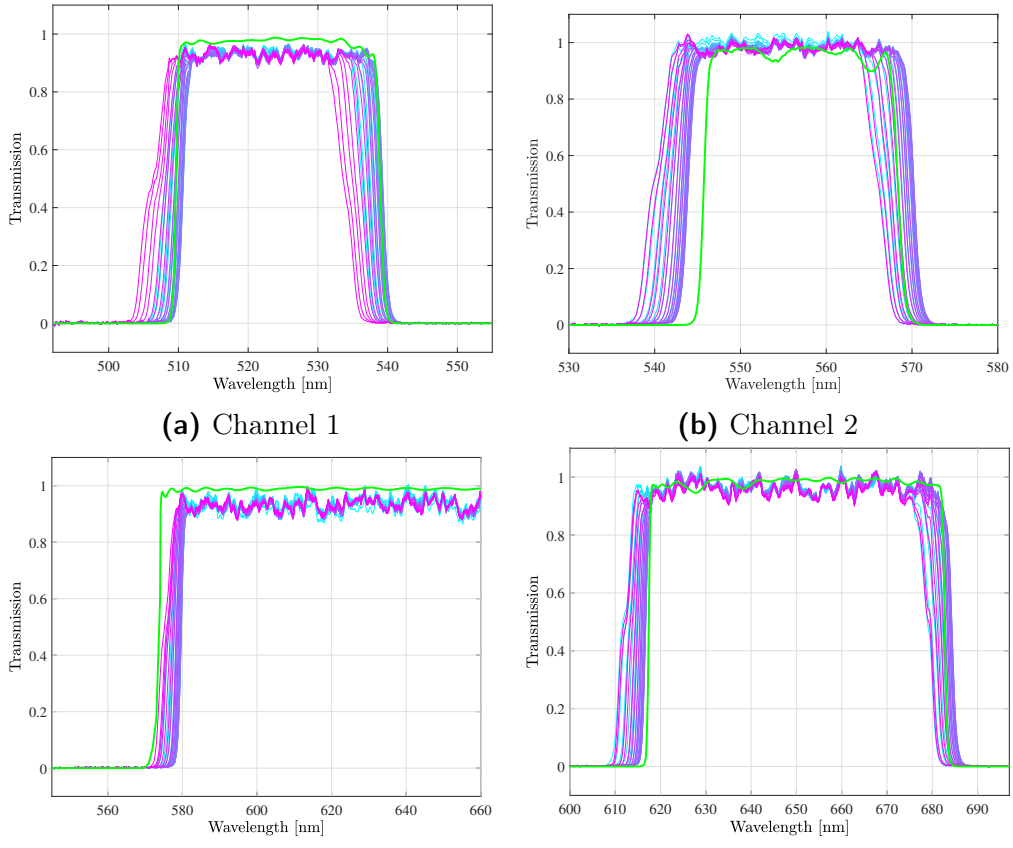
(c) Channel 3 **(d)** Channel 4

Figure 3.16: Acquisitions of the excitation filters transmission as a function of the AOI of light. The green curve plots the datasheet theoretical transmission.

verify the adherence of the filters performances with respect to the datasheet specifications. Fig. 3.18 and Fig. 3.19 report the data extracted for the excitation and emission filters, respectively. Since the filter in Channel 3 is an high-pass filter, Fig. 3.19c shows its 93% high-pass cut.

Those results demonstrate that, in fact, part of the unexpected behavior of the spectra acquired inside the Q3-Plus is due to a certain discrepancy between the datasheet specifications and the effective filter performance in the second and third channel. Repeating the experiments with a number of nominally identical filters has given the same results. Also the acquired data show a good adherence to the theoretical trend of the central wavelength shift described in Eq. 3.4.

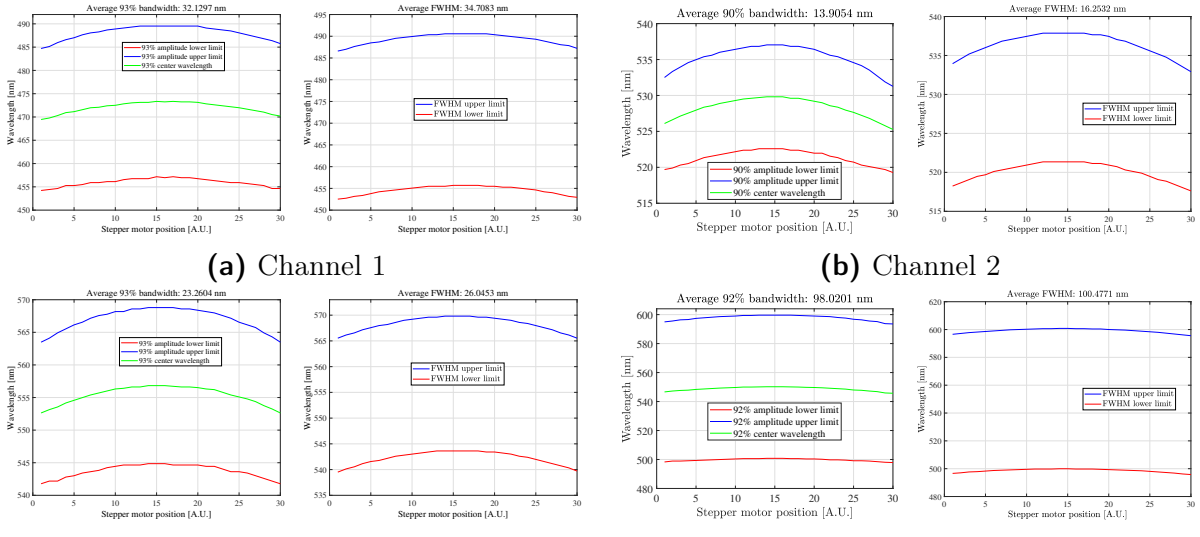
Fig. 3.20a shows in detail the comparison between theoretical and experimental data for FF01-650/60 filter, while Fig. 3.20b collects the same results for all the other filters which had the effective refractive index information available in the datasheet.



(a) Channel 1 **(b)** Channel 2

(c) Channel 3 **(d)** Channel 4

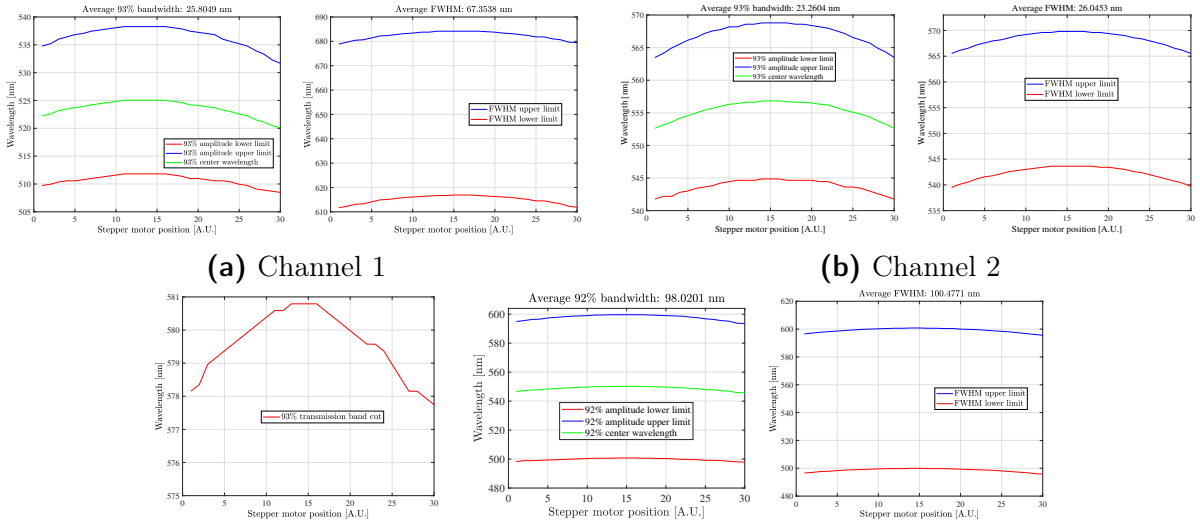
Figure 3.17: Acquisitions of the emission filters transmission as a function of the AOI of light. The green curve plots the datasheet theoretical transmission.



(a) Channel 1 **(b)** Channel 2

(c) Channel 3 **(d)** Channel 4

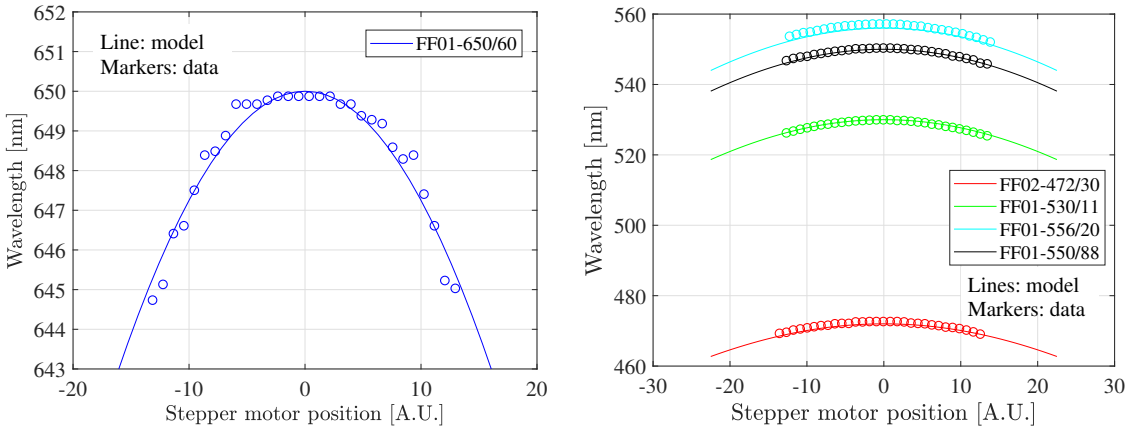
Figure 3.18: Excitation filters 9x% (a) and FWHM (b) band-limits as a function of the AOI of light.



(a) Channel 1 **(b)** Channel 2

(c) Channel 3 **(d)** Channel 4

Figure 3.19: Emission filters 9x% (a) and FWHM (b) band-limits as a function of the AOI of light.

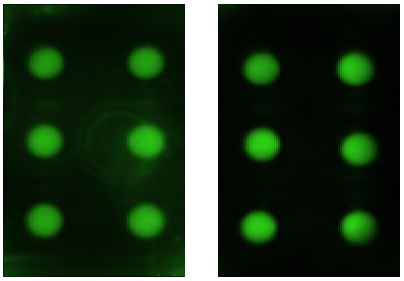


(a) **(b)**

Figure 3.20: Comparison between theoretical and experimental central wavelength trends for the FF01-650/60 filter (a) and the excitation filters set (b).

3.5 Optical system changes for optimized performances

One of the main problems in the original optical system performance was a certain amount of stray light that would always show up in the fluorescence acquired images, with a different weight in each channel. This had a negative impact on the final analysis, since it added noise to the readings, which is particularly relevant for weakly fluorescent samples. Fig. 3.21a gives an example of a strong light interference in the second channel.



(a) (b)

Figure 3.21: Comparison of the fluorescence detected from the same sample in the third channel, with the industrially manufactured filter holder (a) and with the 3D-printed one (b).

The results described in Par. 3.2.1 showed that, once the filter holder was correctly reprinted, the problem of stray light was almost completely gone, and this reopened a previously closed path toward cost optimization. During my first tests, in fact, I tried to see if the system would work in an acceptable way if one of the two LEDs of each channel was obscured. The answer was a straightforward no, the performances would degrade to an undesirable level. This was probably due to the additional crosstalk between the passing bands of the excitation windows, caused by the incorrect positioning of the filters with respect to the LEDs, which let in more light than expected and added extra internal reflections. The new condition allowed me to successfully repeat the tests with half the illumination. When obscuring one LED of a couple, I was able to show that the same fluorescence signal that was detected before could be obtained by appropriately tweaking the acquisition parameters. After I extracted a suitable set of alternative parameters for each channel, it was decided to proceed with the halving of the excitation components. This meant that not only four of the eight LEDs could be removed, but also the relative excitation filters, which are the most expensive components of the system.

The previous system also coped with stray light by adding laser-cut masks to reduce the light passing through the emission filters. The need to reprint the filter holders with the correct design features also gave me the chance to try multiple shapes and sizes for the opening windows. In particular, while the emission apertures remained rectangular and

substantially unchanged, the excitation ones went through a trial-and-error optimization procedure to adjust their shape and tilt. Since half of the excitation light was being removed, it was crucial to be sure to obtain the most uniform illumination possible. Fig. 3.22 shows the result of one wrong design: the shadow of the filter border is visible in the upper and lower wells, while only the central ones are properly lit.

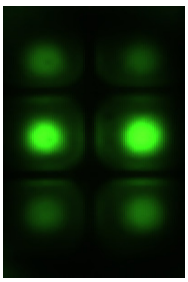


Figure 3.22: Fluorescence detection in the third channel with an incorrectly-sized aperture for the excitation light.

Chapter 4

Looking forward: a system with five wavelengths

As the Q3-Plus reached a certain maturity through interaction and collaborations with public healthcare institutions and medical companies, ideas for a number of new features surfaced over time. When the CMOS camera modules obsolescence prompted the need to find a suitable substitute, the Q3-Plus developing team started to design a whole new version of the device that would implement the new features. This process required the design of a custom lens system for a new sensor. In this chapter, after a description of the new system, I will introduce the Two-Photon Polymerization (2PP) technology I used for the lens fabrication. Then I will focus on the design, objective characterization and fabrication steps for the custom lens.

4.1 The new system

Once the Q3-Plus system proved to be able to successfully manage four wavelengths, the natural evolution was to think of a fifth one, which would bring the total number of possible parallel analyses for six wells from twenty-four to thirty. A limitation to the variety of the available fluorescence windows descended directly from the choice of CMOS camera modules as detectors: those sensors come from the mobile electronics world, so they need to be small, which is the feature that ultimately got them into this system, but their primary function is to take good pictures. Since silicon photodetectors are highly sensitive to near-IR (NIR) light, to accurately reconstruct a color image they need to include an IR cut-off filter, which is usually embedded in the lens system above the sensor

or deposited on the sensor surface.

Cyanines are commonly used dyes for fluorescence detection that work exactly in the near-IR window, and are often preferred to other dyes because of their brighter and more stable fluorescence. From this family the *CY5* and *CY5.5* dyes, that have absorption and emission peak wavelengths summarized in Table 4.1, were chosen for two new channels. The NED dye was instead dropped.

Dye	Absorption peak [nm]	Emission peak [nm]
CY5	650	670
CY5.5	678	694

Table 4.1: Absorption and emission peaks of CY and CY5.5 dyes. Both work in the 650 nm-1350 nm NIR range.

4.1.1 New filters

The filters for the new channels belong to the BrightLine family from Semrock. The excitation/emission sets that are currently employed in the development process are listed in Tab. 4.2, although there is a good chance that this choice will be revised during future optimizations. Fig. 4.1 displays the excitation and emission spectra provided by Semrock.

Dye	Excitation filter	Emission filter
CY5	FF02-628/40	FF01-676/29
CY5.5	FF01-680/22	FF01-692/LP

Table 4.2: Excitation and emission Semrock filters chosen for the new IR channels.

Note the choice of an high-pass filter for the CY5.5 dye, which substituted a band-pass with a similar cut to evaluate if this could enhance the channel performances. In fact the spectrum above the 692 nm cut should not present any relevant crosstalk contribution, and a wider filter can gather more fluorescence signal. The new channels work at longer wavelengths than the others, so to provide them with the proper excitation light two new LEDs from Cree have been added to the system. They have the emission peaks at 625 nm and 660 nm.

4.1.2 New sensor

The sensor choice aimed to strongly improve the optical performances of the system. Desirable features, from the experience gained with the previous version of the device,

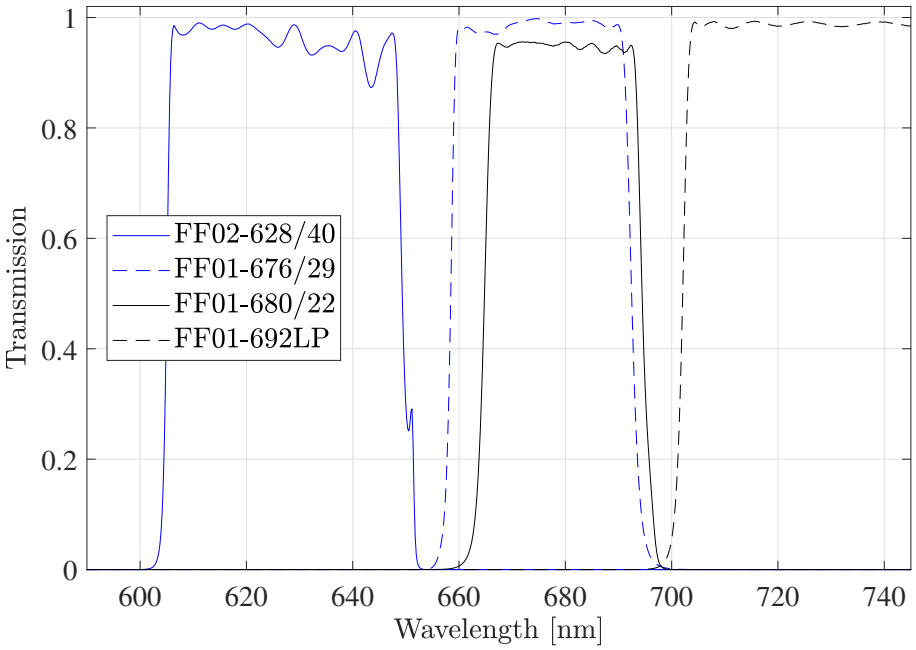


Figure 4.1: Excitation (continuous lines) and emission (dashed lines) theoretical spectra of the new IR filters. The CY5 couple is drawn in blue, the CY5.5 couple in black.

were pixels with a larger area to enhance sensitivity and SNR, and a sensor capable of high dynamic range (HDR) imaging. After evaluating some alternatives of commercially available devices, it turned out that STMicroelectronics was developing an HDR imaging sensor for the automotive industry, with a pixel size of 3.75 µm by 3.75 µm. It was the perfect candidate, the only issue being its considerable size, which would definitely not allow the integration of four separate devices inside the existing Q3-Plus case. Since the actual size of the packaged sensor occupies about the same physical area the old cameras did, it was decided to go for a single detector with four custom separate lens modules integrated on top of it. In this way, each acquisition would produce a picture with replicas of the chip as "seen" by all the channels. While in principle just one fourth of the image needs to be used to extract the relevant data, it could be possible to use the signal detected by the other channels during post-processing to address cross-talk issues. The sensor, shown in Fig. 4.2, is an STMicroelectronics VG6640 high dynamic range (HDR) image sensor for automotive applications. It features a 1280 pixel x 960 pixel resolution, up to 132 dB of dynamic range, 42 dB signal to noise ratio, 10 bit ADCs, and occupies an area of 9 mm by 9 mm. With respect to the old cameras the VG6640 offers a wide range of Bayer image pre-processing functions, and this gives much more control over the acquisition process. In particular each pixel of the array is sampled three times per frame, with different integration times, and these data streams go through an

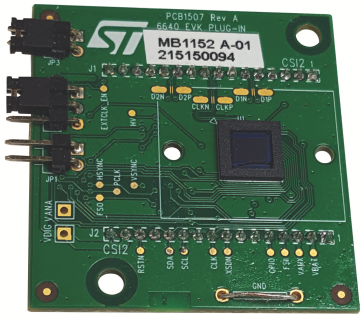


Figure 4.2: STMicroelectronics VG6640 HDR sensor mounted on sample PCB.

HDR pipeline that merges the information into a single data value. Each HDR branch can apply dark calibration algorithms, a digital gain, defect correction and white balance to the data before merging. In the final stages a piecewise linear (PWL) algorithm is applied to compress the information and save bandwidth. All of the algorithms (merging, PWL, those in the HDR pipeline) can be customized by writing into the control registers of the sensor.

Before settling for the VG6640, I tested its compatibility with the Q3-Plus environment by means of a demo board which was provided along with a few samples of the sensor. The demo board connects via a dual-in-line package (DIP) socket with the PCB where the VG6640 is mounted, and is controlled with a dedicated software through USB communication. The software allows for direct programming of almost any control register of the device, either through the graphical interface or by loading a script file. The size of the PCB with the sensor is compatible with the size of the OCS, so I was able to directly test the VG6640 performances adapting a 3D-printed version of the Q3-Plus body to host it, as schematically depicted in Fig. 4.3. I managed to place the new sensor, plugged

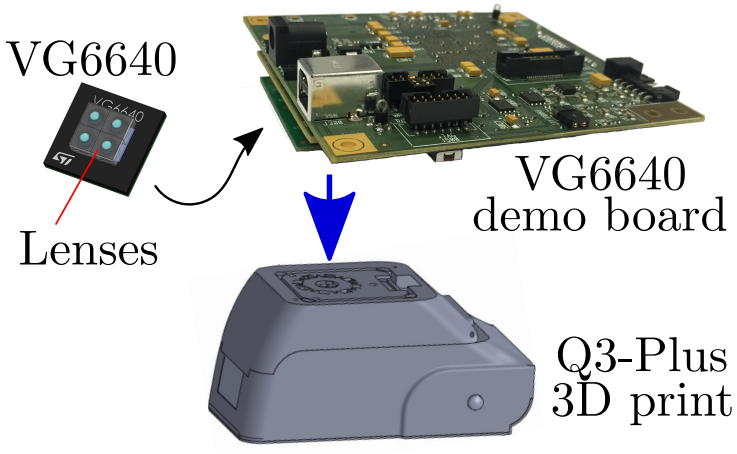


Figure 4.3: Schematic representation of the VG6640 testing setup.

into the demo board, upside-down on the Q3-Plus cut lid. The sensor was hosted in a reprinted version of the lens holder that used to house the four cameras in the original version, adjusted so that the whole sensor frame would fit in it. I was still able to control the Q3-Plus LEDs by ejecting its OCS from the front side, and plugging its flex cable into another Q3-Plus. This meant that I could not control the TCS, therefore a real-time PCR experiment could not be conducted with this configuration. To deal with the absence of a lens system, I recycled the ones that were inside four STMicroelectronics camera modules similar to the old ones. The top layer of the lens stack is an IR filter, which I removed before inserting the lenses inside a 3D-printed holder that fit above the sensor.

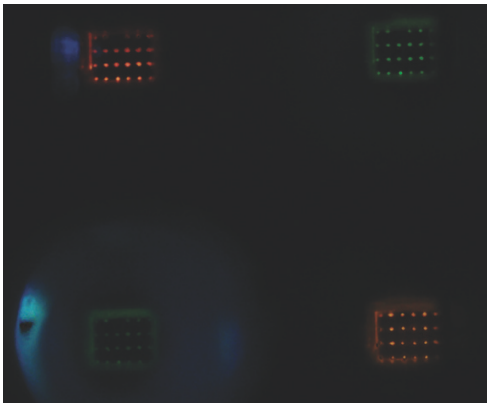


Figure 4.4: Fluorescence image captured with the new sensor in the demo board test setup.

Working with the demo board and its controlling software I was able to demonstrate that this system could, indeed, work. Fig. 4.4 shows an example of fluorescence light emitted from a matrix of quantum dots deposed on a bare silicon chip. The green replicas correspond to the old FAM and VIC channels, while the red ones are taken through the new filters. After determining that an implementation of the new sensor inside the existing structure was feasible, it became necessary to design a new lens system tailored to optimize the sensor performance.

4.2 Optical model

An imaging system is characterized by a number of fundamental parameters, schematically shown in Fig. 4.5. The field of view (FOV) is the portion of the object that is imaged on the sensor through the lens. The working distance (WD) is defined as the space between the front of the lens and the object under inspection. Resolution is the minimum feature size of the object that can be distinguished by the imaging system, and the depth of field (DOF) is the maximum object depth that can be maintained entirely in acceptable

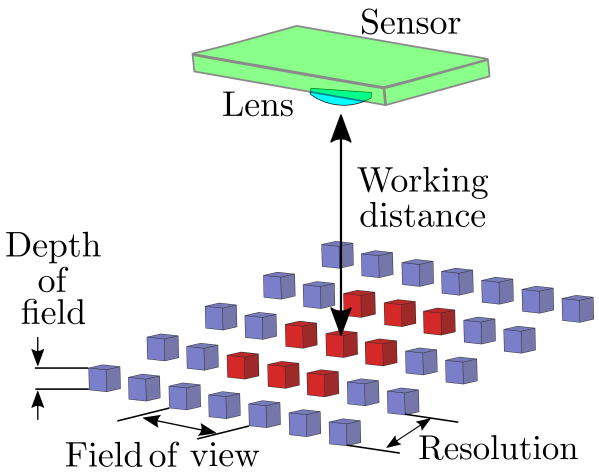


Figure 4.5: Fundamental parameters of an imaging system.

focus. The sensor size is the active area of the camera. In the case of the four separate modules the FOV needed to fill the whole active area of a single camera, with a single shared sensor the area to fill is roughly one fourth of the total. Most of those parameters are mechanical constraints that are already fixed by the Q3-Plus current structure, which should be modified as little as possible to minimize the final cost of the redesign. The size of the chip, and in particular the portion of it that contains the six wells, defines the field of view, while the working distance is roughly fixed by the body of the device to $WD \approx 3$ cm. Resolution and depth of field play a less relevant role, since the purpose of the system is primarily to track fluorescence variations in time rather than to acquire well-defined images.

A rough estimation of a suitable focal length for the system can be obtained assuming a simple structure with a single plano-convex lens between chip and sensor, as shown in Fig. 4.6. The system primary magnification is calculated as $M = \frac{sensor\ size}{field\ of\ view}$. The same quantity can also be expressed as the ratio between the distances of image and object from the lens:

$$M = \frac{d_i}{WD} \approx \frac{f_L}{WD} \tag{4.1}$$

where the distance d_i has been approximated with the focal length assuming $d_i \ll WD$,

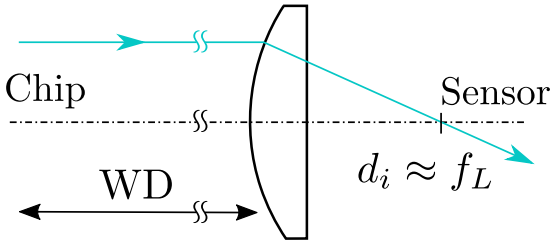


Figure 4.6: Simple model for the focal length estimation.

which is easily verified. When I was modeling this system the new sensor had not been chosen yet, so I took the performance of the old system as a reference. Given the chip size, which is 18 mm by 16 mm, and taking into account the pixel matrix of the VW6558 sensor, which is 640 pixel by 480 pixel, the limiting magnification between horizontal and vertical direction turns out to be the first one. Therefore, also keeping into account that during elaboration about a quarter of the acquired picture is discarded to crop the relevant part, and knowing that the single pixel area is 2.2 µm^2, the focal length f_L from Eq. 4.1 can be written as

$$f_L \approx \frac{0.5 \cdot 2.2\ \mu\text{m} \cdot 640\ pixel}{18\ \text{mm}} \cdot 30\ \text{mm} = 1.17\ \text{mm}. \tag{4.2}$$

Once I had this result, it was necessary to find out if a lens with this focal lens would need to actually be placed at a reasonable distance d_i from the sensor, so that everything could still fit into the existing outer structure. To understand this, I needed to model the optical system in some detail.

The system of lenses that focuses light on the sensor is composed of several elements designed to maximize the performance of the camera in relation to its specifications, and while I did not have access to the detailed information, it is safe enough to picture an equivalent model as sketched in Fig. 4.7. The IR filter is one of those elements, and it is

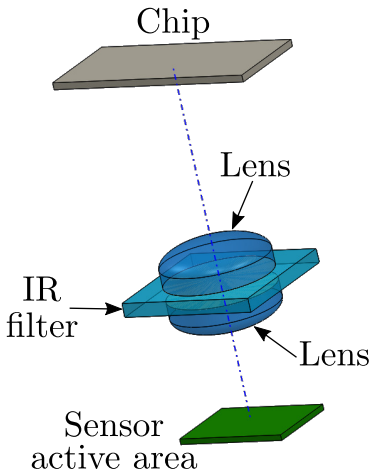


Figure 4.7: Equivalent model of the VW6558 lens system. The IR filter is sandwiched between two sets of lenses here represented with an equivalent one.

sandwiched between two sets of optical components that can always be substituted with a single equivalent lens. A powerful tool to mathematically represent an optical component is its ABCD matrix, or ray transfer matrix. Let us assume that paraxial approximation holds, which implies to consider only rays that travel through the component with a small angle and a small distance with respect to the optical axis of the system. The ABCD

matrix can be constructed considering the path that light rays follow when they go through the optical component, taking as reference input and output two planes orthogonal to the optical axis which delimit the component. Given any input ray, that crosses the input plane at a distance x_1 from the optical axis and makes an angle θ_1 with it, it will travel inside the optical component and emerge from the output plane at a distance x_2 and forming an angle θ_2. These quantities are related by the expression in Eq. 4.3:

$$\begin{pmatrix} x_2 \\ \theta_2 \end{pmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{pmatrix} x_1 \\ \theta_1 \end{pmatrix}. \tag{4.3}$$

A system of any complexity can be represented by the product of all the ABCD matrices of its elements, so that for a given input ray it is always possible to calculate the corresponding ray at the exit plane. For the system in Fig. 4.7 then the total matrix M_{TOT} is the product of seven matrices for the propagation, in this order: in the free space between the sensor and the second lens, in the second lens, in the IR filter, through the first lens and in the free space between the first lens and the chip. The total matrix of the system M_{TOT} reads:

$$M_{TOT} = \begin{bmatrix} 1 & d_i \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{n_{lens}-n_{air}}{R_2 \cdot n_{air}} & \frac{n_{lens}}{n_{air}} \end{bmatrix} \begin{bmatrix} 1 & t_{e_2} \\ 0 & 1 \end{bmatrix} M_F \begin{bmatrix} 1 & t_{e_1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{n_{air}-n_{lens}}{R_1 \cdot n_{lens}} & \frac{n_{air}}{n_{lens}} \end{bmatrix} \begin{bmatrix} 1 & WD \\ 0 & 1 \end{bmatrix} \tag{4.4}$$

where

$$M_F = \begin{bmatrix} 1 & 0 \\ 0 & \frac{n_{filter}}{n_{lens}} \end{bmatrix} \begin{bmatrix} 1 & t_f \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{n_{lens}}{n_{filter}} \end{bmatrix}. \tag{4.5}$$

Here, t_e is the thickness of the cylindrical section of a lens, n is the refractive index in the specified medium, t_f is the filter thickness and R is the radius of curvature of a lens.

To evaluate the only unknown, d_i, I calculated the same M_{TOT} matrix with a different approach. The *Central Theorem of First-Order Ray Tracing* states that M_{TOT} can be calculated knowing any couple of rays entering the system and the relative output rays. Calling m_{ij} the elements of the matrix, they can be calculated from the expression in Eq. 4.6:

$$\begin{bmatrix} x_1^{in} & \theta_1^{in} & 0 & 0 \\ x_2^{in} & \theta_2^{in} & 0 & 0 \\ 0 & 0 & x_1^{in} & \theta_1^{in} \\ 0 & 0 & x_2^{in} & \theta_2^{in} \end{bmatrix} \begin{bmatrix} m_{11} \\ m_{12} \\ m_{21} \\ m_{22} \end{bmatrix} = \begin{bmatrix} x_1^{out} \\ x_2^{out} \\ \theta_1^{out} \\ \theta_2^{out} \end{bmatrix} \tag{4.6}$$

where again a single ray is described in terms of distance and angle as (x_i, θ_i). The easiest rays to consider through a complex system are called *chief ray* and *marginal ray*. Those

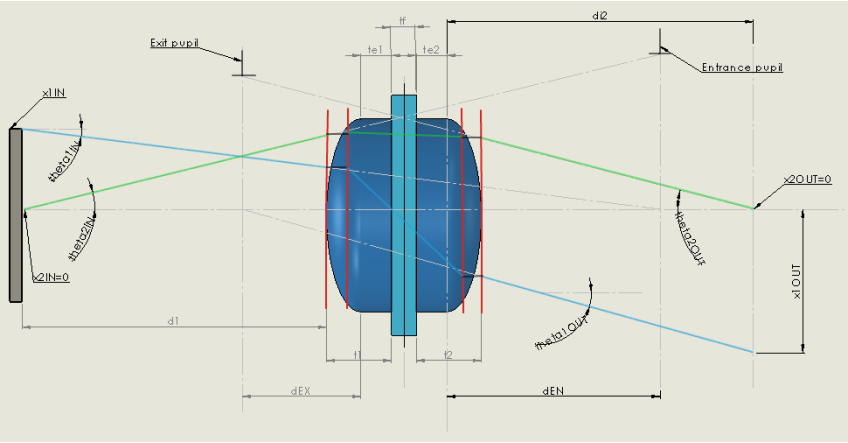


Figure 4.8: Diagram used to calculate the ABCD total matrix of the optical system with the Central Theorem of First-Order Ray Tracing. The lenses are treated as thick lenses, and therefore are represented considering two principal planes (in red) to draw rays paths.

are the image of the *aperture stop*, which is the aperture which limits rays travel the most inside the whole system, in the object space and in the image space. In the first case the image is called *entrance pupil*, in the other *exit pupil*. The aperture stop can be the lens clear aperture, which is twice the radius of the area r_{lens} where the lens meets its specification, or could be smaller, for example the window left open by the lens housing. In this case it does not matter, because it can be shown from the matrix equations that the parameter d_i does not depend on the clear aperture. The marginal ray starts at the point where the object crosses the optical axis, is directed to the edge of the entrance pupil and crosses the optical axis again as if it came from the edge of the exit pupil. The chief ray starts at the edge of the object directed to the center of the entrance pupil, passes through the center of the aperture stop and exits as if it came from the center of the exit pupil.

Fig. 4.8 shows the marginal ray (blue) and the chief ray (green) inside the system, as well as the other variables used to calculate the input and output parameters. The entrance pupil position d_{EN} (Eq. 4.7) and size (Eq. 4.8) can be expressed by means of the lens-maker's and magnification formulas, considering the lens closer to the object:

$$\frac{1}{f_1} = \frac{1}{t_{e_1}} + \frac{1}{d_{EN}} \rightarrow d_{EN} = \frac{t_{e_1} f_1}{t_{e_1} - f_1} \tag{4.7}$$

$$m_{EN} = -\frac{d_{EN}}{t_{e_1}} = \frac{h_{EN}}{r_{lens}} \rightarrow h_{EN} = \frac{|d_{EN}|}{t_{e_1}} r_{lens}. \tag{4.8}$$

For the exit pupil the same logic applies. Solving the equations for the second lens one obtains

$$\frac{1}{f_2} = \frac{1}{t_{e_2}} + \frac{1}{d_{EX}} \rightarrow d_{EX} = \frac{t_{e_2} f_2}{t_{e_2} - f_2} \tag{4.9}$$

$$m_{EX} = -\frac{d_{EX}}{t_{e_2}} = \frac{h_{EX}}{r_{lens}} \rightarrow h_{EX} = \frac{|d_{EX}|}{t_{e_2}} r_{lens}. \tag{4.10}$$

Starting from those results and from the marginal and chief rays definitions, and taking into account the geometry of the system in Fig. 4.8, it is possible to calculate the x and θ values for the two rays. The same information on the exiting rays can be obtained using Eq. 4.4. Then, using those results with Eq. 4.6 and solving for d_i, a value of $d_i = 7.1$ mm is found. This is still a reasonable result, the available space in the lid in the vertical dimension is enough to position the sensor at this distance from a lens.

From the same calculations I obtained a curvature radius $R_c = 1.967$ mm for the lenses, while considering a refractive index equal to $n_{SU-8} = 1.595$, which is an average value in the visible range for the SU82025 photoresist I later used for lens fabrication, and supposing $t_{e_1} = t_{e_2} = 1$ mm to account for the glass lens substrate.

The optical modeling showed that the firs step towards the design of a focusing system for the new sensor was to equip it with four lenses with a focal of about 1 mm and base diameter of roughly 0.8 mm. The Q3-Plus internal structure is such that a lens system could either be directly fabricated on the sensor surface, through a custom industrial process, or it could be a discrete component encased in the space where the emission filters are aligned. The latter option, the only viable in my case, adds tight dimensional constraints which leave a total available area for each lens of about 1 mm^2. All the commercially available solutions for single lenses offer components that cannot satisfy all the requirements together. For example molded glass lenses offer focal lengths of about 1.5 mm and higher, for a base diameter no smaller than 2.5 mm. Microlens arrays have focal lengths of 0.8 mm and below, or 1.6 mm and above, and not only they usually come in a 5 mm x 5 mm size or larger, but they are also quite expensive. Since the needed lens had characteristics that are in between those of commercially available products, it was necessary to find an approach to build a custom component.

Amongst the many alternatives for prototyping a lens, like replica molding with glass structures or ink-jet methods, a versatile technology is two-photon polymerization (2PP). It allows the fabrication of 3D resin micro and nanostructures with a remarkable freedom in the design process.

4.3 Two-photon polymerization (2PP)

Photolithography is a well established research field of science and technology driven by the need to produce structures with increasing smaller features, most notably in the semiconductor industry. Traditional techniques are mostly focused on a 2D perspective, where fabrication is conducted as a sequence of repeated lithographic steps. As the

need for improved spatial resolution, minor complexity and reduced costs increased, however, new materials and procedures have been devised over time. Over the last $15-20$ years a sub-100 nm resolution 3D Direct Laser Writing (DLW) technology, called two-photon polymerization has emerged and found application into a growing number of fields. The most attractive advantage of this technology is the potential to build complex three-dimensional objects without geometrical limitations, with feature sizes beyond the diffraction limit [20].

4.3.1 2PP working principle

Two-photon polymerization is based on the nonlinear absorption of multiple photons by UV photosensitive materials: when the beam of a femtosecond infrared laser is tightly focused inside the volume of such a material, a polymerization process can be initiated by nonlinear absorption within the focal volume. As the laser beam is moved along a desired path, 3D structures of high complexity can be fabricated in the material. The steps of a 2PP fabrication are sketched in Fig. 4.9. First the photopolymer, a resin, is deposed over a transparent substrate by drop-casting or spinning techniques and positioned on a 3D actuation stage, so that during fabrication is the sample that actually moves around the fixed laser spot. It is possible to move the laser beam instead, this leads to faster writing but limited FOV. During fabrication only the laser focal volume, called *voxel*, contraption

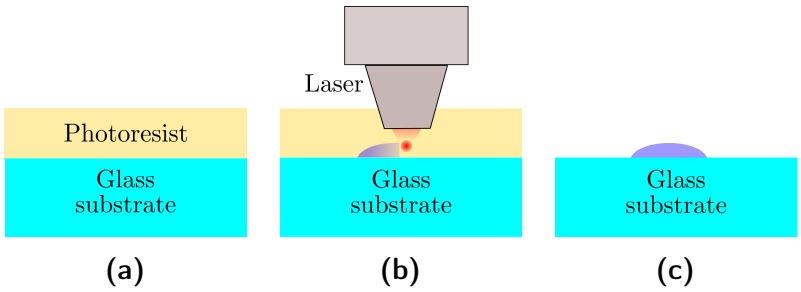


Figure 4.9: Schematic illustration of 2PP fabrication steps: resin deposition (a), laser writing (b), polymerized structure after development (c).

of *volumetric pixel*, has an intensity higher than a certain threshold so that polymerization occurs only in this region. By moving the focal spot it is possible to use the laser as a pen that writes in the resin.

At the end of the writing process the excess material needs to be removed with a chemical developer to reveal the final microstructure. The polymer is usually a negative photoresist, which means that the portion of material that is exposed to the laser beam becomes insoluble to the developer, which instead removes the unpolymerized material. This approach is usually convenient because of the small size of the designs: positive

photoresist will become soluble only where irradiated, and this would require a much higher amount of fabrication time. Development generally involves washing with one or more solvents and can also require additional processing steps as baking.

The resolution of the microstructure is determined by the size of the voxel, that is the minimum feature achievable given certain fabrication conditions. The voxel diameter in the xy plane (Eq. 4.11) and height in the z direction (Eq. 4.12) can be theoretically estimated [21] considering a Gaussian laser intensity profile.

$$d = \frac{\lambda\pi}{NA}\sqrt{ln(\frac{I_0^2\ T_{exp}\ \beta\ \tau_{pulse}\ N_{rep}}{I_{th}})} \tag{4.11}$$

$$h = 2\ \frac{z_R}{n_{poly}}\sqrt{e^{\frac{1}{2}(\frac{d}{w_0})^2} - 1} \tag{4.12}$$

Here λ is the laser radiation wavelength, NA is the numerical aperture of the Gaussian beam, I_0^2 is intensity in the vicinity of focus, T_{exp} is the exposure time, β is an experimentally derived constant, τ_{pulse} is the pulse duration, N_{rep} is the repetition rate, I_{th} is the polymerization threshold intensity, z_R is the Rayleigh range of the Gaussian beam, n_{poly} is the polymer refractive index and $w_0 = \frac{\lambda}{\pi NA}$ is the focal spot size of the Gaussian beam. The equations show that the size of the exposed region changes with laser intensity (and hence, power) and exposure time. Lower intensities and shorter exposure intervals will produce smaller voxels.

4.3.2 Polymerization mechanisms and 2PP materials

The selective polymerization of the laser focal volume happens because, while the sample resin is transparent to the laser operating wavelength, so that single photons would normally just pass through it, the high intensities that can be only reached inside the voxel of a tightly focused beam can trigger two-photon absorption (2PA). This is a nonlinear third-order optical process which happens when a photopolymer absorbs two low-energy photons simultaneously through a virtual intermediate energy level. As shown in Fig. 4.10, while the material is transparent for a single photon of energy $E_{ph} = h\nu_{laser}$, if the photon is followed by a second one within the virtual state short lifetime ($\tau \approx 10$ fs), both can be absorbed as if they were a single photon with energy $E_{2PA} = 2h\nu_{laser}$. This will excite the absorbing molecule.

Two-photon absorption efficiency depends on the square of the laser intensity. It is several orders of magnitude weaker than linear (single photon) absorption, therefore it is actually noticeable only in particular conditions, such as in 2PP processes, where the other phenomenon is absent. Since to have a good efficiency it is necessary to concentrate a

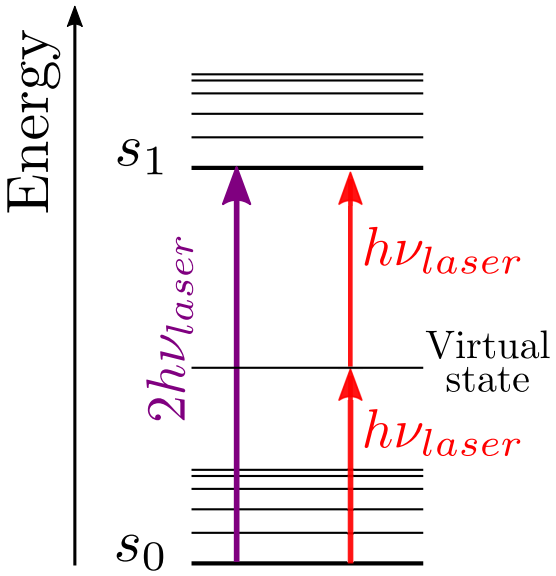


Figure 4.10: Jablonski energy diagram for two photon absorption. The photosensitive molecule is firstly excited from ground state to a virtual state by the first photon with energy $E_{ph} = h\nu_{laser}$, then to an higher, real energy level by the second one. It is the equivalent of the absorption of a double-energy photon.

large number of photons in a small volume, a femtosecond laser is the ideal light source for this. In fact, such a laser produces pulses with high peak power and very short duration, so that the intensity of the associated electric field is high enough for 2PA. The same condition could be replicated with longer pulses and much greater average power, but this would produce thermal accumulation effects that would damage the resin. In general photon absorption efficiency depends on the absorption cross-section, which for a single photon is independent of the beam area A and for 2PA has a A^{-1} proportionality [22]. This explains why if a resin is irradiated with light of a wavelength in its absorption range the whole beam path will be polymerized, while if the wavelength will only allow 2PA there will be only an absorption peak in the regions where the beam is smaller, the voxel.

Voxels are usually ellipsoidal, and the xy plane minor axes perpendicular to the optical axis z are about 3-5 times smaller than the major axis. This shape is determined in part by the point-spread function (PSF) of the laser light intensity near the focus, and since the probability for 2PA is proportional to I^2 this effectively narrows the PSF of the beam near the focal point, so that it can be smaller than the diffraction limit at the excitation wavelength. This happens because the polymerization process is a threshold phenomenon, which means that even though two-photon absorption can in principle happen in all the irradiated volume, the chain polymerization will start only if the density of reactive sites goes beyond a certain level. Therefore only a portion of the narrow PSF peak will be

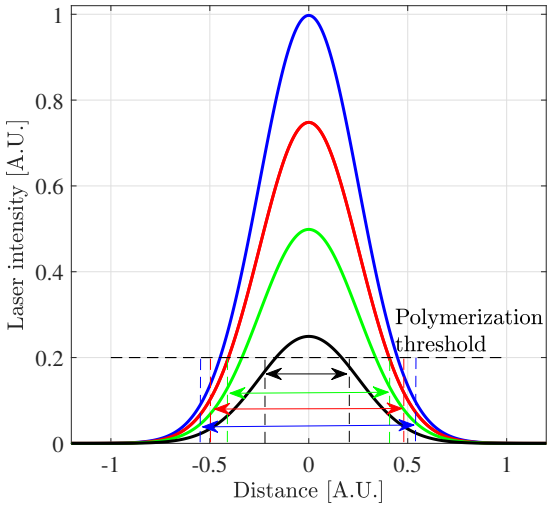


Figure 4.11: Diagram illustrating how voxel sizes change with different laser intensities. Only the volume where the intensity is above threshold will be polymerized.

above the threshold and concur to define the smallest polymerizable volume. Fig. 4.11 gives an intuitive description of this phenomenon.

The photopolymer resin used for 2PP, also called *prepolymer*, can be a viscous liquid, an amorphous solid or a gel. The choice of one option over another takes into consideration the trade-off between easiness of sample preparation and risk of undesired motion of structures during fabrication. There are two fundamental components in a prepolymer resin: the photoinitiator and the monomers that will form the chain. Other components can be added to change the resin characteristics, such as polymerization inhibitors to stabilize it, solvents to assist in casting, filler polymers to increase the viscosity or dyes to add new functions to polymerized structures. Chain-growth polymerization can be described with the chemical equation

$$(-M-)_n + M \longrightarrow (-M-)_{n+1} \tag{4.13}$$

where n is the degree of polymerization (the number of monomeric units) in the M_n polymer and M is a monomer. The process usually consists of four steps: chain initiation, propagation, transfer and termination. Chain initiation happens by means of an *initiator*, that starts the chemical process by creating an active site where single monomers will add one at a time, extending the polymer chain. The addition of each monomer unit regenerates the active site, so that the reaction can propagate. After a certain amount of time the active site can transfer to a new chain, terminating the chain but starting a new one, or can combine and end the polymerization process. Depending on the type of active center, the prepolymer resins can feature *radical* polymerization (the active center is a *free radical*) or *cationic* polymerization (the active center is a *carbocation*) [22], the

former being a more popular choice for characteristics as high reaction rates and ease of processing.

The first step in radical polymerization is the nonlinear excitation of a photoinitiator by two-photon absorption when the laser is focused inside the resin. This generates one or two free radicals, which are highly chemically-reactive molecules with an unpaired valence electron. Then the radicals are transferred from the initiator molecules to the monomer units present in the resin, and the propagation phase starts. The majority of the resins that feature radical polymerization employ acrylate monomers. Commonly used acrylic resins include many photosensitive sol-gel hybrid materials such as ORMOCER [23] or SZ2080 [24], an alternative can be silicon-based polydimethylsiloxane (PDMS).

An alternative mechanism to radical polymerization is cationic polymerization. The reaction begins when the photoinitiator generates a carbocation, which is a strong Lewis acid with positively charged carbon atoms, capable of polymerizing epoxides and vinyl ethers. Propagation proceeds as in the previous case, via addition of monomers to the active species until chain termination. The most widely used materials for cationic polymerization are SU-8 [25] [26] and SCR-701 [27] epoxides.

4.3.3 Single lens design

The model in Fig. 4.12 shows a 3D representation of the lens model described in Par. 4.2. After a structure has been designed for 2PP fabrication, it needs to be translated

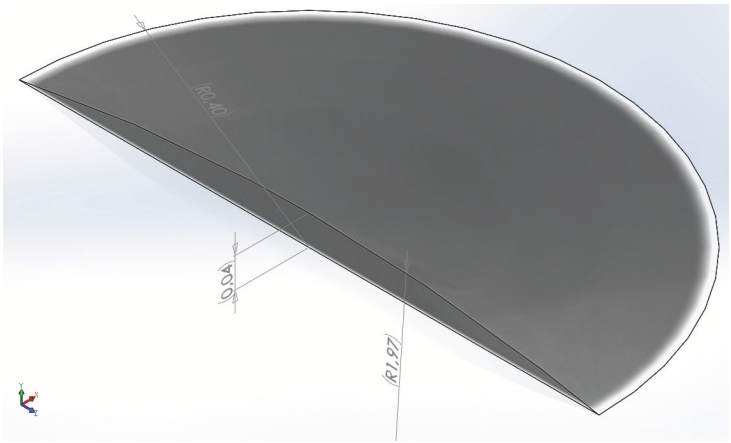


Figure 4.12: SolidWorks CAD for a lens with 1.967 mm curvature radius and 0.4 mm base radius. Note that the actual size for the curvature radius has been approximated to $R_c = 2$ mm during fabrication.

in a program written with a proprietary version of G-code, so that the control software of the fabrication setup can load it and execute the instructions. G-code is the most widely used numerical control programming language, mainly employed in computer-

aided manufacturing to control automated machine tools like milling machines or 3D printers. It essentially describes the position of a mechanical actuator in the working space, and manages the speed and timing of the movements between a position and the next. Therefore, after designing any one of the structures described in Par. 4.5 and 4.6, I had to write the relative G-code programs to proceed to fabrication.

To describe the lens in terms of G-code commands, the 3D structure needs to be sliced into planes which contain a bidimensional path for the laser to follow. During fabrication the laser draws one plane, then the distance in the z direction is changed and the next plane can be written. In the case of this lens, the planes are a sequence of circumferences with radii ranging from zero (at the top) to a certain base size. The physical windows above the Q3-Plus filters are squares of 0.8 mm by 0.8 mm, therefore in principle it is necessary to fabricate a lens with an $r_{base} = 0.4$ mm base radius. From simple trigonometric considerations, it follows that the lens sagittal height will be $z_{MAX} = R_c - \sqrt{R_c^2 - r_{base}^2} = 40$ µm. The design parameters of the lens are summarized in Fig. 4.13.

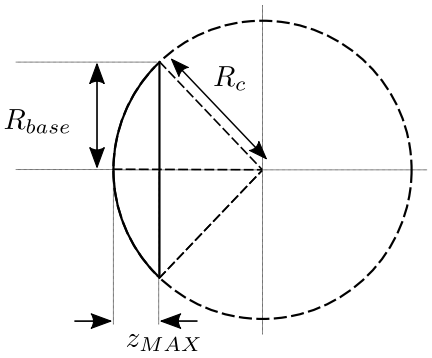


Figure 4.13: Design parameters of the lens.

Considering the layer-by-layer fabrication process just described, it has been shown [28] that the best implementation of the scanning method to draw the whole lens is an annular scanning approach with dynamical z position. This solution allows the fabrication of a smoother surface, Fig. 4.14 shows how the slicing becomes thinner towards the top of the lens. The vertical position of layer i is expressed by Eq. 4.14

$$z_i = R_c - \sqrt{R_c^2 - (i \cdot r_{spacing})^2} \tag{4.14}$$

where the $r_{spacing}$ parameter defines the fixed lateral step that shrinks the slice area in each layer.

Usually in a 2PP process the whole volume of a structure is polymerized, with the laser beam touching all the 3D points inside it. For structures of a considerable size like this lens, this would mean a too long fabrication time. Therefore to reduce it an

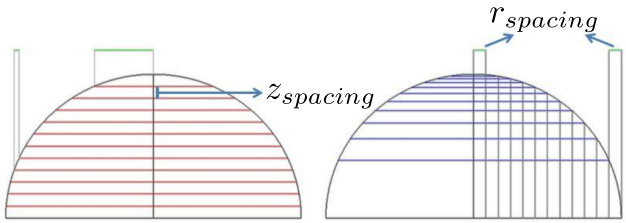


Figure 4.14: Comparison between annular scanning methods with fixed (left) and dynamical (righ) z position. In the first case the area of the top slices has a much bigger radius than in the other one, where a greater detail can be achieved. Image from [2].

useful strategy [29] is to just fabricate the outer shell of the lens, leaving the inner volume filled with unpolymerized material. This structure needs then to be cured by one-photon polymerization under a UV lamp to reach final stability and homogeneous optical quality.

4.4 Experimental setup

The 2PP fabrication process was divided between the IFN-CNR *DiamANT* laboratory at the Department of Physics of Politecnico di Milano, where the actual 2PP setup is located, and the lithography laboratory of the *More Than Moore Labs* at STMicroelectronics, where I prepared the samples.

4.4.1 Sample preparation

I worked with samples that consisted of a microscope cover glass used as a substrate and coated with a photosensitive resin. For the substrate I used 18 mm x 18 mm borosilicate cover glasses, with thickness between 0.13 mm and 0.17 mm and $n = 1.52$. There is an abundant choice of materials for 2PP fabrications (see Par. 4.3.2), between those that were available to use I ended up evaluating SZ2080 and SU-8. The SZ2080 is easier to work with, because its polymerization can be directly monitored during fabrication with a camera. After some considerations I decided to choose a photoresist from MicroChem SU-8 2000 family, which is more complex to work with but is characterized by better optical properties, in particular the surface roughness which is of primary importance for a lens.

SU-8 2000 is an epoxy based negative photoresist which is especially designed for micromachining and similar applications. It is available in twelve viscosities, from 2.49 cSt (centistokes) for SU-8 2000.5 to 22 000 cSt for SU-8 2075. I chose SU-8 2025, which has a viscosity of 4500 cSt, because it could produce coatings between 41 nm and 75 nm thick, which is what the design of the lens required. This resist refractive index after development ranges approximatively from 1.616 and 1.574 in the visible wavelengths. The

resin polymerization process begins with light exposure and terminates with thermal cross-linking, therefore before being developed it always has to undergo a post exposure bake (PEB).

The sample preparation procedure consists of three steps: substrate cleaning, spin coating and soft bake. The two initial operations are conducted on a programmable spin coater. First a new cover glass is placed on the spin coater chuck and locked in place applying a vacuum, then the machine is let spinning for two minutes at 2000 *rpm* while acetone and later isopropyl alcohol are run on it at 25 seconds intervals.

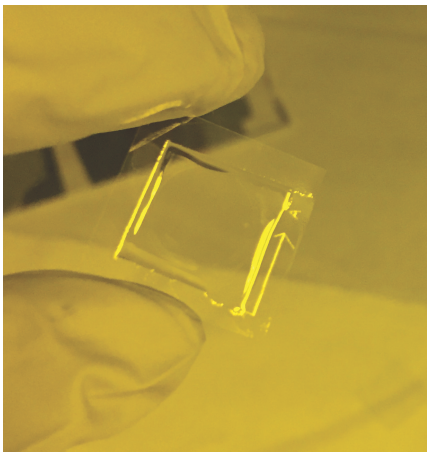
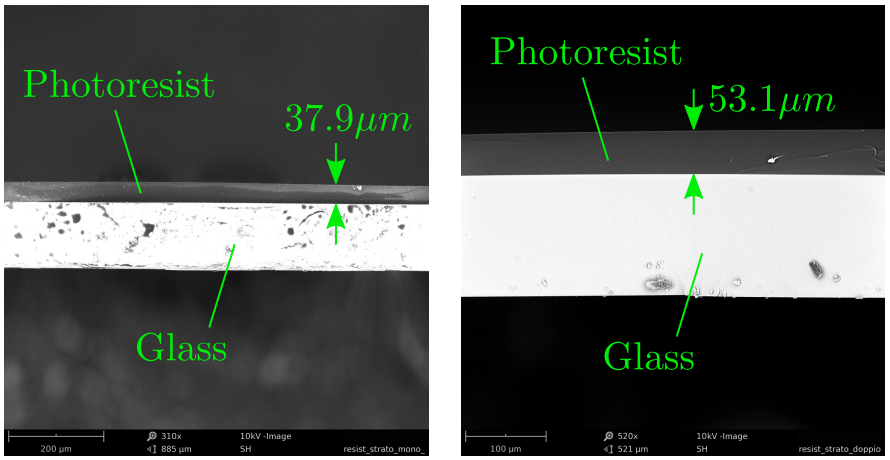


Figure 4.15: Glass substrate coated with a photoresist layer.



(a) **(b)**

Figure 4.16: SEM pictures of single and double photoresist layers on substrate glass.

Then a layer of resist is poured on the clean glass and a coating program featuring two ramps of 5 seconds at 750 *rpm* and 60 seconds at 1500 *rpm* is started; an example of a coated glass is presented in Fig. 4.15. After this the sample is removed from the spin coater it is placed sequentially on two hot plates set at different temperatures for

the PEB: 2 minutes at 65 °C followed by 7 minutes at 95 °C. Both the spin coating and soft bake procedures were devised to obtain a resist layer thickness between 41 µm and 75 µm, from the SU-8 2025 datasheet specifications. A scanning electron microscope (SEM) layer thickness inspection of some unpolymerized samples revealed that the actual result obtained was thinner than planned. Therefore in subsequent sample fabrications I repeated the spin coating and soft bake sequence twice to increase the resist thickness. Fig. 4.16a shows a SEM image for the cross section of a single photoresist layer, which was on average between 14 µm and 43 µm thick, while Fig. 4.16b shows a clearly thicker double layer cross section. It was between 35 µm and 75 µm thick.

4.4.2 Two-photon polymerization fabrication setup

A typical direct laser writing fabrication setup is composed of a laser source, beam steering and focusing optics, power control devices, a beam shutter and a sample/beam motion system with the relative control software. A simplified diagram of the DiamANT 2PP setup, which follows this scheme, is presented in Fig. 4.17. The actual fabrication

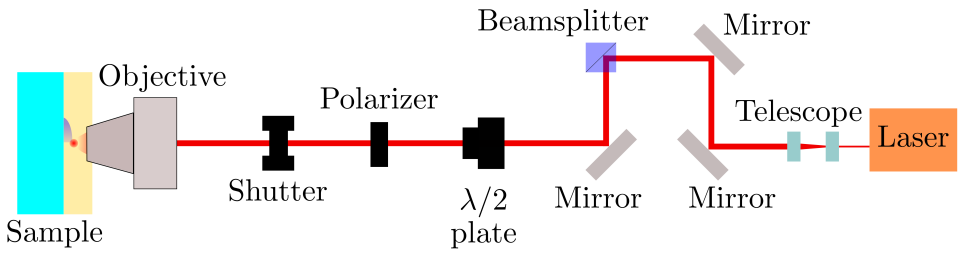


Figure 4.17: Schematic representation of the 2PP setup at the Department of Physics.

of a three-dimensional shape is controlled by the stages, which move the sample around the fixed voxel. The DiamANT setup has three Aerotech ANT130 Nanopositioning stages for motion along the spatial directions. The stages guarantee a 75 nm accuracy over a travel range of 60 mm and a maximum speed of 350 mm/s with no load. The motion of the stages is controlled by the Aerotech A3200 CNC Operator Interface software, shown in Fig. 4.18, that loads and executes G-code programs. The optical path is composed of a number of important components. The origin of the path is of course the light source, which is a Toptica FemtoFiber pro NIR erbium doped fiber laser, consisting in a ring oscillator and an amplifier, with a saturable absorber mirror that enables modelocking operation. The pulsed regime has a repetition rate of 80 MHz, at a wavelength of 1560 nm. The laser is equipped with a second harmonic generation stage, that results in emission at 780 nm, which is the wavelength used for 2PP, with an output power of 160 mW. At the exit of the laser there is a telescope composed of two lenses that is used to expand the traveling beam, so it will have the correct size to overfill the writing objective aperture.

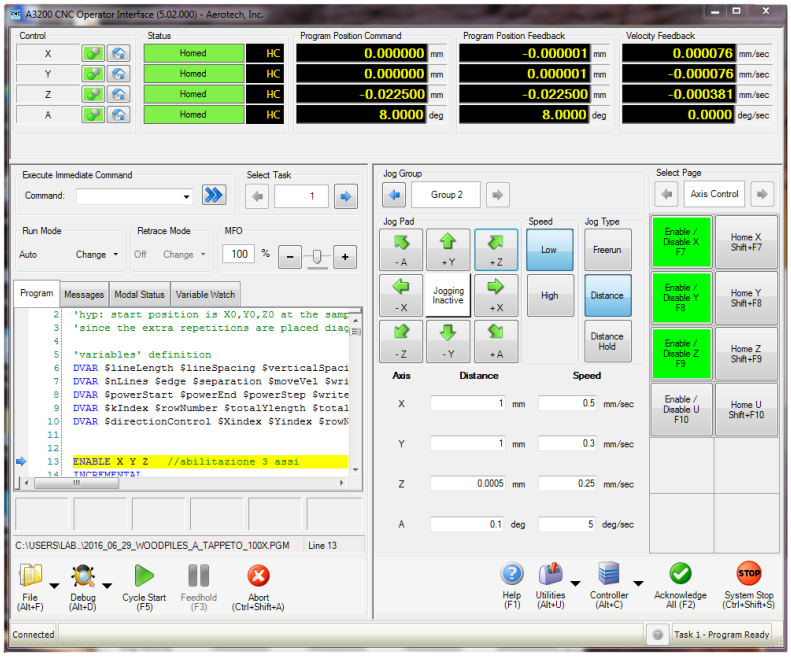


Figure 4.18: Aerotech A3200 CNC Operator Interface software. In the upper section there is the stage status panel, with activation options and positions. The lower section is divided between G-code program loading (left) and manual movement control of the stages (right).

The beam is directed across the optical table with a series of mirrors that are adjusted during the system initial alignment to correct the laser spatial tilt. After the mirrors the laser enters a beam splitter which separates it into two 2PP separate writing lines. The splitting ratio can be regulated to direct more power into a branch or the other if necessary. The line power is regulated through a combination of a half-wave plate and a polarizer. The former rotates the polarization of the electric field of the beam with respect to its optical axis. Rotating the plate of an angle α with respect to the optical axis will rotate the field polarization by 2α. A polarizer allows only the transit of the portion of electric field polarized along its optical axis. Therefore, if beam polarization and polarizer axis are aligned, the whole beam passes, otherwise if they are perpendicular the beam is completely blocked. Selection of an intermediate condition allows a fine regulation of the beam power. The system is equipped with a Uniblitz LS6S2Z1-NL shutter, controlled by a VCM-D1 driver. The shutter blades have a rise time of 700 µs, 150 Hz of maximum operating frequency and can work continuously at 10 Hz. A Thorlabs CCD webcam, used to focus the laser beam on the sample, completes the setup.

The writing operation is always preceded by a few preparation steps. The first operation is to switch on the laser source and let it reach a stable condition, half an hour can be a reasonable time to wait. After that, the beam path needs to be fine-tuned to obtain the

best possible alignment, with particular care to check that the laser completely fills the objective aperture and arrives orthogonally on the writing plane. When the alignment is done, a new sample can be cleaned and mounted on its holder. The laser reflection on the first air-glass interface of the sample is then used to correct the sample tilt, so that when the XY stages move the voxel will stay on the same horizontal plane. To do this, it is necessary to find the laser focus in one point of the glass with the help of the Thorlabs webcam described before. As the laser focus approaches the glass surface the camera will begin to detect its reflection. The tilt is corrected by manually adjusting a rotary stage below the sample to guarantee that the focus spot remains the same along the x and y directions, over a distance that covers at least half of the glass cover dimensions. The setup preparation is done after this last operation; a G-code program can be loaded and executed after the laser focus is brought at the glass-resin interface.

4.4.3 Inspection instruments

To evaluate the results of 2PP fabrications I used an optical microscope and a scanning electron microscope (SEM), both available in the laboratories of the Department of Physics. The optical microscope is a Nikon Eclipse ME600 that can work in both light transmission and reflection mode, and is equipped with objectives of magnification between 4x and 40x, and numerical apertures from 0.13 to 0.75. This microscope can also be used to inspect a sample before development, using the light transmission mode and covering the aperture of the light source with a red filter to prevent further polymerization. The SEM is a compact tabletop microscope by PhenomWorld, Phenom Pro, with acceleration tension between 5 and 10 kV and a resolution of 10 nm in optimal conditions. It can work in low vacuum mode for non-conductive samples, and has an optical camera for sample inspection before switching to electronic image mode. Coating of samples with a conductive layer of metal is often required to enable or improve the imaging of a SEM, because it inhibits charging and reduces thermal damage. Because of this the laboratory is equipped with a Cressington 108auto sputter coater, with argon gas supply and MTM-20 thickness monitor. The metal target used for the coating is gold. The MTM-20 measures the thickness of the metal layer through a capacitor placed in the deposition chamber, that changes its capacity when gold is deposited.

4.5 Objective and material characterization

As it has been described in Par. 4.3.1, the smallest feature size of a 2PP fabrication is defined by the voxel, which in turn depends on many factors. Once the laser source is

fixed, the choice of the objective is of primary importance, because it has a strong impact on the final result. The resin used for fabrication is chosen depending on the type of structures and their desired characteristics, but it is important to notice that a different material will also influence the working range of fundamental fabrication parameters as laser power and writing speed. In fact objective, material and fabrication parameters all play a role in the photochemical and thermal contributions to the polymerization reaction. Therefore to choose the best fabrication parameters it is important to characterize the performances of different configurations.

To investigate the available ranges of laser power and writing speed I fabricated for each objective a matrix of woodpiles, which are simple structures devised to evaluate the voxel size and mechanical stability, for various operational conditions. A woodpile consists of layers of one-dimensional rods stacked over one another, written alternatively in x and y direction. The woodpile matrix is built repeating this structure in rows, where each row is written with a different power and each woodpile of a row with an increasing speed. This will give a two-dimensional space that tracks the dependence on those two parameters, with a central zone of well-defined structures that represent the viable combinations of power and speed. This region will be delimited by the woodpiles that have been burnt because of high power or low writing speed, and by those that had insufficient mechanical stability to survive development because of the opposite conditions.

4.5.1 100x oil immersion objective

I fabricated the first woodpiles with a Carl Zeiss immersion oil microscope objective, with 100x magnification and a 1.4 NA. An oil immersion objective is best suited for a high fabrication resolution. The layer of oil, which is placed between the objective lens and the glass surface of the sample, avoids the strong refractive index change between glass and air, and therefore allows to achieve a numerical aperture higher than one, necessary to maintain a low aspect ratio of polymerized lines. I used a range of powers from 2 mW to 26.8 mW, and writing speeds between 0.031 25 mm/s and 4 mm/s. Fig. 4.19 shows the fabrication results after development.

Each woodpile is composed of ten layers of parallel rods, alternately drawn in the x and y directions. A plane is separated from the next by 1.2 µm in the vertical direction. There were only seven combinations of power and speed that produced stable structures, highlighted in Fig. 4.19. The surrounding woodpiles either collapsed or burned.

Fig. 4.20 shows a well formed, standing woodpile and one that was fabricated with the same power but higher speed. The higher speed does not allow the correct polymerization of the resin, therefore the structure begins to form but lacks in structural stability and

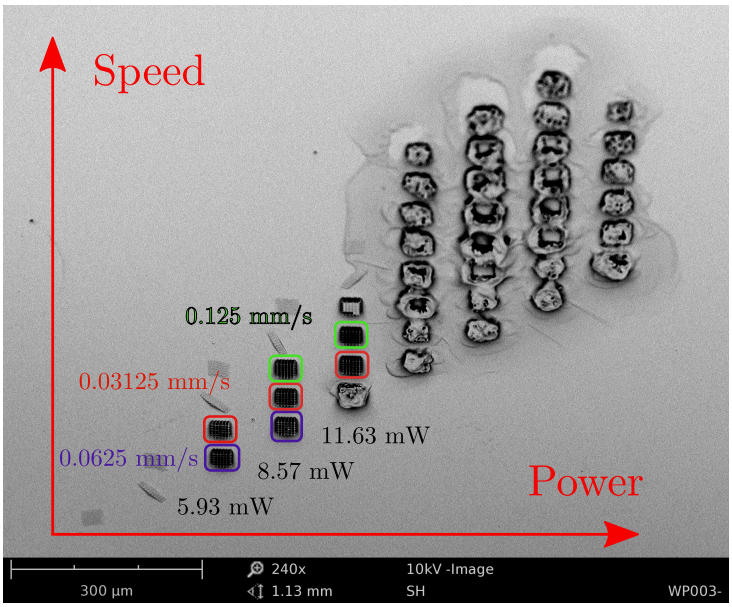
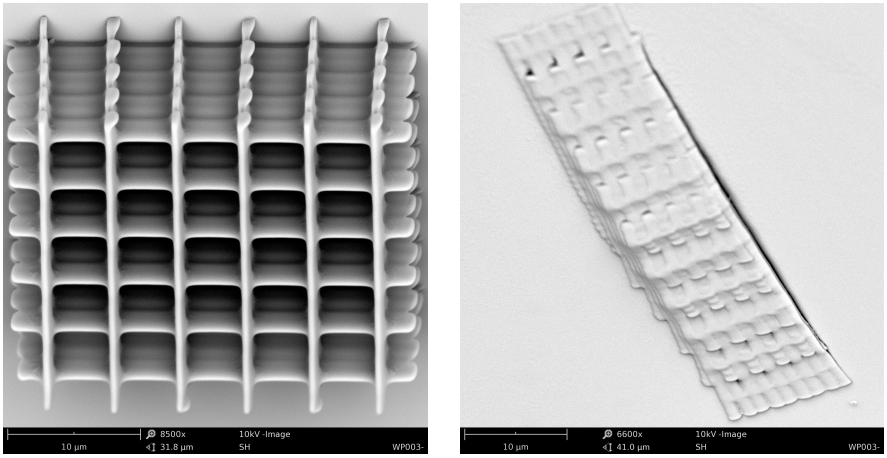


Figure 4.19: SEM picture of woodpiles fabricated with the 100x objective.



(a) **(b)**

Figure 4.20: Details of a well formed woodpile (power 8.57 mW, speed 0.0625 mm/s) (a) and a collapsed one (b) fabricated with the 100x objective.

collapses. This is related to the voxel size: the standing structures have x and y dimensions ranging from about 0.3 µm to about 0.6 µm, and z dimensions ranging from about 1.7 µm to about 2.3 µm, growing from low power/high speed to high power/low speed in the woodpile matrix. The woodpiles that fall have much smaller voxel dimensions, and collapse during the development process.

While an oil immersion objective can be the best choice when fabricating small and detailed structures, to reduce the writing time for a large structure such as a lens I needed to use an objective that delivered a bigger voxel, that would guarantee volume overlap between reasonably displaced scans. With the 63x objective, at a distance of just 1.2 µm

between scans the polymerized lines would barely touch each other, and only at the highest available powers and for very low speeds.

4.5.2 63x objective

Another available option was a Carl Zeiss microscope objective, with 63x magnification and a 0.68 NA. For its characterization I fabricated a woodpile matrix scanning powers from 20 mW to 50 mW and writing speeds between 0.015 625 mm/s and 0.5 mm/s, the results are shown in Fig. 4.21. The structure of the woodpiles is the same as in the

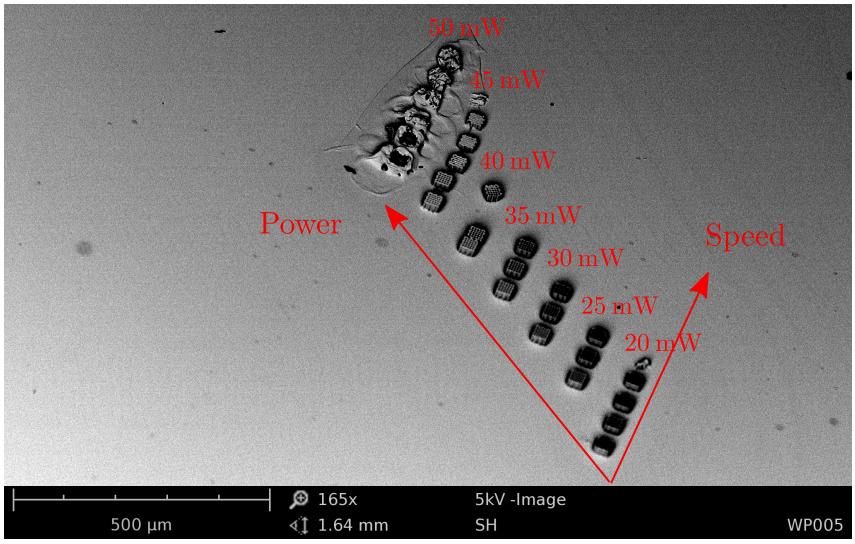
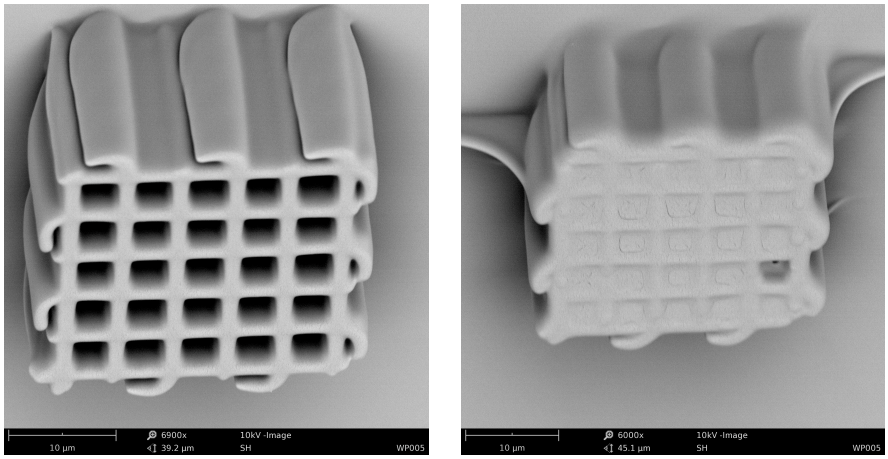


Figure 4.21: SEM picture of woodpiles fabricated with the 63x objective.

previous case. Note that this time only the highest power, 50 mW, could burn a line, while this never happened for low speeds. All the collapsed structures have been washed away with the development process. The woodpiles written at high power and low speed ended up filled with photoresist, despite a prolonged development to try and empty them. It is possible that the thermal effects associated with an high power laser beam slowly passing many times on the same area managed to polymerize the whole volume. One example is shown in Fig. 4.22b. The transversal sizes of the voxel ranged from about 0.7 µm to about 1.4 µm, and while this time measuring the vertical dimension proved more complicated, it was always about three times larger than the x and y sizes. This result pointed that the 63x objective could be a good option to fabricate the lens large design: using an high power value would allow to write at a reasonably high speed while avoiding the risk of burning the resist.

Fig. 4.23 shows the SEM image of a test structure written with the 63x objective, fabricated with 55 mW of power and writing at 0.031 25 mm/s. I made the 3D design in SolidWorks, and then used a dedicated software that translates the CAD file in a G-code



(a) **(b)**

Figure 4.22: Details of a well formed woodpile (power 30 mW, speed 0.031 25 mm/s) (a) and a standing but resist-filled one (b) (power 45 mW, speed 0.125 mm/s) fabricated with the 63x objective.

script. The software firstly recomposes the design as a group of horizontal slices of the three-dimensional figure, and then describes each of them with a series of parallel lines that represent the path for the laser voxel. The test structure was written with a 1 µm

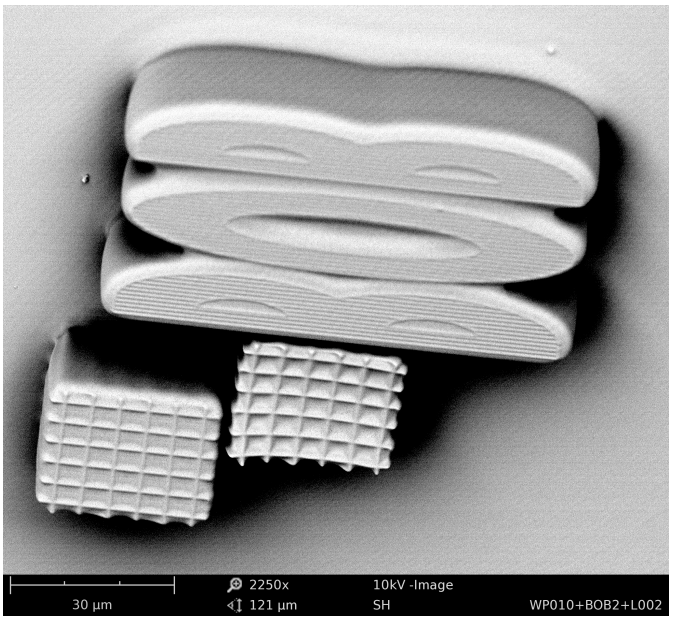


Figure 4.23: SEM picture of an 80 µm x 50 µm x 10 µm test structure fabricated with the 63x objective.

spacing in x, y and z directions, and although the lateral surfaces seem to be smooth and well defined, the top one clearly shows the pattern followed by the laser during fabrication. This is probably because there is less overlap of the voxels when the last slice is written.

This result confirmed that the 63x objective can be a good choice to fabricate large (for 2PP standards) and compact structures, although the surface unevenness hinted that a 1 µm gap between the paths followed by the laser is too much, and should be reduced.

4.6 Lens fabrication and results

The design process highlighted how the structural stability of the lens could pose a significant challenge, given its large base radius and its really small sagittal height, that could lead to a collapse of the central part of the lens. Therefore I decided to try different approaches to see if there was a design that proved to be more resilient than the others, and to shorten fabrication time for these trials I reduced the lens base radius to 250 µm, keeping the other design parameters unchanged.

Since the idea was to write only the outer shell of the lens, leaving the inner volume filled with unpolymerized photoresist, the strongest concern was to build a strong enough structure that would not collapse during development. So I decided to fabricate lenses with a single spherical cap, with a double cap, and with the addition of horizontal planes intersecting the shell. The relevant details about the fabrications are gathered in Table 4.3, where I have labeled the lens here described with an L and a progressive number for the sake of clarity. The N_{caps} parameter refers to the number of spherical caps, N_{rep} is

Lens code	N_{caps}	N_{rep}	$r_{spacing}$ [µm]	P [mW]	Notes
L1	1	3	0.2	30	Intersecting planes
L2	2	4	0.2	57.3	Hard baked
L3	2	3	0.2	57.8	
L4	1	3	0.1	57.8	

Table 4.3: Lens fabrication design details.

the number of passes of the laser over the same circle and $r_{spacing}$ is the vertical spacing parameter defined in Par. 4.3.3. All the lenses were written moving the sample at 500 µm/s, which gave a total fabrication time ranging from three and an half hours to almost seven, depending on the complexity of the design. When there is a second spherical cap, its base radius is always 200 µm.

The fabrication procedure was completed, after the writing step, with a post-exposure bake of 10 minutes at 95 °C. Each sample was then slowly immersed into SU-8 developer for repeated intervals of 30 s, usually three. After a quick microscope inspection this last operation could be reiterated to thoroughly clean the samples. Then, to polymerize the resin sealed inside the lens cap, I irradiated each lens with an Hamamatsu 365 nm

UV lamp for about two minutes and then proceeded with a second PEB, identical to the previous one. Since a SEM inspection requires the lens gold coating, I needed to perform the focal length characterization first. After that I coated the sample with a 20 nm - 25 nm thick gold layer, depending on the case, which proved to be enough for an optimal image acquisition. Fig. 4.24 shows the detail of a sample with five lenses after the UV

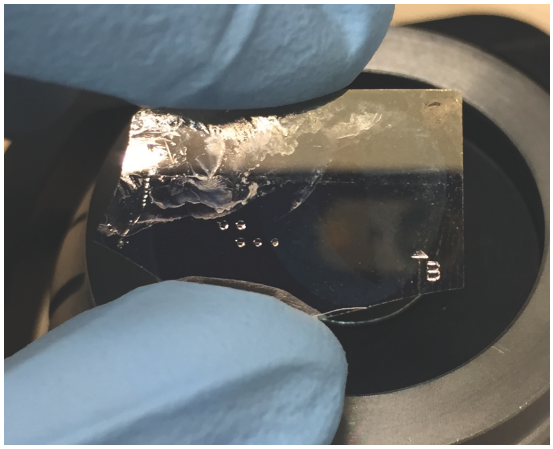


Figure 4.24: Sample with five lenses before gold coating.

lamp irradiation and baking.

L1 was fabricated with one spherical cap, intersected by grids running alternatively along the x and y planes. The extra lines had to be written every other plane and the whole fabrication needed lower power than in the other cases, because this scanning pattern accumulated a lot of heat inside the resist. Fig. 4.25 shows an optical microscope 4x image of the lens after development. Measuring the lens diameter I obtained a value of 500.4 µm,

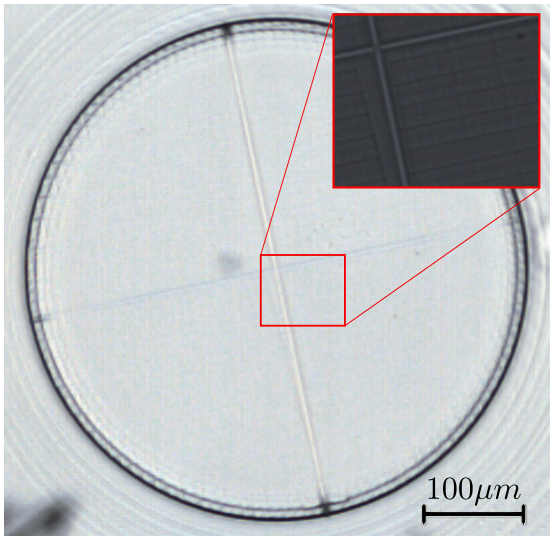


Figure 4.25: L1 optical microscope image.

an almost perfect match to the design specification. This was verified for all the other

fabrications. Alas, the characterization procedure for L1 showed no ability of the lens to form a focus. This has been most probably caused by the square pattern visible in Fig. 4.25, which is due to the supporting planes.

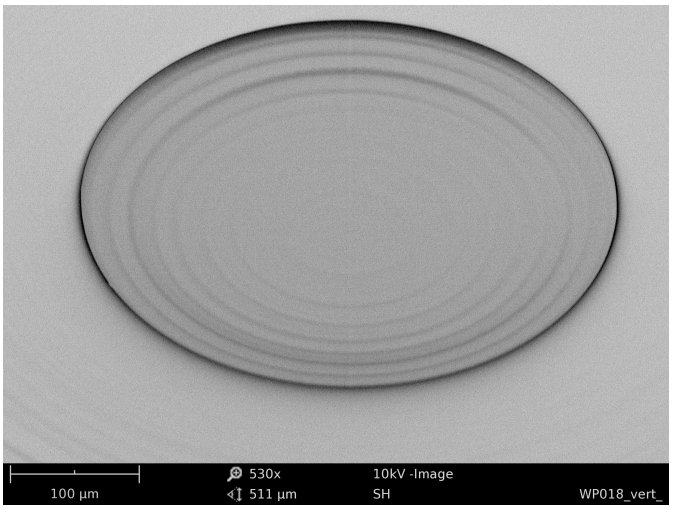


Figure 4.26: L2 SEM image.

L2 was fabricated with two spherical caps and four passes of the laser beam for each circle that composes the lens. It underwent a second PEB of 20 minutes at 200 °C, aimed to verify if this procedure, called *thermal reflow*, could improve the lens performances. As a matter of fact the focused spot appeared more defined after the bake than before,

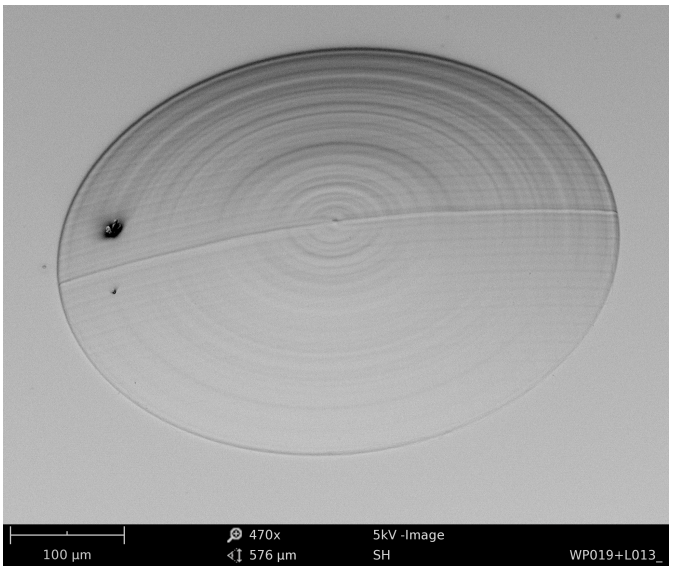


Figure 4.27: L3 SEM image.

and the surface evenness seems better than for the other lenses, as is noticeable from Fig. 4.26. Also the typical diametrically crossing lines, which appear in each fabrication due to the laser shutter opening and closing time intervals, appear softened. I could not have

access to a profilometer to verify the actual roughness improvement associated with the additional PEB.

L3 features two spherical caps and three passes of the laser beam on the circular patterns. As Fig. 4.27 shows, the surface does not look as smooth as it should be for an optical component. It is divided into concentric sections that could be the result of the structure gradually collapsing during fabrication or development.

L4 (Fig. **??**) is a variation of the L1 design, with a tighter $r_{spacing}$ to compensate for the absence of the horizontal planes.

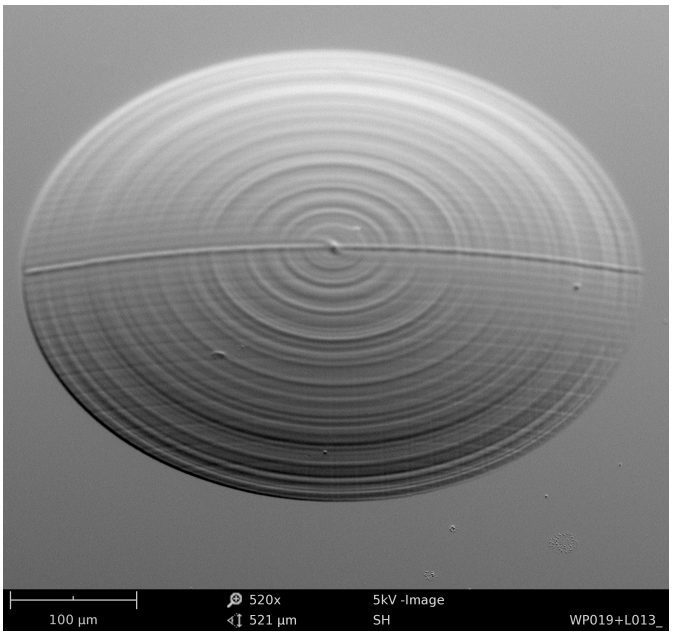


Figure 4.28: L4 SEM image.

To characterize the focusing properties of the fabricated lenses I used a HeNe laser, which has $\lambda_{HeNe} = 632.8nm$. The focal length is measured by focusing the collimated laser beam with the lens and then imaging the focal spot on a white screen with a 20x objective. Then, translating the sample closer to the objective, the surface of the lens itself is imaged on the screen. The distance between the two positions provides the focal length of the lens. Fig. 4.29 shows a step of the characterization procedure.

All the lenses were designed to have a focal length of 1.17 mm. During characterization all of them revealed a bright focal point, followed and preceded by multiple foci of lesser brightness, which are probably caused by the surface unevenness that is visible in the SEM images of all the lenses. Since the lens thickness is quite small, it makes sense that the superficial roughness has a considerable effect on the final behavior. Table 4.4 summarizes the focal lengths measured for the fabricated lenses.

Those results have led me to consider that the structures might be behaving as Fresnel

Lens code	Measured focal length [mm]
L2	1.28
L3	1.82
L4	1.18

Table 4.4: Focal lengths of the fabricated lenses.

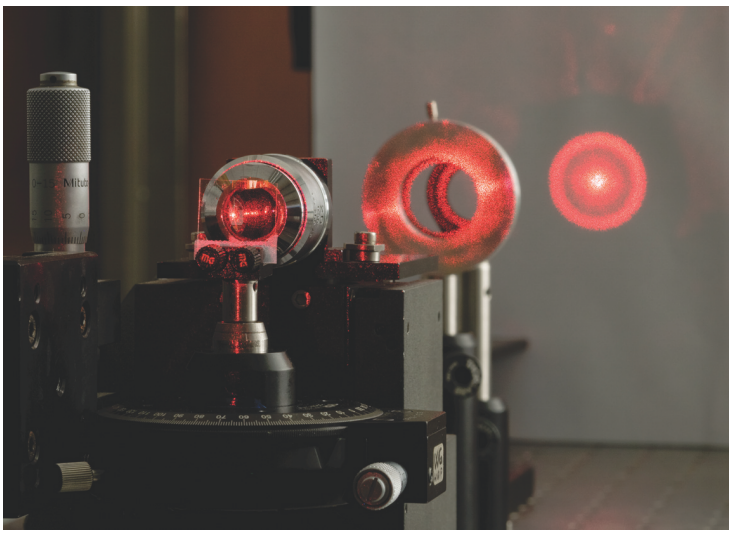


Figure 4.29: L4 characterization. The lens is visible on the cover glass where it has been fabricated.

lenses, both because of the presence of multiple foci and for the peculiar surface shape. An ideal Fresnel lens divides the volume of an equivalent refractive lens into infinite concentric annular sections, and reduces the total thickness by just preserving the curved edge of each section. Fig. 4.30 compares the cross sections of a conventional plano-convex lens and its equivalent Fresnel implementation. An actual Fresnel lens could very well have a

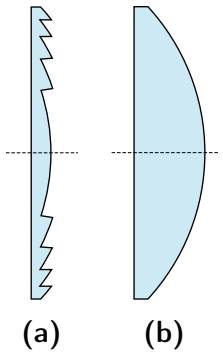
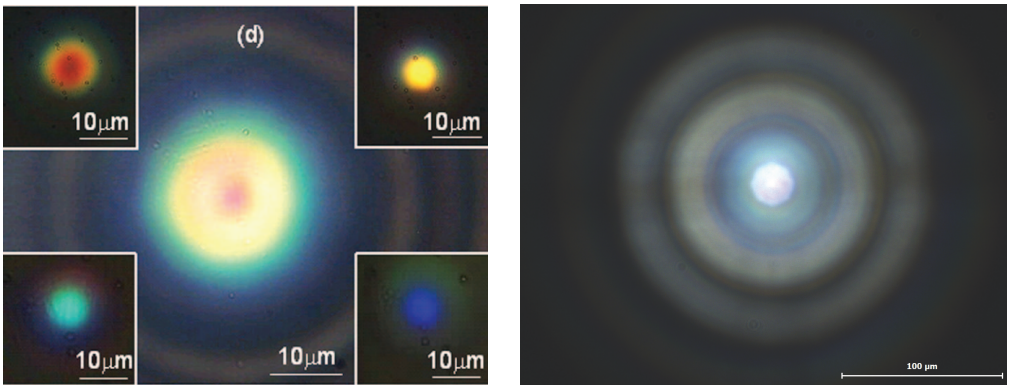


Figure 4.30: Cross sections of a spherical Fresnel lens (a) and of an equivalent conventional spherical plano-convex lens (b).

thickness close to the theoretical sagittal height of these structures, which from geometrical

considerations is about 16 µm.

The similarity is even stronger considering the chromatic aberration, also called *color fringing*. Chromatic aberration is an optical effect in which there is a failure of a lens to focus all colors to the same convergence point. It occurs because the refractive index value of a material decreases for increasing wavelengths. Therefore, different wavelengths are focused at different points on the optical axis and/or in the focal plane, resulting in fringes of color surrounding objects, especially in high-contrast situations. Fresnel lenses have stronger chromatic aberration than simple refractive lenses do, because of the focal length dispersion characteristic of the structure. This design pays the reduction in lens thickness with a degradation of the imaging quality of the lens, and is therefore rarely used for imaging purposes. By observing the light focused by the L3 lens with a microscope in light transmission mode, one can notice the strong resemblance with the chromatic aberration of a Fresnel lens. A comparative example can be seen in Fig. 4.31. Note how



(a) **(b)**

Figure 4.31: White light focused by a Fresnel lens [3] (a) and L3 (b). Focal length dispersion separates the different spectral components.

the lens is actually focusing the shape of the microscope light aperture.

In conclusion, the single cap lens proved to be the closest to the design requirements, although the presence of chromatic aberration makes it not optimal for imaging applications. Thermal reflow is also helping with the lens behavior, and an even longer hard bake could be an important step to add to such a large fabrication. The double cap strategy instead seems to add more irregularities to the lens surface, and therefore a worse performance in terms of focal length. Considering that the lenses have been fabricated with a base diameter which is half the required size, the experimental results suggest that fabrication of the complete design could prove challenging.

A viable alternative could be an optimized implementation of a Fresnel lens instead, that would not have the stability issues that were of concern in this design. In this case the resolution becomes of the uttermost importance, because the rings that compose the

Fresnel lens zones need to be precisely defined. An oil immersion objective would then be the best choice for the fabrication. This design would strongly reduce the total fabrication time, but would also probably require some optimization steps to minimize the chromatic aberration in the fluorescence emission windows.

Conclusions and future developments

In this thesis I have presented my work on the characterization and optimization of the STMicroelectronics four channels real-time PCR Point of Care handheld device called Q3-Plus. The field of POC devices is relatively new but highly promising, because of a growing demand of such a fast and patient-centered diagnostic solution.

In the first part of my work I assembled two experimental setups to characterize the optical subsystem of the Q3-Plus, which is deputed to perform the fluorescence detection during a PCR reaction. An extended analysis of the spectral behavior of the optical section led to the discovery of manufacturing defects that were undermining the device performances. This study brought to structural modifications of the optical subsystem that led to much better performances and opened the way for further optimization. A performance analysis of the redesigned system allowed the reduction of the total number of optical components of the exciting section, which is particularly relevant for the filters since those have a considerable weight on the final device cost. The optimization process was completed with a redesign of the mechanical parts that encase the optical components.

The new generation of the STMicroelectronics real-time PCR instrument features the addition of a fifth channel and the introduction of a new HDR imaging sensor, while also expanding the operating wavelengths towards the near infra-red (NIR) region. I demonstrated the feasibility of those modifications with a new set up that showed fluorescence reading in the visible and IR spectra. The successful proof of concept was followed by the modeling of an optical lens that could be paired with the sensor. A lens with a 1.17 mm focal length and a 1.967 mm curvature radius has been chosen, based on the model results. Those specifications did not have a commercially available match in the optics industry, so I looked for an alternative to fabricate it with a prototyping technology.

I decided to fabricate an SU-8 plano-convex lens with the two-photon polymerization (2PP) technology, which allows direct laser writing of complex micron- and nano-structures inside a photosensitive resin. The characterization process suggested the employment of a microscope objective with 63x magnification and a 0.68 NA. Here I presented four lenses

with different designs to test the structure stability. It emerged that a single cap lens is the best solution, since it has no structural disadvantages with respect to a double cap design and has better performances, close to the theoretical behavior. A thermal reflow process proved to improve the results. From the characterization emerged the presence of multiple foci for each lens. By microscope inspection the smoothness of lens surfaces resulted to be insufficient, showing a circularly patterned roughness probably derived from the fabrication strategy. The hypothesis is that, because of this roughness, the lens acts similarly to a Fresnel lens. Considering that the lenses have been fabricated with a base diameter which is half the required size, those results suggest that the complete design could prove challenging. On the basis of the obtained results, a viable alternative to a traditional lens design could be the realization of an equivalent Fresnel lens, that would need some optimization but has the structural advantage of being composed of simple circular plates with small vertical dimensions.

The insight gained from this experimental activity has paved the way for the future development of the new real-time PCR system, that will need to go through some further optimizations before becoming a fully functional industrial product.

Bibliography

[1] J. Evans. BCC Research - Point of Care Diagnostics, Nov 2016. vi, 26, 28

[2] A. Merlini. Fabbricazione di lenti micrometriche tramite polimerizzazione a due fotoni, 2016. x, 90

[3] Q. D. Chen, H. B. Sun, et al. Phase lenses and mirrors created by laser micronanofabrication via two-photon photopolymerization. *Applied Physics Letters*, October 2007. xi, 104

[4] J. M. S. Bartlett and D. Stirling. *PCR Protocols*. Methods in Molecular Biology. Humana Press Inc., 2nd edition, 2003. 4

[5] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, 4th edition, 2002. 4

[6] A. Chien, D. B. Edgar, and J. M. Trela. Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *Journal of Bacteriology*, 127:1550–1557, September 1976. 8, 11

[7] Smithsonian Institution Archives. History of the Polymerase Chain Reaction Videohistory Collection 1992-1993, 2016 (accessed November, 2016). `http://siarchives.si.edu/collections/siris_arc_217745`. 8

[8] C. M Leutenegger. The real-time taqman pcr and applications in veterinary medicine. *Veterinary Sciences Tomorrow*, January 2001. 19

[9] International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 15 February 2001. 24

[10] J. B. Jackson, S. Y. Kwok, J. J. Sninsky, J. S. Hopsicker, K. J. Sannerud, F. S. Rhame, K. Henry, M. Simpson, and H. H. Jr. Balfour. Human Immunodeficiency Virus Type 1 Detected in All Seropositive Symptomatic and Asymptomatic Individuals. *Journal of Clinical Microbiology*, 28(No.1):16–19, January 1990. 24

[11] New England Biolabs, Inc. PCR Troubleshooting Guide, 2016 (accessed November, 2016). `https://www.neb.com/tools-and-resources/troubleshooting-guides/pcr-troubleshooting-guide`. 25

[12] World Health Organization. The availability and use of diagnostics for HIV: a 2012/2013 WHO survey of low- and middle-income countries., December 2014 (accessed November, 2016). `http://apps.who.int/iris/bitstream/10665/147213/1/9789241507905_eng.pdf`. 26, 27

[13] E. Wynberg, G. Cooke, S. Amir, S. D. Reid, and N. Ford. Impact of point-of-care CD4 testing on linkage to HIV care: a systematic review. *Journal of the International AIDS Society*, January 2014. 27

[14] A. P. Dhawan, W. J. Heetderks, M. Pavel, S. Acharya, M. Akay, A. Mairal, B. Wheeler, C. C. Dacso, T. Sunder, N. Lovell, M. Gerber, M. Shah, S. G. Senthilvel, M. D. Wang, and B. Bhargava. Current and Future Challenges in Point-of-Care Technologies: A Paradigm-Shift in Affordable Global Healthcare With Personalized and Preventive Medicine. *IEEE Journal of Translational Engineering in Health and Medicine*, March 2015. 27

[15] J. Howick, J. W. L. Cals, et al. Current and future use of point-of-care tests in primary care: an international survey in Australia, Belgium, The Netherlands, the UK and the USA. *BMJ Open*, 2014. 28

[16] Semrock. Semrock Introduces BrightLine® Fluorescence Filter Sets for Quantum Dots, 2016 (accessed March, 2017). `https://www.semrock.com/PressReleases.aspx?id=39`. 46

[17] T. Erdogan and V. Mizrahi. Thin-Film Filters Come of Age. *Photonics Spectra*, July 2003. 46

[18] H. A. Macleod. *Thin-Film Optical Filters*. Institute of Physics Publishing, 3rd edition, 2001. 67

[19] Turan, E. Optical Filters: Non-normal Angles of Incidence, 2011 (accessed March, 2017). `https://www.semrock.com/Data/Sites/1/semrockpdfs/nonnormalanglesofincidence.pdf`. 67

[20] M. Malinauskas, V. Purlys, M. Rutkauskas, and R. Gadonas. Two-photon polymerization for fabrication of three-dimensional micro- and nanostructures over a large area. *Proceedings of SPIE*, 2009. 84

[21] Y. Liu, L. Pyrak-Nolte, and D. Nolte. General 3d microporous structure fabricated with two-photon laser machining. *Proceedings of SPIE*, 6886, 2008. 85

[22] C. N. LaFratta, J. T. Fourkas, and T. Baldacchini. Multiphoton fabrication. *Angewandte Chemie (International ed. in English)*, 2007. 86, 87

[23] J. Serbin, A. Egbert, A. Ostendorf, et al. Femtosecond laser-induced two-photon polymerization of inorganic–organic hybrid materials for applications in photonics. *OSA*, 28(5), 2003. 88

[24] M. Malinauskas, A. Žukauskas, G. Bičkauskaitè, R. Gadonas, and S. Juodkazis. Mechanisms of three-dimensional structuring of photo-polymers by tightly focussed femtosecond laser pulses. *OSA*, 18(10), May 2010. 88

[25] W. H. Teh et al. Effect of low numerical-aperture femtosecond two-photon absorption on (su-8) resist for ultrahigh-aspect-ratio microstereolithography. *Journal of Applied Physics*, 2005. 88

[26] K. Ohlinger, Lin Y., Z. Poole, and K. P. Chen. Undistorted 3d microstructures in su8 formed through two-photon polymerization. *OSA*, 2011. 88

[27] S. Maruo, A. Takaura, and Y. Saito. Optically driven micropump with a twin spiral microrotor. *OSA*, 17(21), 2009. 88

[28] G. Rui, X. Shizou, Z. Xiaomin, et al. Micro lens fabrication by means of femtosecond two photon polymerization. *OSA*, January 2006. 89

[29] M. Malinauskas, H. Gilbergs, A. Žukauskas, et al. A femtosecond laser-induced two-photon photopolymerization technique for structuring microlenses. *Journal of Optics*, February 2010. 90