UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

The application and analysis of single-molecule spectroscopy data in photosynthesis

by

Joshua L. BOTHA

13290152

Supervisor: Prof. Tjaart P.J. KRÜGER

Submitted in fulfilment of the requirements for the degree

Philosophiae Doctor (Physics)

in the Faculty of Natural and Agricultural Sciences
University of Pretoria

9 December, 2024

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

UNIVERSITY OF PRETORIA

Abstract

Faculty of Natural and Agricultural Sciences

Physics

Philosophiae Doctor

The application and analysis of single-molecule spectroscopy data in photosynthesis

by Joshua L. BOTHA

Most of the energy that sustains life on earth can be traced back to photosynthesis, and so its importance is difficult to overstate. Our understanding of this crucial process is still growing as we develop ever-advanced techniques to investigate the world of proteins and their interactions with light. Single-molecule spectroscopy (SMS) has emerged as a transformative technique for studying molecular systems at the nanometre scale, offering unparalleled insights into dynamic processes and heterogeneity at the individual molecule level. However, effective utilisation of SMS data often requires sophisticated analytical tools and stable experimental setups, which can pose significant challenges to researchers. This thesis addresses these challenges through the development of Full SMS, a versatile analysis software suite and a custom-designed sample cell holder, which together provide an integrated framework to advance SMS-based studies.

Full SMS is a comprehensive graphical user interface (GUI)-driven software package for processing and analysing multiparameter SMS datasets. These datasets typically include fluorescence intensity and lifetime, as well as spectral properties of single dye molecules, quantum dots, or fluorescent active biomolecules. The software enables detailed analysis through tools for statistical evaluation of fluorescence intensity, clustering, and level identification; lifetime decay fitting; and second-order correlation function calculations. Visualisation features include the display of fluorescence intensity and spectral traces, as well as raster-scan images, with robust filtering options to tailor data processing. A custom HDF5-based file structure ensures efficient storage, while flexible export capabilities facilitate integration into broader research work flows. Written in Python, Full SMS is open source and accessible to users without programming expertise, leveraging a multiprocess architecture for enhanced computational performance. Full SMS is used to analyse three sample types as case studies to showcase its utility and ease-of-use.

Complementing the software, this thesis also presents a custom-designed sample cell holder, which extends the capabilities of the SMS experimental setup designed and assembled in-house at the University of Pretoria. This hardware innovation enhances stability and versatility, allowing for more robust and reproducible single-molecule measurements under a wider range of experimental conditions. Together, these advancements provide a powerful platform for SMS, bridging the gap between complex experimental data and resulting insights, and significantly contributing to the field of single-molecule biophysics.

Declaration of Authorship

I, Joshua L. BOTHA, declare that the thesis, which I hereby submit for the degree *Philosophiae Doctor* at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signed: Joshua

Date: 24th of March, 2025

Acknowledgements

To Professor Tjaart P.J. Krüger, my supervisor, I want to offer my deepest and sincerest *thank you* for all the guidance and support, patience and encouragement through the years. It has been an honour and privilege to learn from, and be led by you. You are an inspiration to the group you lead, to your students, and to me. I am convinced that I would not have been able to achieve what I have here without you.

To the Physics department, colleagues, and friends, your friendship and support proved to be a vital ingredient throughout this process. A special thanks to the Head of Department, Professor Chris Theron, for your belief in me and for your willingness to go above and beyond since I first joined the Physics department. To Bertus Van Heerden, thank you for joining me in the final stretches and being instrumental in getting this across the line. Likewise, to Towan Nöthling, words fail me. Thank you for every messy soft-serve ice cream on the benches outside our department, every late night call and shared experience. The journey meant more because you were part of it.

To my dearest family. It has been a hard road to get this point, and I know that without a shadow of a doubt that I would never have got here without you. My father Leon Botha, my mother Karin Botha, my sister Jenilee Boshoff, my brother-in-love Freddie Boshoff, and my uncle Claude Botha, amongst the many other incredible members of my family. From the bottom of my heart: thank you.

Kirsten Charisse Botha, my wife. My word *"...slip, slide, perish..."* under the weight of love I place on them. In you I have the fullness of the *"rose garden"*, for I have not a single timeless moment, but all of them. I have not only *"echoe[s]"* only disturbed by memory, but a lifetime of song in itself endless and eternal. No one will ever know in full the sacrifices you made for and with me to have these words written on these pages. And that is good, for they are our own, as our story is our own. But here I thank you, with my every atom. Thank you.

Lastly, I thank my God, my Saviour Jesus Christ. Every good and precious thing I have comes from Him, including whatever ability I had to produce this thesis.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Contents

Abstract **i**

Declaration of Authorship **iii**

Acknowledgements **iv**

1 Introduction **1**

1.1 Photosynthesis 3

1.2 Light-Harvesting Complex II 6

1.3 Photophysical Processes 7

1.4 This Thesis 11

2 Single Molecule Spectroscopy **13**

2.1 Theory 14

2.1.1 Diffraction Limit 14

2.1.2 Confocal Microscopy 15

2.2 Experimental Setup 16

2.2.1 Light Source 17

2.2.2 Acousto-Optic Tunable Filter 18

2.2.3 Spatial Filter 18

2.2.4 Dichroic Beam Splitter 19

2.2.5 Objective Lens 20

2.2.6 Translation Stage 20

2.2.7 Fluorescence Filter 21

2.2.8 Detectors 22

2.2.9 Single Photon Avalanche Diode 22

2.2.10 Diffraction grating 23

2.2.11 Electron Multiplying CCD 23

2.3 Sample Cell Holder 25

2.4 Materials 31

2.4.1 Quantum Dots 31

2.4.2 Alexa 31

2.4.3 Light Harvesting Complex II 31

3 Analysis: Theory **33**

3.1 Introduction 33

3.2 Change Point Analysis 35

3.2.1 Photon-by-Photon Log Likelihood Test 36

3.2.2 Recursive Segmentation 40

3.3 Intensity Level Grouping 40

3.3.1 Agglomerative Hierarchical Clustering 41

3.3.2 Expectation Maximisation Clustering 42

3.3.3 Number of States 44

3.3.4 Global Grouping 44

4 Analysis: Software **46**

4.1 Overview of *Full SMS* 46

4.2 Analysis 48

4.2.1 Intensity Level Resolution 49

4.2.2 Intensity-Level Grouping 55

4.2.3 Fluorescence Decay Fitting 60

4.2.4 Second-Order Correlation Function 62

4.2.5 Data Filtering 64

4.2.6 Display of Spectra 64

4.2.7 Display of Raster Scan Images 65
4.3 Additional Tools 66
4.3.1 Measurement Data Conversion 66
4.3.2 Trace Region of Interest (ROI) 67
4.3.3 Photon Burst Detection 68
4.4 Application Architecture 68
4.4.1 General Architecture 68
4.4.2 Multiprocessing 69
4.4.3 Distribution and Compatibility 70
4.4.4 Measurement File Structure 70
HDF5 Format 71
4.4.5 HDF5 for SMS 73
4.5 Data Export 75
4.6 Conclusion 77

5 Case Studies **79**
5.1 Analysis of Light Harvesting Complex II 79
5.2 Comparative Study 80
5.3 Fluorescence Intermittency Models 83
5.4 Conclusion 87

6 Summary **88**

A Code Sections **90**
A.1 Run Change Point Analysis 90
A.2 Find Change Points 91
A.3 Find Segment 92
A.4 Weighted Likelihood Ratio Test 93
A.5 Run Grouping 95
A.6 Agglomerative Hierarchical Clustering 96

A.7 Expectation Maximisation 97
A.8 BIC 99

B Graphical User Interface **100**

C Sample Cell Holder **105**

Bibliography **111**

List of Figures

1.1 Photosynthesis overview 4
1.2 LCHII structure 7
1.3 Jablonski diagram of LHCII 9

2.1 Confocal principle 16
2.2 Single Molecule Spectroscopy setup 17
2.3 Sample cell holder overview 24
2.4 Exploded view of sample cell holder 25
2.5 Sample cell holder chamber bottom 26
2.6 Sample cell holder chambers 28
2.7 Sample cell holder chamber being filled with liquid 28
2.8 Sample cell holder with gas flow 29

4.1 Flowchart of change point analysis process 50
4.2 Decision tree for change point segmentation 51
4.3 Example of intensity trace 52
4.4 Change point detection breakdown 53
4.5 Flowchart of grouping process 56
4.6 Flowchart of grouping calculations 57
4.7 Example of intensity level grouping 59
4.8 Example of lifetime fit 61
4.9 Parameters for lifetime fitting 62
4.10 Example of antibunching analysis 63
4.11 Data filtering example 65

4.12 Example of spectra 66
4.13 Exporting example 76

5.1 Distribution of number of photons in lifetime fit 80
5.2 Intensity distributions of LHCII 81
5.3 Sample analysis comparison 82
5.4 Progression of intensity distribution 84
5.5 Distributions of 2-state and 3-state models 85
5.6 Switching rate model comparison 86

B.1 Intensity trace example of Alexa 100
B.2 Two-channel intensity trace example 101
B.3 Group selection example using the groups shown in Fig. 4.7. The group's intensity, dwell time, and number of photons are displayed. 101
B.4 Global grouping example 102
B.5 Example of a raster-scan image 102
B.6 Screenshot of the trace trimming tool dialog 103
B.7 Screenshot of the photon burst dialog 104

Chapter 1

Introduction

This chapter establishes the foundational context required to understand the research presented in this thesis. It is divided into two main sections. The first gives a broad introduction to the research area of this work: photosynthetic energy dynamics. The second addresses the interaction between light and matter in the context of photosynthesis, with emphasis on the fundamentals of fluorescence.

The nature of light and its interaction with matter has intrigued scientists since the mid-16th century [1, 2]. In 1865, Maxwell's equations unified the concepts of electricity and magnetism, providing a comprehensive mathematical framework for understanding light as an interplay of electromagnetic waves [3]. Five years earlier, Kirchhoff and Bunsen demonstrated that elements could be identified by their unique spectral properties [4]. Einstein's introduction of stimulated emission in 1916 [5] ultimately enabled Maiman's construction of the first laser in 1960, revolutionising the field of spectroscopy [6].

Three decades later, the first single molecules were measured—in what are now known as single-molecule spectroscopy (SMS) experiments—by Moerner and Kador (based on absorption) [7] and Orrit and Bernard (based on fluorescence) [8], both working at ultra-cold temperatures. Shortly thereafter, room temperature detection of single chromophores was pioneered by the Keller group [9–11], after demonstrating the same approach to a single multichromophoric light-harvesting protein complex in 1987 [12]. Often referred to as "single-particle spectroscopy" when not strictly limited to a single molecule, SMS is applicable to nanoscale systems comprising one or a limited number of quantum emitters. The level of selectivity and sensitivity of SMS techniques provides access to numerous properties of nanoscale emitters that are usually masked in conventional ensemble averaging techniques, offering a wide range of applications in physics, chemistry, and biology [13]. For example, time-dependent SMS measurements provide information about the dynamics and kinetics of nanoscale systems and can resolve time-dependent processes without the need to synchronise these processes in a large collection of identical molecules. Today, SMS techniques are broadly applied across fields of biophysics, chemistry, and materials science, enabling detailed studies of processes like protein folding, enzyme function, and photosynthesis at the single-molecule scale. Techniques like real-time single particle tracking [14] and single-molecule Förster resonance energy transfer [15] have become essential tools for real-time investigations of molecular dynamics, conformational changes, and molecular interactions.

The biological applications of SMS are particularly vast and include observations of protein conformational dynamics [16], enzyme reactions [17, 18], transcription in single DNA molecules [19, 20], light-harvesting complexes switching between different functional states [21–24], changes in the oligomeric states of macromolecules [25], and changes in the diffusive states of nanoparticles [26]. Real-time 3D tracking of biological (macro)molecules allows SMS to be performed with high spatio-temporal resolution in live cells [27].

Applications in non-biological settings include the use of quantum dots [28], single organic molecules [29], and nitrogen vacancy centres in diamond [30] in photonic quantum technologies and quantum sensing.

Two typical features exhibited in the photon emission from individual nanoscale emitters are photon antibunching and fluorescence intermittency. These phenomena have been used to verify the presence of a single quantum emitter but have also found interesting applications. Photon antibunching can be used to count the number of fluorescent labels in confocal microscopy, investigate multiple coupled emitters (by analysing higher-order photon correlations), and quantify photophysical processes through the width of the antibunching dip [31]. Although fluorescence intermittency (often called ‘blinking’) tends to be a hindrance in most applications, it provides a source of information to improve spatial resolution beyond the diffraction limit [32, 33] and—for light-harvesting complexes—is a useful means to study photoprotection [34, 35].

1.1 Photosynthesis

Photosynthesis occurs in plant cells within organelles known as chloroplasts (see Figure 1.1), which are characterised by their complex internal membrane structures. The chloroplast stroma is a fluid matrix that houses thylakoids, which are organised as bilayered lipid membrane structures enclosing the thylakoid lumen. Advanced imaging techniques, such as electron microscopy, have shown that thylakoid membranes form two distinct structural domains: grana (stacked thylakoid membranes) and stroma lamellae (unstacked thylakoid membranes connecting the grana) [36].

Photosynthesis involves two major phases: the light-dependent and light-independent reactions. During the light-dependent phase, photons are absorbed by approximately 200 pigment molecules within each photosystem, photosystem I (PSI) and photosystem II (PSII). Upon light absorption by PSII, photon energy is transferred between pigments via excitation energy transfer (EET) until it reaches the reaction centre (RC), where a specialised pair of pigments becomes excited and releases a free electron. This electron is subsequently transported across the membrane, resulting in charge separation that prevents recombination. The vacancy created by the electron is filled through a series of redox reactions involving the oxidation of the oxygen-evolving complex (OEC). With each successful photon absorption, an electron is extracted from the OEC. After four successful electron extractions, the

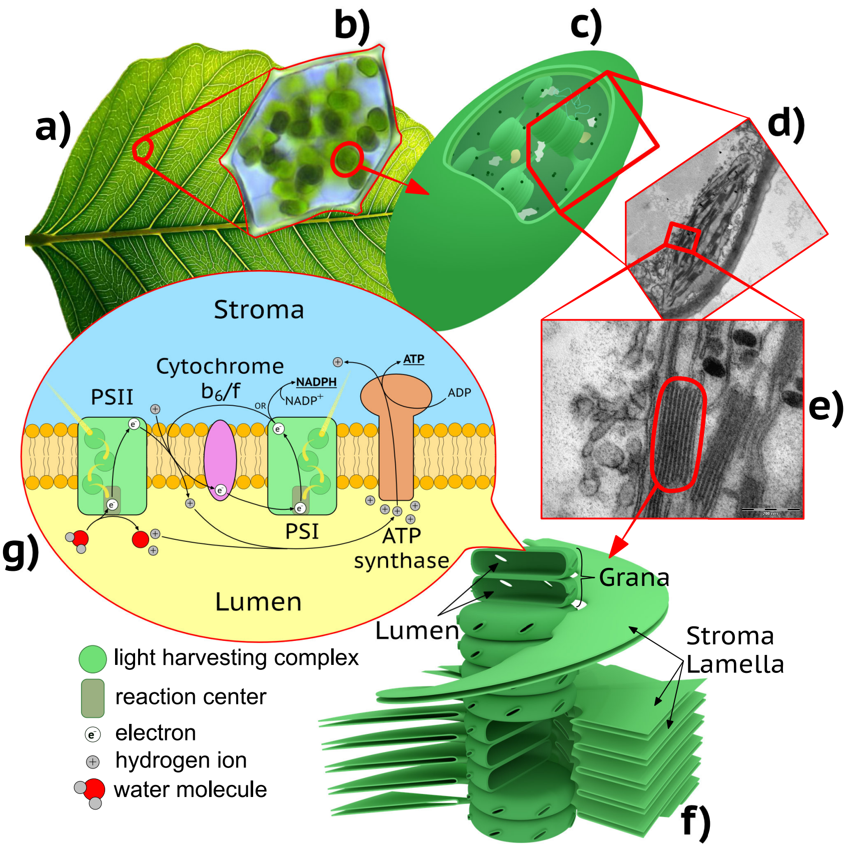


Figure 1.1: The journey from the macroscopic to the heart of photosynthesis: beginning with a) a photo of a leaf, to b) a plant cell, to c) an illustration of a chloroplast, to d) a 12000x and e) a 30000x transmission electron microscope image of a chloroplast, to f) an illustration of a thylakoid, and finally to g) an illustration of a thylakoid membrane. Sections of the stroma lamellae in f) have been cut out to reveal the manner in which they connect to the grana stacks. Illustration g) includes membrane proteins involved in the light-dependent steps in photosynthesis, such as photosystem I & II that contain light harvesting complexes and reaction centres. Note that the illustration does not hold true to the actual stoichiometric factors. Image b) was adapted from Kristian Peters, images d) and e) were adapted from 'and3k' and 'caper437', and c) and f) were adapted from Kelvinsong, all under CC BY-SA 3.0.

OEC catalyses the reaction of two water molecules to yield four hydrogen ions and an oxygen molecule, represented as follows:

$$\text{photon} + \text{OEC} \longrightarrow e^- + \text{OEC}^+$$

$$\text{photon} + \text{OEC}^+ \longrightarrow e^- + \text{OEC}^{+2}$$

$$\text{photon} + \text{OEC}^{+2} \longrightarrow e^- + \text{OEC}^{+3}$$

$$\text{photon} + \text{OEC}^{+3} \longrightarrow e^- + \text{OEC}^{+4}$$

followed by,

$$2\,H_2O \xrightarrow{\text{OEC}^{+4}} 4\,H^+ + O_2$$

or, in short,

$$4\,\text{photons} + H_2O \xrightarrow{\text{OEC}} 4\,H^+ + 2\,O_2 + 4\,e^-$$

In the light-independent phase, the adenosine triphosphate (ATP) and nicotinamide adenine dinucleotide phosphate (NADPH) generated in the stroma are used to fix carbon and create carbon-carbon bonds in carbohydrates through the Calvin-Benson cycle. This means that the energy captured by photosynthesis is temporarily stored in the form of sugars and then ultimately stored as starches, which are subsequently transported and used throughout the plant.

Given that photosynthetic organisms are optimised for survival, they are, in most cases geared towards low-light conditions, during which the photoelectron quantum efficiency approaches nearly 100%. Consequently, in high-light conditions, the rate of photon absorption may exceed the rate of energy utilisation, leading to the formation of highly reactive chlorophyll triplet states that react with molecular oxygen to produce toxic reactive oxygen species (ROS) [37]. To prevent damage from ROS, oxygenic photosynthetic organisms are equipped with processes, collectively termed photoprotection, which occur over various timescales—ranging from seconds to hours or even years—to mitigate the formation of these harmful radicals. One of the key mechanisms of photoprotection is non-photochemical quenching (NPQ), specifically the energy-dependent component known as qE, which dissipates excess excitation energy on ultrafast timescales after absorption [38].

1.2 Light-Harvesting Complex II

Light-Harvesting Complex II (LHCII) serves as the principal light-harvesting component in the photosynthetic systems of green plants, algae, and cyanobacteria. This complex is pivotal in capturing solar energy and facilitating the transfer of the subsequent excitation energy within PSII. As the Earth's most prevalent membrane protein assembly, LHCII is vital for the conversion of solar energy into chemically utilisable forms, making it a significant subject of study within biophysics, structural biology, and bioenergetics.

LHCII comprises a trimeric pigment–protein structure—illustrated in Figure 1.2—embedded in the chloroplast thylakoid membrane [39], where each monomeric subunit features a polypeptide backbone tethered with non-covalently bound pigments: 8 chlorophyll a pigments, 6 chlorophyll b pigments and 4 carotenoid pigments. The chlorophylls are mostly responsible for light absorption across a wide visible spectrum, whereas carotenoids enhance light capture and provide photoprotection. These pigments collaboratively enable LHCII to effectively capture and direct energy towards the PSII reaction centre via neighbouring light-harvesting complexes.

Energy transfer takes place within and between neighbouring LHCII via Förster resonance energy transfer and eventually reaches the PSII reaction centre. The chlorophyll and carotenoid molecules are carefully positioned within LHCII to form energy delocalisation states, referred to as excitons [40], which contribute significantly to highly efficient energy transfer [41]. Beyond energy transfer, LHCII is

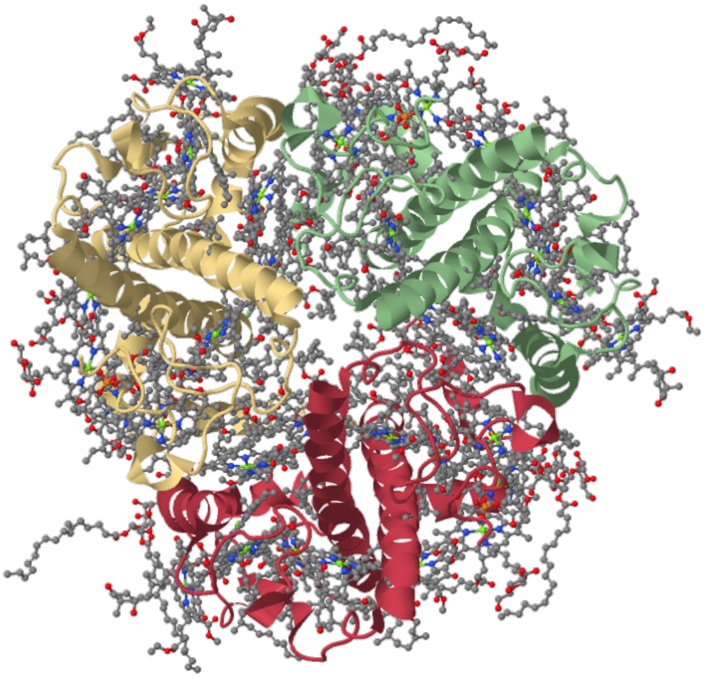


FIGURE 1.2: Light harvesting complex II as resolved by [39]—available on the Protein Data Bank under ID 1RWT—and rendered by JMol. Each colour ribbon represents a monomeric protein scaffold. The remaining ball-and-stick molecules represent the chlorophyll and carotenoid pigments.

critically involved in the *qE* component of NPQ, and is thus responsible for the dissipation of surplus energy as heat under intense light conditions. The reversible transition of LHCII from a light-harvesting to an energy-dissipating state underscores its dual role in energy transfer and photoprotection, allowing plants to adjust rapidly to varying light conditions.

1.3 Photophysical Processes

Absorption and fluorescence represent two fundamental optical phenomena that manifest when molecules interact with light. These processes underpin numerous spectroscopic methodologies employed to examine molecular architecture, motion, and interactions across diverse disciplines, encompassing physics, chemistry, biology, and materials science. Notably, fluorescence spectroscopy has emerged as an invaluable tool in biophysical research, facilitating the exploration of molecular environments, conformational shifts, and interactions at both the collective ensemble and individual single-molecule scales.

Absorption occurs when a molecule residing in its ground electronic state absorbs a photon whose energy precisely corresponds to the energy difference between the ground state and an excited electronic state. The vibrational state of ground and excited states also contributes to the energy differences that are available for absorption, as explained by the Franck-Condon principle. Some electronic transitions, however, are not allowed; the spin selection rule only allows transitions in which the total spin state is conserved. Additionally, the Laporte's rule prohibits transitions between states with the same parity in systems that have a centre of symmetry. Absorption spectra, depicting the probability of absorption as a function of the photon's energy (often expressed in units of wavelength), reveal information such as electronic energy levels, chemical compositions, and molecular conformations.

After the excitation of a molecule, it may revert to the ground state by the spontaneous emission of a photon, which is known as fluorescence. If N molecules are excited at the same time, the rate at which the number of molecules in the excited state, S_1, move into the ground state, S_0, due to spontaneous emission is given by,

$$\frac{\delta N(t)}{\delta t} = -A_{S_1 \to S_0} N(t).$$

$A_{S_1 \to S_0}$ is known as the rate of spontaneous emission and is governed by the electronic configuration of the excited molecule and the interaction of the excited electrons with the molecule's nuclear motions as well as vibrations of the molecule and its environment. In cases where this process is the only radiative process, $A_{S_1 \to S_0}$ is the same as the radiative emission rate k_R (see Figure 1.3). Assuming $N(0)$ to be the initial number of excited molecules, the above equation can be solved to give the

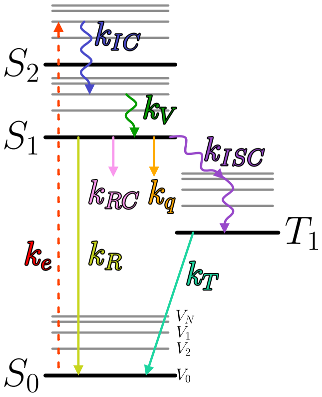


FIGURE 1.3: This figure, which is called a Jablonski diagram, illustrates the potential pathways for the relaxation processes of excited states within LHCII. In this context S_0, S_1 and S_2 denote the electronic ground state, the first and the second electronically excited state, respectively. Each electronic state has vibrational states associated to it, as denoted by $V_0, V_1, \ldots, V_N$ and illustrated by the grey lines. The parameter k_e refers to the rate of excitation introduction into the system. k_{IC} refers to the energy relaxation process known as internal conversion from one state into another, and k_V refers to the rate at which vibrational relaxation occurs. k_R refers to the radiative emission rate, manifesting as fluorescence, and k_{ISC} describes the rate of energy-induced electron spin flips, known as triplets and denoted by state T_1. The decay of triplet states occurs at a rate specified by k_T; this rate influences the repopulation rate of S_0 rather than affecting the depopulation rate of S_1. Furthermore, k_q represents the rate of quenching of excitations due to qE, and k_{RC} refers to the rate at which energy transfer occurs to a neighbouring system which eventually results in the energy being transferred to the reaction centre of a photosystem complex.

following equation:

$$N(t) = N(0)e^{-k_R t} = N(0)e^{-t/\tau_f}, \tag{1.1}$$

where τ_f is known as the fluorescence lifetime and is defined as $\tau_f = 1/k_R$.

In most molecules, fluorescence competes with several nonradiative mechanisms, such as internal conversion and intersystem crossing. Internal conversion is the process in which a system undergoes excitation relaxation from a low vibrational state in a high electronic state, to a high vibrational state in a low electronic state, culminating in vibrational relaxation to the lowest vibrational state. Intersystem crossing involves transitioning to a triplet state, potentially leading to phosphorescence rather than fluorescence. These processes have been shown to be sensitive to environmental conditions, such as solvent polarity, pH, temperature, and any competing process, making fluorescence a keen probe for the local environment surrounding the fluorophore.

When a photon is absorbed by one of the pigments in LHCII, the excitation quickly relaxes to the site possessing the lowest energy and remains there until it undergoes one of several potential de-excitation pathways.

The mean rate at which excitations reach a particular fate (illustrated in Figure 1.3) is described by the following rate equation:

$$k_L = \overbrace{\underbrace{k_R + k_{ISC} + k_{IC} + k_q}_{\text{Isolated}} + k_{RC}}^{\text{Coupled to neighbouring complexes}}$$

where $\tau_L = 1/k_L$ denotes the lifetime of excitations. The rate k_R describes how quickly excitations are dissipated via radiative processes. Intersystem crossing, which involves the spin flip of electrons leading to an electronic triplet state, occurs at a rate denoted by k_{ISC}. Additionally, the rate of non-radiative de-excitation via vibrational levels, while maintaining spin, is represented by k_{IC}. These three intrinsic processes—k_R, k_{ISC}, and k_{IC}—are inherently tied to the structural constitution of the complex, rendering their average rates effectively constant. The quenching rate associated with the qE process, if present in the system, is represented by k_q. The transfer of energy to the reaction centre occurs at a rate known as k_{RC}. Moreover, in isolated LHCII complexes where k_{RC} is absent, indirect quantification of the system's quenching rate (k_q) can be resolved by measuring the effective fluorescence lifetime (τ_L). This can be done through the use of time-correlated single photon counting (TCSPC), where the distribution of the relative time between excitation by, and re-emission of, photons is measured for multiple events. This resulting decay curve can then be fit to determine fluorescence lifetime.

Notably, the quenching channel linked to qE in the system is frequently toggled on and off. This reversible quenching is a phenomenon known as fluorescence intermittency [34], which manifests itself as abrupt and significant changes in the fluorescence intensity of isolated LHCII complexes during continuous illumination. Originally documented in single dye molecules [42], fluorescence intermittency, also termed blinking, was subsequently observed in quantum dots [43]. Multiple mechanisms have been discovered and suggested in various systems [44–47] Intriguingly, more intricate systems such as pigment-protein complexes—LHCII as an example—exhibit similar behaviour [48].

1.4 This Thesis

This thesis describes how I expanded the capabilities of SMS by designing and prototyping a Sample Cell Holder, a novel physical component that enhances the experimental setup by allowing precise control and manipulation of the environment in which the samples are measured. This development enables an enriched range of measurements that were unattainable with the previous configuration. The enhanced control provided by the Sample Cell Holder not only extends the scope of observable phenomena in SMS but also opens avenues for more nuanced investigations into photosynthetic complexes at the single-molecule level.

Second, I have developed specialised software tailored to analyse data generated by SMS experiments, called Full SMS. This software incorporates multiple advanced analytical methods, including change point analysis, level grouping, fluorescence lifetime fitting, and antibunching correlation studies. It provides an efficient and streamlined solution for organising and processing SMS data, significantly improving both the speed and accuracy of the analysis.

To validate its efficacy, I applied this software to three different fluorescent systems, with a particular focus on LHCII. Through a secondary analysis, I demonstrate the utility of the software's grouping algorithm, showcasing its capability to reveal intricate patterns within these systems.

In summary, this thesis contributes both a physical enhancement in the form of the Sample Cell Holder to the SMS setup, as well as a robust analytical software suite called *Full SMS*, thereby advancing both experimental and computational approaches in the study of photosynthetic mechanisms. The integration of the Sample Cell Holder and the specialised software marks a significant step forward in enabling more comprehensive investigations into the complex dynamics of photosynthetic systems at the single-molecule level.

Chapter 2

Single Molecule Spectroscopy

Throughout the years, significant strides have been made in elucidating the mechanisms of photosynthesis, prompting the development of more sophisticated methodologies. Traditionally, ensemble measurements yield aggregate data, potentially masking the intricacies of the processes involved and the diversity in the dynamics of the system being studied. However, single molecule spectroscopy (SMS) stands out as a powerful tool, permitting the examination of individual systems which has led to the discovery of many observations that would otherwise not have been possible [16–20, 22–27]. This technique can therefore highlight the variations and patterns in the behaviour of single molecules, which are pivotal in comprehending the quenching phenomena within of Light-Harvesting Complex II (LHCII). SMS facilitates the observation of unique molecular attributes, which may differ greatly from one another and offer indispensable insights into their specific functional contributions. This chapter delves into a brief review of the foundational theoretical concepts, the inherent challenges faced, and the detailed experimental methodologies pertinent to SMS.

Additionally, I introduce and discuss my contribution of a newly designed and prototyped Sample Cell Holder, which provides precise control and manipulation of the environment in which the samples are measured.

2.1 Theory

In order to fully comprehend the principles underlying SMS, it is essential to examine the restrictions presented by the intrinsic properties of light; hence, this section begins with the exploration of the diffraction limit as a fundamental challenge to achieving high spatial resolution. Subsequently, it delves into descriptions of confocal microscopy and its significant contribution toward enhancing the quality of the signal as a methodology for minimising noise within measurement processes. Additionally, the discussion extends to cover various crucial methods for further signal optimisation and strategies for reducing background noise. All of these considerations are vital to ensure the precise detection of single-molecule events, especially within increasingly complex environments.

2.1.1 Diffraction Limit

The Abbe diffraction limit, first introduced by Ernst Abbe in the year 1873 [49], delineates the theoretical limitation on the degree to which a light beam can be focused. In the field of microscopy, this limitation results in the formation of a diffraction pattern rather than an infinitesimally small point, with the smallest discernible feature of a circular aperture being known as the Airy disk. The size of the Airy disk is determined by its radius[1], which can be calculated using the equation

$$r = 1.22\frac{\lambda}{2 \cdot n \cdot \sin\theta} = \frac{\lambda}{2\,\mathrm{NA}}, \tag{2.1}$$

where λ represents the wavelength of the light, n is the refractive index of the medium, θ is the semi-aperture angle of the lens system, and the term $n\sin(\theta)$ is known as the numerical aperture (NA). Contemporary oil-immersion microscope objectives often achieve NA in the range of 1.4 to 1.6. For a light wavelength

---

[1] The radius definition used here is the distance from the point of maximum brightness to the first minima. There are however other definitions, one being the distance from the point of maximum power to half power, in which case the factor of 1.22 is then 1.02.

of 630 nm, the smallest radius of the Airy disk that can be practically attained is approximately 260 nm. Given that proteins generally measure only a few nanometres across, they must be spaced apart by at least several hundred nanometres to be effectively resolved using traditional spectroscopic methods that rely on visible light. This intrinsic limitation holds particular significance in SMS, where it dictates the ability to distinguish individual molecular entities, thus necessitating the employment of advanced techniques to transcend these inherent physical barriers. Numerous super-resolution methodologies, including techniques such as STED (Stimulated Emission Depletion) microscopy and PALM/STORM (Photoactivated Localisation Microscopy, and Stochastic Optical Reconstruction Microscopy), have been innovatively designed to circumvent these constraints, thereby achieving resolutions that surpass the diffraction limit. As such, these advanced methods enable the detailed observation of single molecules at an enhanced spatial resolution, which is pivotal for certain cutting-edge applications in the realm of microscopy.

2.1.2 Confocal Microscopy

The process of observing the fluorescence from a single protein encounters specific hurdles primarily due to the inherently low light intensity it emits in comparison to the surrounding background noise. In order to enhance the signal-to-noise ratio (SNR), confocal microscopy is applied, which involves directing excitation light through a pinhole aperture, permitting only light emanating from the focal volume to pass through (see Figure 2.1). This meticulous focusing facilitates the distinction of fluorescence signals from a single molecule against the background, which is a critical requirement in SMS. By effectively eliminating out-of-focus light, confocal microscopy markedly diminishes background noise, thereby improving the sensitivity of detection for individual fluorophores. Moreover, the technique plays a pivotal role in minimising photobleaching by confining the volume of light exposure, which is vital for generating accurate and reliable time-resolved measurements in SMS. By

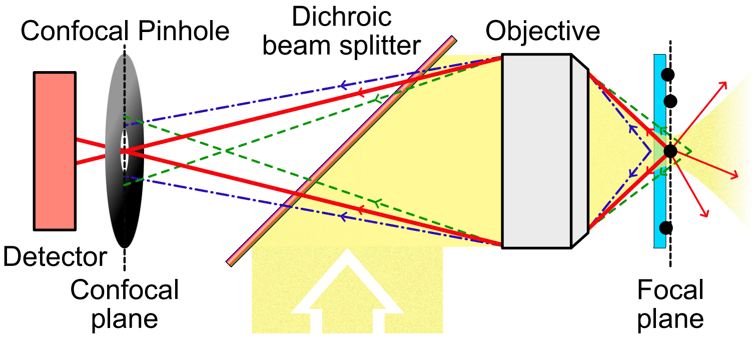


FIGURE 2.1: An illustration of confocal microscopy and how it prohibits light from being detected that does not originate on the focal plane of the objective.

restricting excitation to a smaller region, the likelihood of fluorophores transitioning into a non-fluorescent state is significantly minimised, resulting in more stable and consistent fluorescence readings.

A further notable benefit of utilising confocal microscopy lies in its capacity for optical sectioning, allowing for the reconstruction of three-dimensional structures through the acquisition of images across a series of focal planes. This capability proves particularly advantageous when examining biological samples where individual molecules may be located within larger constructs.

In the context of SMS, the application of confocal microscopy can be further integrated with time-correlated single-photon counting (TCSPC) to assess fluorescence lifetimes, thereby offering enhanced insights into the local environmental conditions and the dynamics of energy transfer.

2.2 Experimental Setup

The SMS experimental setup has been meticulously designed and built—which started during my MSc and has continued on during my PhD—to conduct a thorough examination of isolated LHCII complexes. This sophisticated setup comprises several critical components, each playing a pivotal role in ensuring the attainment

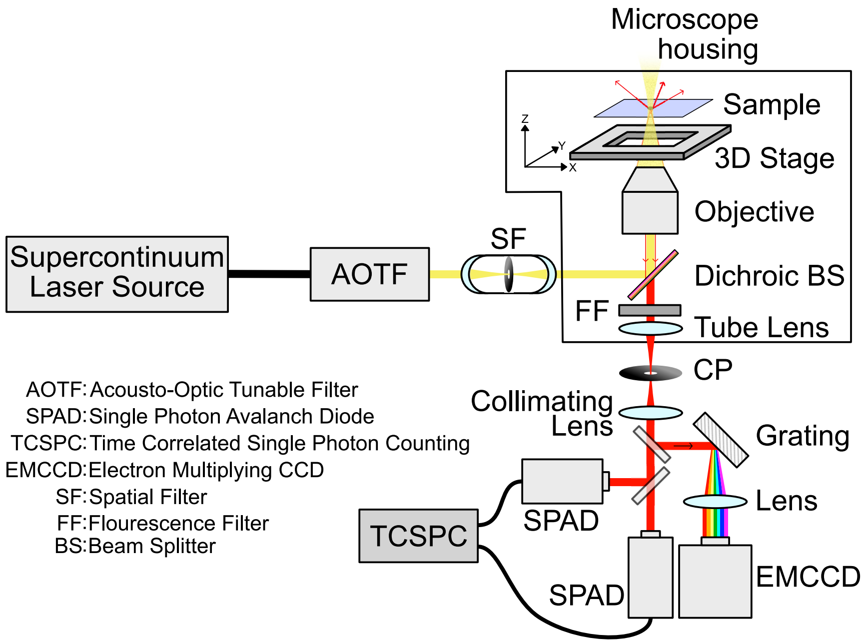


FIGURE 2.2: A diagram of the single molecule spectroscopy setup.

of measurements with high sensitivity, precision, and accuracy. The setup incorporates the following elements, each contributing to the detection and measurement of single molecule fluorescence, achieving high spatial and temporal resolution.

2.2.1 Light Source

The system employs a Fianium Supercontinuum laser (C400-4-PP), which is capable of emitting light across a broad spectrum ranging from 420 nm to 2200 nm, with a pulse width of 6 ps. It offers a selectable pulse repetition rate varying between 0.1 MHz and 40 MHz. This supercontinuum laser provides an extensive range of excitation wavelengths, facilitating the selective excitation of diverse fluorophores within LHCII. By offering the capacity to fine-tune the pulse repetition rate, it allows for the optimisation of excitation conditions tailored to specific experimental objectives, thereby significantly augmenting its versatility for SMS applications. Furthermore, utilising a pulsed laser source permits precise timing of photon emission events, an aspect critical for accurate fluorescence lifetime measurements, and the elucidation of the dynamic behaviour of single molecules. The versatility offered by the supercontinuum laser renders it highly suitable for a spectrum of applications ranging from spectral characterisation to lifetime measurement.

2.2.2 Acousto-Optic Tunable Filter

The Acousto-Optic Tunable Filter (AOTF) serves as a key device for selecting specific excitation wavelengths from a broadband laser source by leveraging the Bragg diffraction principle intrinsic to an acousto-optic crystal. This component grants precise control over the diffraction angle, thus facilitating the simultaneous manipulation of up to eight wavelengths within the spectral range extending from 400 nm to beyond 650 nm. Such meticulous control is critical for the selective excitation of distinct chromophoric entities, thereby significantly reducing undesired background noise. The AOTF's capability for rapid wavelength switching is advantageous in multi-wavelength experimental setups and is instrumental in conducting spectral characterisations of LHCII.

This capability is critical for probing energy transfer pathways within LHCII, as the various pigments within the complex display absorption and emission at different wavelengths. Targeted excitation of these pigments allows for a deeper understanding of the energetic dynamics inherent to the photosynthetic machinery. Furthermore, the AOTF not only permits wavelength selection but also the modulation of the intensity of each chosen wavelength, facilitating carefully controlled excitation conditions.

An alternative—albeit less convient and flexible—option is to make use of bandpass filters, which filters out all but a narrow band of the excitation light.

2.2.3 Spatial Filter

To achieve a spatially uniform Gaussian intensity profile, a spatial filter (SF) is employed to eliminate an optical aberrations, as well as higher-order components of the laser light. By focusing the laser beam through a pinhole, the spatial filter effectively removes higher-order modes, purifying the beam's spatial profile. This ensures that the excitation efficiency is maximised, while minimising intensity fluctuations that can detract from the precision of fluorescence measurements. Enhancing the uniformity of the excitation beam, the spatial filter is instrumental in securing reliable and consistent data during SMS experiments. Spectral artefacts can introduce extraneous noise into measurements and complicate the accurate quantification of fluorescence signals at the single-molecule level. By effectively mitigating these artefacts, the spatial filter contributes to a cleaner excitation beam, enhancing the reliability and precision of the resulting fluorescence measurements.

2.2.4 Dichroic Beam Splitter

The dichroic mirror, which often serves the role of a beam splitter, is designed to reflect wavelengths of shorter nature while allowing longer wavelengths to pass through unimpeded. The selection of a suitable cutoff wavelength for the dichroic mirror is crucial, as it ensures a clear differentiation between the excitation and emission light, thereby significantly reducing any potential interference or overlap. This effective separation is critical for achieving high sensitivity in detecting fluorescence from individual molecules, as it aids in isolating the weak fluorescence signal from the substantially stronger excitation light. The efficiency of the dichroic mirror has a direct influence on the sensitivity of the entire system; even a minimal amount of leakage from the excitation light can overshadow the fluorescence signal. Beyond merely separating excitation and emission light, the dichroic mirror also plays a role in guiding the emitted fluorescence to the appropriate detectors. Depending on the particular experimental requirements, this emitted light may be routed towards a spectral detector for wavelength analysis or a time-correlated single-photon counting (TCSPC) detector aimed at enabling lifetime measurements. The adaptability of the dichroic mirror facilitates diverse detection methodologies, enabling an extensive investigation of the photophysical properties of single molecules.

2.2.5 Objective Lens

The Nikon CFI Plan Apochromat Lambda 100 × oil immersion objective lens, possessing a numerical aperture (NA) of 1.45, is proficient in concentrating the light beam to a diffraction-limited focal point with an approximate radius of 220 nm at a wavelength of 630 nm. This substantial NA is essential for maximising the effectiveness of light collection and achieving a minimal focal volume, thereby improving the spatial resolution of the SMS setup. The use of oil immersion further optimises the objective's NA by aligning the refractive index between the sample and the objective, consequently reducing light refraction and thus enhancing resolution.

Furthermore, objective lens can introduce optical aberrations into the laser beam. By employing an apochromatic design the objective ensures that minimal chromatic and spherical aberrations are caused, thereby guaranteeing that the excitation light is sharply focused and that the resultant emitted fluorescence is effectively captured. This aspect is particularly significant within the field of SMS, where even slight deviations in the focal volume can lead to substantial variations in the detected fluorescence signal.

2.2.6 Translation Stage

The sample is loaded into a sample cell holder (see Section 2.3) that is affixed to a Mad City Labs Nano-LPS200 nanopositioning stage, which facilitates movement with a precision of 0.4 nm in the X, Y, and Z axes. Control over the positioning stage is accomplished through USB communication, thereby enabling the precise placement of the sample within the focal volume. The stage's high precision is crucial for the accurate localisation and measurement of individual LHCII complexes while preventing interference from neighbouring molecular entities. Such capability proves to be particularly advantageous during time-resolved measurements or when delineating the spatial distribution of complexes over a surface. Additionally, the 3D positioning stage supports automated scanning of the focal plane, thus allowing for the identification of the positions of measurable sample particles. Measuring multiple particles is crucial to accumulate sufficient statistical information regarding the behaviour of individual LHCII complexes, as data can be gathered from a substantial number of molecules. The capacity to scan the sample in a three-dimensional manner also assists in the development of high-resolution spatial maps of fluorescence intensity, yielding insights into the organisation and diversity of LHCII complexes present on a surface.

2.2.7 Fluorescence Filter

Subsequent to the emission of fluorescence, the light is collimated by the objective and transmitted through the optical system. The light is then focused by the tube lens onto a confocal pinhole (see Section 2.1.2), after which another lens is used to collimate the light once more. The collimated beam is then passed through one or a combination of several fluorescence filters, which are optical components meant to absorb any remaining excitation light while transmitting light emitted due to fluorescence. The cutoff wavelength of the filters are chosen to be as close to the excitation wavelength as possible. The effectiveness of these filters directly influences the SNR of measurements, since even minor leakage can introduce considerable noise. By utilising superior quality fluorescence filters, the experimental setup ensures that the detected signal predominantly originates from the targeted fluorophores, resulting in more precise and dependable data outcomes.

2.2.8 Detectors

In the study of emitted fluorescence, several types of detectors are used. For the SMS setup built at the University of Pretoria, the Single Photon Avalanche Diode (SPAD) and the CCD Camera are used due to their high sensitivity. The SPAD is used to measure fluorescence intensity and—when coupled with a TCSPC module—fluorescence lifetime, which provides insight into the energy dynamics of the system. The CCD Camera, on the other hand, is used to measure the spectra of the fluorescence emitted after it has been spatially separated as a function of wavelength through the use of a diffraction grating. Spectral measurements can unveil any shifts in peak wavelengths or variations in spectral profiles, which may indicate fluctuations in energy transfer routes, or reveal changes within the chromophores' immediate surroundings.

2.2.9 Single Photon Avalanche Diode

The Single Photon Avalanche Diode (SPAD) is designed for high-sensitivity photon counting. Utilising a Micro Photon Device SPAD (PD-050-CTE), which is Peltier-cooled and maintains a dark count rate of 10.5 counts per second, this detector has a time resolution of 128 ps—as measured by the instrument response function—which is sensitive enough to study a wide range of molecular dynamics. Such capability supports TCSPC measurements, which are essential for evaluating the fluorescence lifetime and dynamic behaviour of LHCII complexes. This lifetime analysis reveals behaviour relevant to the rates of radiative and non-radiative decay mechanisms (as discussed in Section 1.3), which underpin the understanding of energy transfer efficiency and the occurrence of quenching phenomena in photosynthetic complexes.

2.2.10 Diffraction grating

The emitted fluorescence is also dispersed into varying wavelengths by implementing a diffraction grating. Using the Optometrics 3-8880 grating—which is fine-tuned for 800 nm and features 830 grooves per millimetre with a diffraction efficiency nearing 70% across the 700 nm to 900 nm region—facilitates spectral separation of the fluorescence. This allows for concurrent measurements across multiple emission wavelengths. Such capability is particularly beneficial for examining spectral variance among individual LHCII complexes, yielding insights into their inherent energy transfer mechanisms.

2.2.11 Electron Multiplying CCD

The Andor iXon3 897 EMCCD camera is used for the detection of diffracted light and leverages electron multiplication technology to enhance sensitivity to single photons. The device is equipped with a sensor array measuring 512 by 512 pixels, each pixel having a size of 16 μm, and operates with reduced background noise by cooling the sensor to a temperature of -100°C. This is achieved by a combination of a Peltier cooler and a water cooling system and is critical for reducing dark noise, an essential factor in SMS measurements, where the fluorescence signals emitted by individual molecules are typically very weak. The EMCCD is particularly suited for capturing weak fluorescence signals owing to its electron multiplication process, which boosts the signal before it is read out, which minimises the effect of readout noise. This feature is crucial in SMS, where single molecules may emit very few photons. By enhancing the signal, the EMCCD camera guarantees that even dim fluorescence events are detected accurately, allowing for the examination of individual molecules under conditions of low excitation intensity. Besides its exceptional sensitivity, the EMCCD camera also boasts a rapid readout speed compared to other CCD cameras, which is advantageous for obtaining time-resolved fluorescence data. Moreover, the ability to capture time-resolved data facilitates the exploration of fluorescence intermittency, a phenomenon where a single molecule's fluorescence toggles between on and off states.

Utilising both the CCD Camera and SPAD allows a complementary approach to analysing the fluorescence characteristics of LHCII complexes. While the CCD Camera offers detailed spectral insight, the SPAD contributes precise temporal data, together enabling a holistic analysis of the complexes' photophysical properties. This dual detection methodology is advantageous for investigating the dynamic processes within single molecules, as it integrates spatial and temporal resolution facets.

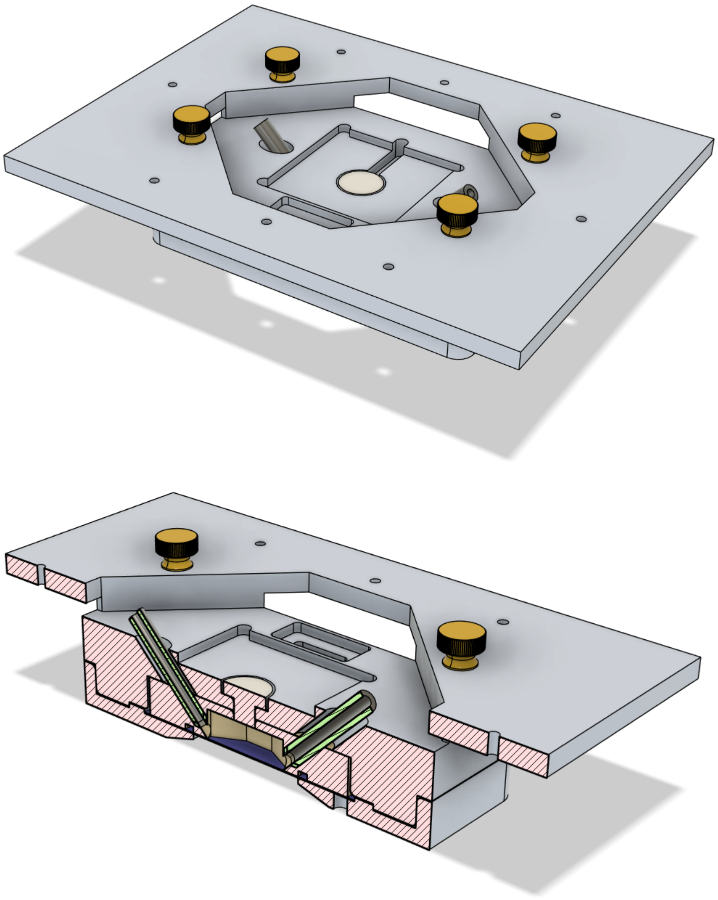


FIGURE 2.3: An isometric view of the sample cell holder model. The top panel shows the full model, whereas the bottom panel shows a half-section view of the model.

2.3 Sample Cell Holder

One of the important parts of an SMS setup is the housing in which the sample is loaded. Depending on the requirements of the study being done, it needs to fulfil many roles, stability being first and foremost. Another important consideration has to do with the environment the sample needs to be kept in. If the sample's nature is dry (i.e. spin coating, non-aqueous thin films, etc.) most of the requirements are simpler. However, if the sample is to be kept in an aqueous medium and if the environment is not to be static and needs to be changed before or during the measurement, the sample cell holder (SCH) needs to be able to facilitate this. The device provided with the stage was very limited in its ability to meet requirements of the latter kind (which will be expanded on below). For this reason, a new SCH

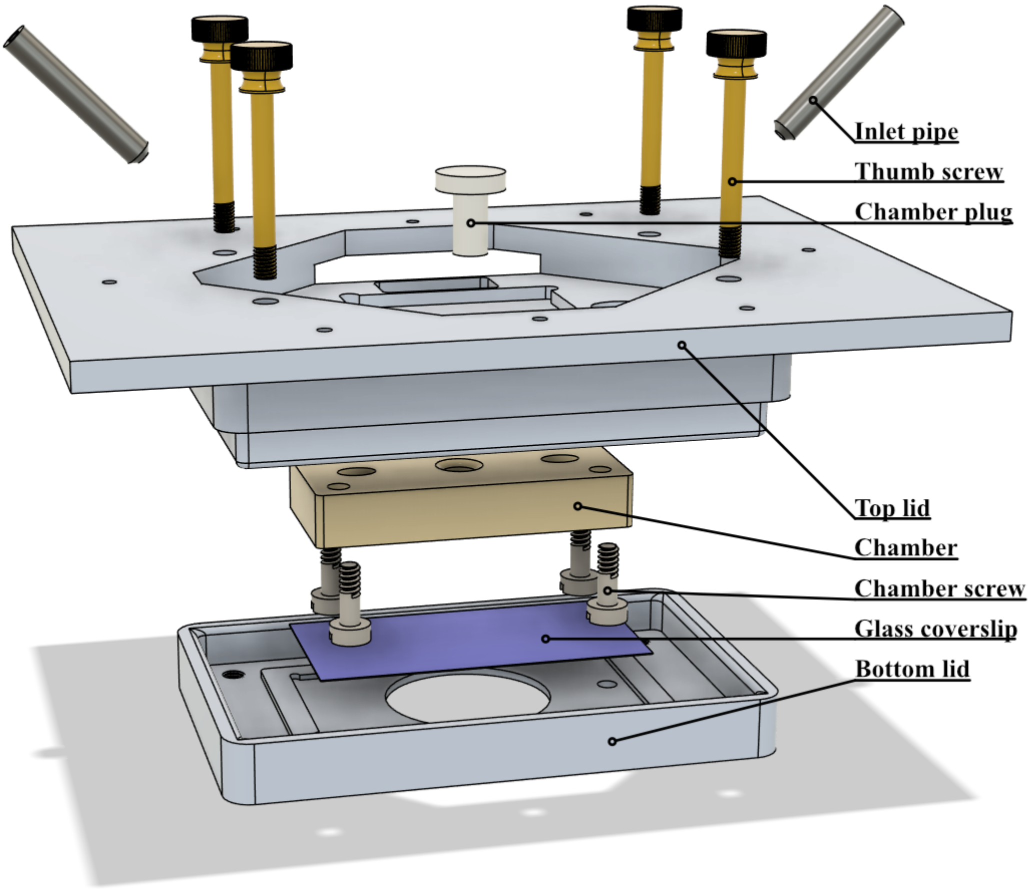


FIGURE 2.4: Exploded view of the sample cell holder and all the parts it consists of.

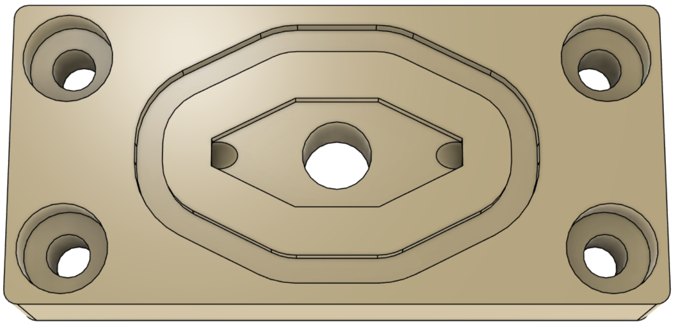


FIGURE 2.5: A bottom view of the 1 mm chamber which shows the shape of the volume that it creates, the inlets and the o-ring slot.

was designed and prototyped.

The primary requirements and limitations of the new SCH were as follows:

- The SCH needed to provide structural stability to avoid translational movements during measurements.
- It needed to allow the introduction of new—and consequently the ejection of the previous—chemical environment, and had to do so in a manner that would prevent the samples from being washed away.
- With the previous item in mind, it needed to provide a controlled and isolated environment that would prevent contamination or introduce unwanted external factors.
- It had to have a sufficiently large thermal reservoir to combat temperature changes due to the introduction of new chemical environments.
- It had to allow the flow of a nitrogen gas[2] over the aqueous medium from before or during the whole measurement.
- It had to facilitate as many kinds of studies and future expansions as possible, and for that reason, would be well suited to a modular design.

---

[2] The introduction of nitrogen into the SCH chamber, either for an incubation period or for the duration of the measurement, was to achieve an oxygen-scarce environment.

- It had to remain within the confines of what the in-house machinists could achieve given the time and tools at their disposal, as well as the inherent limitations provided by the materials used.
- The SCH also needed to be easy to use.
- The major limitation that these requirements had to adhere to was size-related: the stage provided length and breadth limitations, and the distance to the object provided depth limitations.

Given these requirements, during the course of my studies I developed the prototype that will be shown below.

The SCH was modelled with Autodesk Fusion 360 which, at the time, provided a reasonable free academic licence. An isometric view of the model can be found in Figure 2.3. The figure also shows a half-section of the model to provide some insight into how the model achieves the design requirements.

The SCH design consists of the following parts, as illustrated in Figure 2.4, and their primary functions:

- **Bottom lid**: To seat the coverslip and align it with the rest of the SCH.
- **Glass coverslip**: To provide optical access to the sample.
- **Chamber screw(s)**: To affix the chamber to the top lid.
- **Chamber**: Together with the coverslip, it defines the volume in which the sample will be housed.
- **Top lid**: Aligns with and fastens to the stage.
- **Thumb screw(s)**: Aligns with, and screws into the bottom lid.

Due to the design, when the thumb screws are tightened, they exert a compression force between the chamber and the glass coverslip. This design ensures that the compression is localised between the coverslip and the chamber, rather than causing

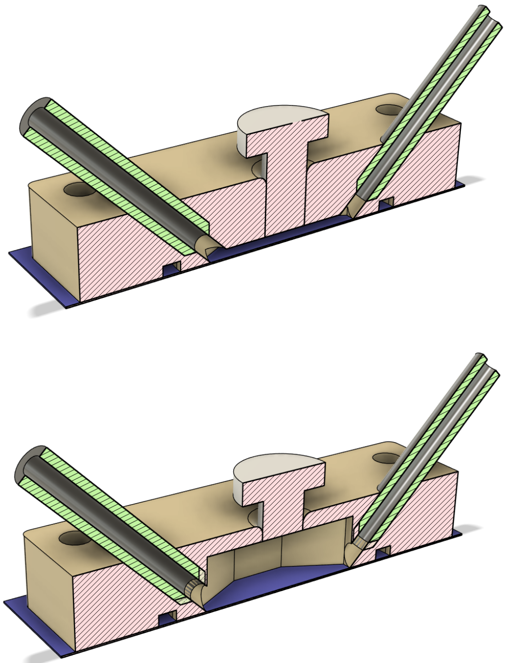


FIGURE 2.6: Two modular chambers that have been designed and prototyped. The top chamber has a sample volume defined by a height of only 1mm, which is intended to be used in cases where no gases are being used during measurement. The bottom has a large sample volume, specifically for cases where a gas flow is to be introduced.

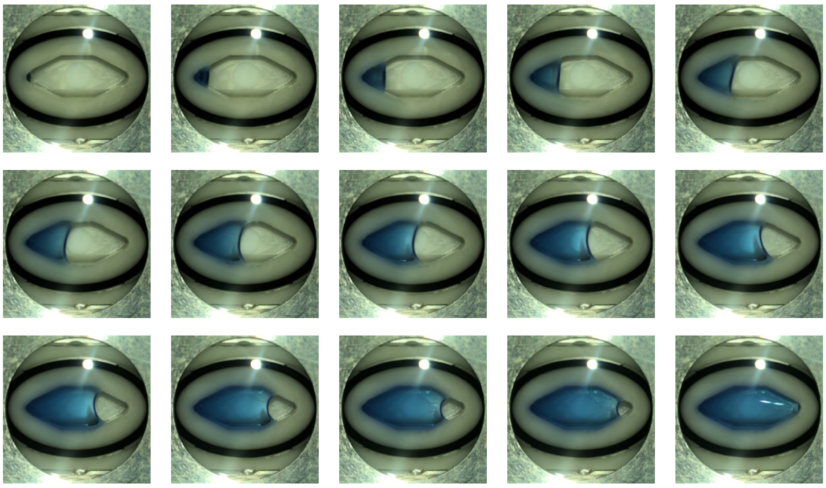


FIGURE 2.7: A series of photos, starting with the top left and progressing to the right and from top to bottom, that shows the chamber being filled with liquid. This serves to show that the shape of the chamber ensures that the liquid fills it evenly and without air pockets.

the top lid to press directly against the bottom lid. Together with the o-ring housed in the chamber, shown in Figure 2.5, a stable and sealed interface is created at the coverslip, allowing for effective isolation of the chamber contents without requiring excessive force between the top and bottom lids.

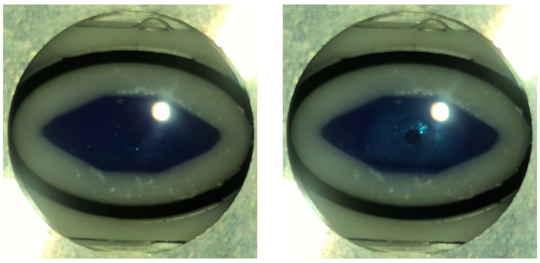


FIGURE 2.8: Photo of the sample cell holder as the gas chamber was filled with gas. The photo on the left was taken before the gas was introduced, and the photo on the right was taken while a constant flow of gas was being introduced into the sample volume. The presence of gas is indicated by the slight lightening of the coloured fluid. This shows that the gas flows into and out of the chamber without displacing all the liquid.

There are multiple ways in which the chemical environment can be replaced, but the primary way is to make use of the two inlet pipes that slot into the chamber, as seen in Figure 2.6. Fluid is introduced into one of the inlets and exits the other. The shape of the chamber prevents the formation of bubbles as the fluid fills the volume. Figure 2.7 illustrates how, in the built prototype, the chamber is filled without trapped air pockets, and is a successful outcome of the prototype.

In the case where a gas needs to be introduced into the chamber, the gas is introduced into the top of the chamber by removing a chamber plug. An adapter, not shown here, is used to form a seal with the gas tubing and the chamber to prevent leakage. After the chamber has been filled using the inlets, as in the case described above, a positive pressure is applied. By closing one of the inlets with a plug and leaving the other open, the gas pushes out an initial volume of the fluid in the chamber, but due to the size of the gas chamber and the positions of the inlet, a viable amount of fluid is left to cover the sample. The gas can be left to flow for as long as needed, even for the duration of the measurement, or the gas flow can be stopped after some time. Figure 2.8 shows that this process can be successfully performed without removing all the fluid.

The design is shown in Appendix C as series of Computer Aided Design (CAD) drawings. Additionally, an exploded view animation of the SCH can be watched on YouTube with the following link: https://youtu.be/AkC5tpZXx9U.

In conclusion, the SCH design meets the requirements in the following ways:

- As the SCH is mostly made out of a single body the structural rigidity ensures that the sample is held in place during measurements.
- The chamber design allows for effective and gentle introduction of a new chemical environment and/or the flow of a gas over the sample.
- Once assembled, the SCH is a closed system, which prevents contamination and allows for careful control of the environment the sample is exposed to.
- The weight of the SCH is roughly ten times more than the previous solution which improves the thermal stability considerably.
- Due to the modular design of the SCH it is relatively easy to add additional chambers, which allow for various future studies.
- The design was able to be prototyped successfully by the in-house machinist given the limitations of time, tools and materials at their disposal.
- While the complexity of the SCH is greater than that of the alternative, it remains easy to assemble and utilise.
- Lastly, the SCH achieves the above while remaining compact enough to fit into the stage and microscope housing.

2.4 Materials

In Chapter 5 three samples are analysed as case studies. What follows is a description of how each sample was prepared.

2.4.1 Quantum Dots

Qdot 605 (carboxylic acid conjugate) (Thermo Fisher Scientific) was diluted to ~80 pM in 10 mM MES buffer (pH 7) containing 0.2 mM $MgCl_2$ and 0.05% (w/v) Tween-20. A small droplet was applied to a poly-L-lysine (PLL)-treated coverslip, and another coverslip was placed on top. The excitation power and wavelength were 140 nW and 488 nm, respectively, which corresponds to approximately 26 W/cm^2. The dichroic mirror and fluorescence filter were 605dcxt (Chroma Technology Corp.) and 600LPF (Edmund Optics), respectively.

2.4.2 Alexa

Alexa Fluor 647 carboxylic acid (Thermo Fisher Scientific) was diluted to ~25 pM in 6 mM 2-(N-morpholino)ethanesulfonic acid (MES) buffer (pH 7) containing 4% (w/w) PLL, and spin-coated onto a glass coverslip. The excitation power and wavelength were 200 nW and 633 nm, respectively, which corresponds to approximately 36 W/cm^2. The dichroic mirror and fluorescence filter were FF650-Di01-25x36 (Semrock) and FELH0650 (Thorlabs), respectively.

2.4.3 Light Harvesting Complex II

LHCII from spinach, isolated using the protocol in Ref. [50], was diluted to ~3 pM in 20 mM buffer of 4- (2-hydroxyethyl) -1-piperazineethanesulfonic acid (HEPES) (pH 8) containing 0.03% (w/v) n-dodecyl-α-D-maltoside (α-DM) and 1 mM $MgCl_2$. A small droplet was applied to a PLL-treated coverlip, and the coverlip was placed in a custom sample chamber filled with an additional buffer-detergent-salt solution. To remove oxygen and prolong measurement times, a glucose/glucose oxidase/catalase mixture was used along with flushing of the sample chamber with N_2 gas. The excitation power and wavelength were 119 nW and 633 nm, respectively, which corresponds to approximately 22 W/cm^2. The dichroic mirror and fluorescence filter were TX660 and ET665lp, respectively (both from Chroma Technology Corp.).

Chapter 3

Analysis: Theory

This chapter discusses the theoretical basis for the analysis and application of single-molecule spectroscopy (SMS) data. Although the methods used to analyse the data are not novel, the framework that facilitates, organises, and streamlines the analysis is of considerable importance. Chapters 3-5 are significant expansions on the work presented in Joshua et al., (2024) [51].

3.1 Introduction

The first part of this chapter addresses the analysis of SMS data. It delves into the complexities of data capture and the types of data involved in SMS. In addition, the methods used to analyse SMS in this setup are discussed.

A key aspect of understanding the theoretical analysis of SMS involves recognising the phenomena that hinder and facilitate the analysis of SMS data.

Fluorescence intensity can be measured by counting the number of photons detected in a predefined bin size over a series of bins for the duration of the measurement, which is stored for analysis. However, measuring intensity in this manner involves averaging over a period of time, which limits the ability to detect transient events shorter than the bin size. Alternatively, the series of photon arrival times during the measurement can be stored by using a Time Correlated Single Photon Counting module, from which the intensity can be inferred, either by binning or by measuring the inter-photon duration, the inverse of which is intensity. While the binning method is simpler than measuring photon arrival times, the strengths of measuring photon arrival times far outweigh this simplicity, as they are highly beneficial for detecting and understanding special phenomena. Analysing single photon arrival times is a more complex task than intensity measurements.

In either case, a critical step in analysing SMS data is resolving the discrete intensity states present within fluorescence intensity measurements. A common approach involves hidden Markov models [52] (HMMs), where multiple underlying models are proposed and evaluated using the Bayesian information criterion. However, a major limitation of HMMs is their reliance on a priori knowledge about the system, which can significantly impact their effectiveness in cases where the number or nature of states is uncertain.

To mitigate this limitation, alternative methods have been explored that still require some insight into the underlying system but to a lesser extent. For instance, the aggregated Markov model [53] reduces the complexity of state-space representations by grouping microstates into metastable macrostates, preserving essential dynamical behaviour while reducing the need for detailed prior knowledge.

Another class of approaches involves non-Markovian memory kernel models [54], which attempt to bypass explicit state-space representations altogether. Instead of assuming memoryless transitions between predefined states, these models incorporate history-dependent effects, dynamically accounting for past states. This makes them particularly useful for analysing systems with hidden or slow degrees of freedom, where conventional Markov models may struggle to capture long-lived correlations.

The method developed by Watkins and Yang (2005) [55] provides a strong theoretical foundation for analysing intensity-driven dynamics in experiments like SMS which does not require a priori knowledge of the underlying system. This was one of the primary considerations for its choice as the method used in this study and forms the foundational basis of the analysis, providing a rigorous framework for detecting intensity changes with high temporal resolution.

An important phenomenon that can occur during SMS measurements is fluorescence intermittency. Fluorescence intermittency occurs within single-photon-emitting systems where the emission of photons switches radically between an ‘on’ state, where photons are being emitted, and an ‘off’ state, where photon emission ceases. This phenomenon can be problematic in SMS, as it adds complexity to the analysis.

Resolving this behaviour is crucial to gaining insight into the fluorescence system. An example of the utility of understanding fluorescence intermittency can be found in photosynthetic compounds, where non-photochemical quenching has been linked to photoprotective mechanisms [56]. In such cases, understanding fluorescence intermittency can shed light on photoprotection processes.

3.2 Change Point Analysis

The method developed by Watkins and Yang (2005) [55] uses a log likelihood test to identify discrete changes in intensity based on photon arrival times. This approach has significant advantages over traditional binning methods and allows for model-free analysis of molecular transitions.

Change points refer to specific moments in the data where a significant shift occurs in the observed signal, indicating a transition from one state to another. In SMS, detecting these points is crucial, as they often signify important molecular events or changes in the system’s dynamics.

Identifying these change points provides a detailed understanding of the underlying processes, helping to distinguish between different molecular states and their transitions. This approach improves the accuracy of the analysis, revealing intricate details that traditional methods may not.

3.2.1 Photon-by-Photon Log Likelihood Test

As detailed in Watkins and Yang (2005) [55], the photon-by-photon log likelihood test provides a robust mechanism to identify intensity change points by evaluating individual photon arrival events. This method offers several benefits in terms of temporal resolution and resilience against noise, particularly in time-resolved SMS experiments. This is because the method makes use of the Poisson distribution, which is an inherent aspect in photon counting, and therefore can elucidate dynamics within noisy measurements.

Consider a measurement in which the state of the sample does not change and is being excited by a constant light source. It follows that the resulting fluorescence will take place at a constant mean rate—see Equation (1.1) in Section 1.3—and therefore the measured intensity will be constant. Considering the fact that each fluorescence event is independent of previous events, this means that the measurement of photons emitted by a single system undergoing fluorescence will follow a Poisson distribution.

With the expected number of events denoted by λ we can calculate the probability of n events using the probability density function of a Poisson distribution given by

$$\mathrm{Pois}(\lambda) = \frac{\lambda^n e^{-\lambda}}{n!}.$$

In a fluorescence measurement with a period T and an average intensity[1] of I the expected number of detected photons is given by $\lambda = IT$. We can therefore use the following equation to calculate the probability of detecting n photons given an intensity I in a period T with

$$g(n;I,T) = \frac{(IT)^n e^{-IT}}{n!}. \tag{3.1}$$

---

[1] During spectroscopic measurements the absorption cross section usually remains constant. Furthermore, in the case of fluorescent studies, where the emission spectra also remains constant, the definition of intensity is simplified from power per unit area—or energy per unit time per unit area—to photon per unit time.

Consider a series of N detected photons, where the state of the fluorescent system changes from state α to state β before photon n_k. In the generalised likelihood ratio test, as proposed by Ref. [55], we construct two hypotheses: the null hypothesis H_0 that no state change took place before photon n_k, and the positive hypothesis H_P that a state change took place before photon n_k. To relate these hypotheses with the intensity of the states, we can say that the intensity remains constant during the null hypothesis, whereas for the positive hypothesis, the intensity before and after n_k is not the same. Or, stated mathematically:

$$H_P: \quad I(t_1) = \cdots = I(t_k) = \quad I_\alpha \neq I_\beta \quad = I(t_{k+1}) = \cdots = I(t_N) \text{ and} \tag{3.2}$$

$$H_0: \quad I(t_1) = \cdots = I(t_k) = \quad I_0 \quad = I(t_{k+1}) = \cdots = I(t_N). \tag{3.3}$$

where I_α is the average intensity before the state change, I_β is the average intensity after the state change, and I_0 is the average intensity if no state change occurs.

To calculate the likelihood of H_P in Equation (3.2) we need to calculate the probability of detecting n_α photons before the state change and n_β photons after the state change. Using Equation (3.1) we find

$$L_P = g(n_\alpha; I_\alpha, T_k)g(n_\beta; I_\beta, T_{N-k}) = \frac{(I_\alpha T_k)^k e^{-I_\alpha T_k}}{k!} \frac{\left(I_\beta T_{N-k}\right)^{N-k} e^{-I_\beta T_{N-k}}}{(N-k)!}, \tag{3.4}$$

where $n_\alpha = k$ and $n_\beta = N-k$. The same can be done for the likelihood of H_0 in Equation (3.3) using Equation (3.1) with

$$L_0 = g(n_\alpha; I_0, T_k)g(n_\beta; I_0, T_{N-k}) = \frac{(I_0 T_k)^k e^{-I_\alpha T_k}}{k!} \frac{(I_0 T_{N-k})^{N-k} e^{-I_\beta T_{N-k}}}{(N-k)!}. \tag{3.5}$$

Using Equations (3.5) and (3.7) we can then calculate the log likelihood that a state change occurred after photon n_k by calculating the ratio of L_P over L_0 with

$$\mathscr{L}_k^0 = \ln \frac{L_P}{L_0} = \ln \frac{(I_\alpha T_k)^k e^{-I_\alpha T_k} \left(I_\beta T_{N-k}\right)^{N-k} e^{-I_\beta T_{N-k}}}{(I_0 T_k)^k e^{-I_\alpha T_k} (I_0 T_{N-k})^{N-k} e^{-I_\beta T_{N-k}}}.$$

Finally, given that $I_\alpha = k/T_\alpha$, $I_\beta = (N-k)/T_\alpha$, and $I_0 = N/T$ the above can be simplified to

$$\mathscr{L}_k^0 = 2k \ln \frac{k}{V_k} + 2\,(N-k) \ln \frac{N-k}{1-V_k} - 2N \ln N, \tag{3.6}$$

where $V_k = T_k/T$.

To prevent errors when k approaches the boundaries of the segment, two improvements are made to Equation (3.6). A weighting W_k is added, called the Henderson weighting [57], given by

$$W_k = \frac{1}{2} \ln \frac{4k\,(N-k)}{N^2}.$$

Secondly, a standardised log likelihood ratio is calculated using

$$\bar{\mathscr{L}}_k = \frac{\mathscr{L}_k^0 - E(\mathscr{L}_k^0)}{\sigma_k},$$

where $E(\mathscr{L}_k^0)$ is the expectation value of Equation (3.6) and σ_k is the standard deviation of Equation (3.6). Given the following

$$\begin{aligned}
V_k &= T_k/T, \\
\mu_k &= E(\ln V_k) = -\sum_{j=k}^{N-1} \frac{1}{j}, \\
\mu_{N-k} &= E(\ln\,(1-V_k)) = -\sum_{j=N-k}^{N-1} \frac{1}{j}, \\
\xi &= \frac{\pi}{6} - \sum_{j=1}^{N-1} \frac{1}{j^2}, \\
v_k^2 &= \sum_{j=k}^{N-1} \frac{1}{j^2}, \text{ and} \\
v_{N-k}^2 &= \sum_{j=N-k}^{N-1} \frac{1}{j^2},
\end{aligned} \tag{3.7}$$

the final equation for the calculation of the weighted standardised log likelihood ratio is given by

$$\begin{aligned}\mathscr{L}_k &= \frac{\mathscr{L}_k^0 - E(\mathscr{L}_k^0)}{\sigma_k} + W_k \\ &= \frac{-2k\ln V_k + 2k\mu_k - 2(N-k)\ln(1-V_k) + 2(N-k)\mu_{N-k}}{4k^2 v_k^2 + 4(N-k)^2 v_{N-k}^2 - 8k(N-k)\xi} + \frac{1}{2}\ln\frac{4k(N-k)}{N^2}.\end{aligned} \tag{3.8}$$

The null hypothesis is rejected when

$$\max_{1\leq k\leq N}\{\mathscr{L}_k\} \geq \tau_{1-\alpha}, \tag{3.9}$$

where $\tau_{1-\alpha}$ is calculated with Noé's algorithm [58] and α is the probability of identifying a change point erroneously. In practical terms, this means that the higher the value of $\mathscr{L}_k$, the more confidence there is in a change point at that position in the data. This statistical interpretation ensures that the detected change points are not merely artefacts of random fluctuations or noise, but are indeed indicative of true changes in the molecular state.

A key advantage of the photon-by-photon approach is its ability to achieve far greater time resolution by analysing individual photon arrival times. This level of detail enables the detection of rapid state transitions that might otherwise be missed when using traditional binning techniques, where averaging effects can result in the loss of critical information. Additionally, the method is model-free, which means it does not rely on assumptions about the underlying kinetic models. This makes it broadly applicable to various molecular systems without imposing constraints on the data. Furthermore, it is robust against the Poisson noise that inherently accompanies photon detection, allowing for accurate identification of change points even in conditions with low signal-to-noise ratios. These strengths make the photon-by-photon approach a powerful tool for understanding the dynamics of molecular transitions.

3.2.2 Recursive Segmentation

Recursive segmentation is a crucial step in the CPA process that enables the detection of multiple change points within a single emission trajectory. Additionally, the effectiveness of the CPA method is highest when the number of data points considered is 1000 or less. Once a change point is identified, Equation (3.9) is applied recursively to the segments of the data that lie before and after the detected change point. This approach systematically divides the data into smaller intervals, continuously searching for additional transitions that may occur at different points in time.

The recursive nature of this method is essential because SMS data often contain a series of rapid transitions between different states. By iteratively segmenting the data, the analysis does not limit itself to a single change point but rather maps out a complete sequence of state changes, revealing the dynamic behaviour of the molecular system in fine detail. This iterative process stops when no further statistically significant change points are detected within the segmented intervals, ensuring that all relevant state transitions are accurately captured.

3.3 Intensity Level Grouping

Intensity level grouping is a crucial step in the analysis of SMS data. Watkins and Yang [55] outlined a sophisticated approach for grouping intensity levels based on photon arrival times, which allows for accurate identification of molecular states and the transitions between them. This approach has been particularly valuable in identifying discrete intensity levels in a model-free manner, which avoids many of the pitfalls of traditional clustering methods that rely on binning or arbitrary thresholds. One major advantage of this grouping method is its ability to handle states where very few data points were measured. By grouping such sparsely represented states with others that share similar characteristics, the analysis can effectively reduce noise and improve the statistical robustness of the identified states, ensuring that rare but meaningful transitions are not overlooked.

3.3.1 Agglomerative Hierarchical Clustering

Agglomerative Hierarchical Clustering (AHC) is a widely used unsupervised machine learning algorithm designed to group data points into clusters based on their similarity. It is a bottom-up approach, which means that it starts by treating each data point as an individual cluster and then progressively merges the most similar clusters until all points belong to a single cluster. Unlike partition-based methods such as K-means, which require the number of clusters to be predefined, AHC generates a hierarchical structure that allows for flexible cluster selection at different levels of granularity. This method is particularly useful in high-dimensional datasets, where complex relationships between data points exist. Because AHC captures nested relationships, it is especially beneficial when data exhibit natural hierarchical organisation.

For a series of measured photons with times $\{t_0, t_1, t_2, t_3, \cdots\}$ there is a corresponding series of inter-photon durations $\{dt_1, dt_2, dt_3, \cdots\}$ so that

$$\{t_0 + dt_1 = t_1; t_1 + dt_2 = t_2; t_2 + dt_3 = t_4; \cdots\}.$$

It follows from Equation (1.1) that the probability density function of measuring an inter-photon duration of dt in a fluorescence state with an intensity I is given by

$$f(dt; I) = I e^{-I\,dt}. \tag{3.10}$$

Consider a fluorescent measurement where there are J number of distinguishable intensity levels as resolved by the method discussed above so that J_j represents the set of measured photons that belong to level j. The classification log likelihood for

G number of groups can be calculated using the following expression:

$$\mathscr{L}_C(\{dt_i\};\{I_j\}) = \sum_{j=1}^{G}\sum_{i\in J_j} \ln f(dt_i;I_j) = \sum_{j=1}^{G}\sum_{i\in J_j} \ln\left(I_j e^{-I_j dt_i}\right), \tag{3.11}$$

where there are J number of distinguishable intensity levels as resolved by the CPA analysis. The AHC algorithm iteratively evaluates the above while merging two groups, and does so for each possible pair of grouping. Suppose that two groups, group j and group i, are to be merged, a log likelihood merge-merit function can be calculated by using Equation (3.11)

$$M_{\langle j,m\rangle} = \mathscr{L}_C(\{dt_i\};\{I_1,\cdots,I_G\}) - \mathscr{L}_C(\{dt_i\};\{I_1,\cdots,I_{\langle j,m\rangle},\cdots,I_{G-1}\}).$$

This finally simplifies to

$$M_{\langle j,m\rangle} = \left(n_m + n_j\right)\ln\left(\frac{n_m+n_j}{T_m+T_j}\right) - n_m \ln\left(\frac{n_m}{T_m}\right) - n_j \ln\left(\frac{n_j}{T_j}\right). \tag{3.12}$$

The group pair that results in the greatest log-likelihood merge-merit value is then the most likely to belong to the same group and is merged. One way that this can be denoted is to construct a matrix of p_{mj} elements such that

$$p_{mj} = \begin{cases} 1 & \text{if } I_j \text{ belongs to the group } m \\ 0 & \text{if not} \end{cases} \tag{3.13}$$

This can be performed iteratively until all groups are merged into a single group. The most likely state is then decided on, as explained in Section 3.3.3.

3.3.2 Expectation Maximisation Clustering

AHC suffers a high sensitivity to the initial conditions, as well as an inability to regress. For this reason, the authors suggest the implementation of an expectation maximisation clustering (EMC) that uses the results of AHC as an initial state. To this end, beginning with p_{mj} from Equation (3.13) and using Equation (3.1), the probability density of measuring I_j is given by

$$\prod_{m=1}^{G} g(n_j; I_m, T_j)^{p_{mj}}.$$

From this follows that the total likelihood function is

$$L(I_j; p_{mj}, p_m, I_m, T_j) = \prod_{j=1}^{J+1} \prod_{m=1}^{G} \left[p_m g(n_j; I_m, T_j)\right]^{p_{mj}}.$$

Calculating the log-likelihood of the above—which is maximised by EMC—gives

$$\mathscr{L}_{\text{em}}(I_j; p_{mj}, p_m, I_m, T_j) = \sum_{j=1}^{J+1} \sum_{m=1}^{G} p_{mj} \ln \left[p_m g(n_j; I_m, T_j)\right]. \tag{3.14}$$

The EMC cycles over two steps until p_{mj} converges: a maximisation step (M-step), and an expectation step (E-step). The M-step consists of calculating the following maximum likelihood estimate parameters[2]:

$$\begin{aligned}
\hat{T}_m &= \sum_{j=1}^{J+1} \bar{p}_{mj} T_j, \\
\hat{n}_m &= \sum_{j=1}^{J+1} \bar{p}_{mj} n_j, \\
\hat{p}_m &= \frac{\hat{T}_m}{T}, \text{ and} \\
\hat{I}_m &= \sum_{j=1}^{J+1} \frac{\bar{p}_{mj} n_j}{\hat{T}_m},
\end{aligned} \tag{3.15}$$

which is followed by the E-step, in which the next iteration of p_{mj} is calculated with

$$\bar{p}_{mj} = \frac{\hat{p}_m g(n_j; \hat{I}_m, T_j)}{\sum\limits_{m=1}^{G} \hat{p}_m g(n_j; \hat{I}_m, T_j)} = \frac{\hat{p}_m \frac{(\hat{I}_m T_j)^{n_j}}{n_j!} e^{-\hat{I}_m T_j}}{\sum\limits_{m=1}^{G} \hat{p}_m \frac{(\hat{I}_m T_j)^{n_j}}{n_j!} e^{-\hat{I}_m T_j}}. \tag{3.16}$$

---

[2] On the first iteration M-step the value of $\bar{p}_{mj}$ is given by the resulting p_{mj} of the preceding AHC step.

Once the EMC has converged, the grouping state is given by p_{mj}. Iterating over multiple AHC-EMC steps ultimately provides the most accurate grouping sequence.

3.3.3 Number of States

Determining the appropriate number of intensity states is a critical part of the analysis. This is achieved by using the Bayesian Information Criterion (BIC) to assess model complexity against data fit quality. The BIC helps to prevent overfitting by penalising models with too many parameters, ensuring that only meaningful states are identified. The approximate BIC value is calculated with

$$\mathrm{BIC}(\mathscr{L}_{\mathrm{em}}, N_G, N_{\mathrm{cp}}) \approx 2\mathscr{L}_{\mathrm{em}} - (2N_G - 1)\ln N_{\mathrm{cp}} - N_{\mathrm{cp}} \ln N, \tag{3.17}$$

where $\mathscr{L}_{\mathrm{em}}$ is the resulting AHC-EMC log-likelihood in Equation (3.14), N_G is the number of groups, and N_{cp} is the number of detected change points. This value can be interpreted as an indication of the amount of information the solution manages to represent, and therefore the solution with the highest BIC value is taken to be the best representative of the grouping achieved.

3.3.4 Global Grouping

Global grouping refers to the process of analysing multiple trajectories or datasets collectively to identify common intensity levels across different molecules. In SMS, data is often obtained from numerous molecules, each with its unique trajectory. These differences can arise from various factors, such as variations in the focus point, differences in the intensity of the laser power, or the orientation of the measured particle. Such factors can lead to variability in the measured intensity, making consistent identification of intensity levels challenging.

Global grouping involves analysing all these trajectories together to find common intensity levels, thus ensuring that molecular states are consistently defined throughout the entire dataset. This approach is especially critical when comparing multiple experimental conditions or combining data from various molecules to enhance statistical power.

Global grouping can potentially resolve the actual number of states that the system comprises. By considering data from multiple molecules collectively, global grouping can identify subtle but consistent states that might not be evident from individual trajectories, providing a more comprehensive understanding of the molecular system's complexity. However, for global grouping to be applied successfully, the systematic variation in the maximum intensity of the unquenched states between measurements needs to be acceptably small. In the cases where this variability cannot be prevented global grouping could still be applied if a method can be found to normalise the intensities. In the case of light-harvesting complex II (LHCII) this can be successfully achieved due to the linear relationship between the intensity of the state and its fluorescence lifetime.

Chapter 4

Analysis: Software

4.1 Overview of *Full SMS*

SMS is a powerful investigation technique, as the measured data contain a wealth of information. Numerous approaches have been developed during the past three decades to provide access to diverse types of dynamic information about various kinds of nanoscaled systems. However, since the developments have been made across different research groups, each measurement type often requires a different data analysis programme, which necessitates the analysis of each measurement type to take place typically in isolation from the rest. In addition, different software packages usually require different file formats and often lack versatility, having been designed for a particular application. As a result, it usually takes a significant amount of effort to obtain a multiparameter SMS data set through the combination of different existing software packages.

Although numerous open-source programmes provide access to their scripts, it is cumbersome to adapt those scripts to enable other applications, and it is only a viable solution for experienced programmers. Commercial software is also limited in its application, incompatible with other types of hardware, and expensive.

Therefore, there is a need for generalised open-access analysis suites to facilitate and standardise the analysis of multiparameter SMS data. Some work in developing such software has been done, and examples include the Multiparameter Fluorescence Detection (MFD) software of the group of Seidel [59], *PAM* from the group of Lamb [60], *Globals* from the group of Gratton [61], *iSMS* from the group of Birkedal [62], and a change-point analysis toolbox from the group of Koenderink [63]. A general-purpose file format for SMS and other photon-by-photon data has also been developed, namely Photon-HDF5 [64].

Here, we present *Full SMS*, a single workspace where data from multiple concurrent measurements of the same subject can be viewed and analysed. Unlike the MFD, *PAM*, *Globals*, and *iSMS* software mentioned above, our software is largely based on the analysis of intensity time traces using a statistically robust change point analysis (CPA). CPA, along with subsequent clustering of the intensity levels and fluorescence lifetime fitting, is a powerful approach to analysing fluorescence intermittency and identifying the different emissive states of nanoscale emitters. *Full SMS* has also been open-sourced to the community. It was furthermore entirely written in Python which is known for its ease-of-use and open-source community, separating it from MFD and *Globals*, which are closed source, while *PAM* and *iSMS* are open source but written in MATLAB, which is proprietary software. Unlike the toolbox developed by Palstra and Koenderink, our software uses a graphical user interface (GUI). This has several advantages over simple scripts or command-line programmes, the main ones being that easy visual exploration of the data is enabled and that users need very little technical knowledge to get started using the software. In this regard, our software is similar to *Glotaran* [65, 66], a widely used tool to analyse time-resolved spectra. Additional advantages of *Full SMS* over the Palstra and Koenderink toolbox are the display of raster scan images and spectral time traces, automated trimming of traces, and built-in data filtering based on intensity or lifetime distributions. We also developed a custom file format based on the HDF5[1] format and similar to Photon-HDF5, but which includes specific additional

---

[1] The choice of HDF5 as the file format used for SMS measurements is motivated in Section 4.4.5.

measurement capabilities suitable to measurements of single immobilised particles. *Full SMS* is available at `http://github.com/BioPhysicsUP/Full_SMS`.

4.2 Analysis

We first give a summary of the analyses and tools that the software provides and follow this with a discussion of the main features, using mainly the data from one measured particle as a stepwise illustration of the analysis capabilities. For detailed information, including practical usage instructions, see the user guide at `https://up-biophysics-sms.readthedocs.io/en/latest/`. As the primary use case of the software involves point measurements on individual particles, each individual measurement set is called a ‘particle’ in our software, as well as in the discussion below (this is also reflected in the name *Full SMS*).

Many kinds of analysis can be performed on SMS data, depending on the type(s) of measurements made. A predominant SMS measurement modality is fluorescence intensity (also called ‘brightness’), either directly by measuring the photon flux within predetermined time windows or indirectly by measuring photon arrival times and then calculating the intensity in time bins (see Section 4.2.1). The latter, called time-correlated single-photon counting (TCSPC), typically provides a statistical distribution of photon arrival times relative to the moment of excitation of the sample by a pulsed light source, a method commonly known as TTTR data collection. Further analysis allows the resolved intensity levels to be grouped by employing a clustering algorithm, and the statistically most probable number of states to be determined (see Section 4.2.2). *Full SMS* additionally enables global grouping of resolved intensity levels from multiple measurements to resolve the underlying states with greater confidence (see Section 4.2.2). *Full SMS* can analyse TCSPC data to extract fluorescence lifetimes for the whole measurement, but also for each resolved intensity level or intensity group (see Section 4.2.3). It can also calculate the second-order photon correlation function $g^{(2)}(\tau)$(see Section 4.2.4).

Additional features, available in separate tabs, are data filtering options (see Section 4.2.5), the display of measured fluorescence spectral sequences and raster scan images, and options for exporting the analysed data (see Section 4.2.6, Section 4.2.7, and Section 4.5). Raster scan images are useful for spatial mapping of measured particles in concurrent imaging modalities, or simply to locate individual particles before measurement.

The software also provides several other features (see Section 4.3), such as data format conversion, photon burst detection, and the ability to define regions of interest (i.e., trimming of intensity traces) for each measurement individually, either manually or procedurally, by defining intensity and temporal thresholds.

4.2.1 Intensity Level Resolution

Analysis of an emitter's intensity trace is useful not only for assessing its intrinsic brightness but also for investigating time-dependent processes that give rise to fluorescence intensity fluctuations, which may result from static or dynamic quenching, photobleaching, or other photophysical changes. Among these, fluorescence intermittency, a tell-tale sign of a single emitter, is evidenced by abrupt and reversible decreases in the photon emission rate for periods anywhere between submilliseconds and tens of minutes [67, 68]. To correctly analyse intensity fluctuations, the points at which the intensity changes occur must be identified. An additional benefit of such a capability is the ability to filter out unwanted segments of the intensity trace, for example, those corresponding to photobleaching.

Identifying the points at which a change in the photon statistics occurs is typically done using change-point analysis (CPA)—an approach to locate abrupt changes in a time-dependent variable such as emission intensity [55, 69]. This is commonly used to analyse fluorescence intermittency, as in the present work, but it can also be used, for example, to detect changes in the diffusion behaviour of a single particle [70]. CPA can be done using time-binned intensity traces or directly using the photon

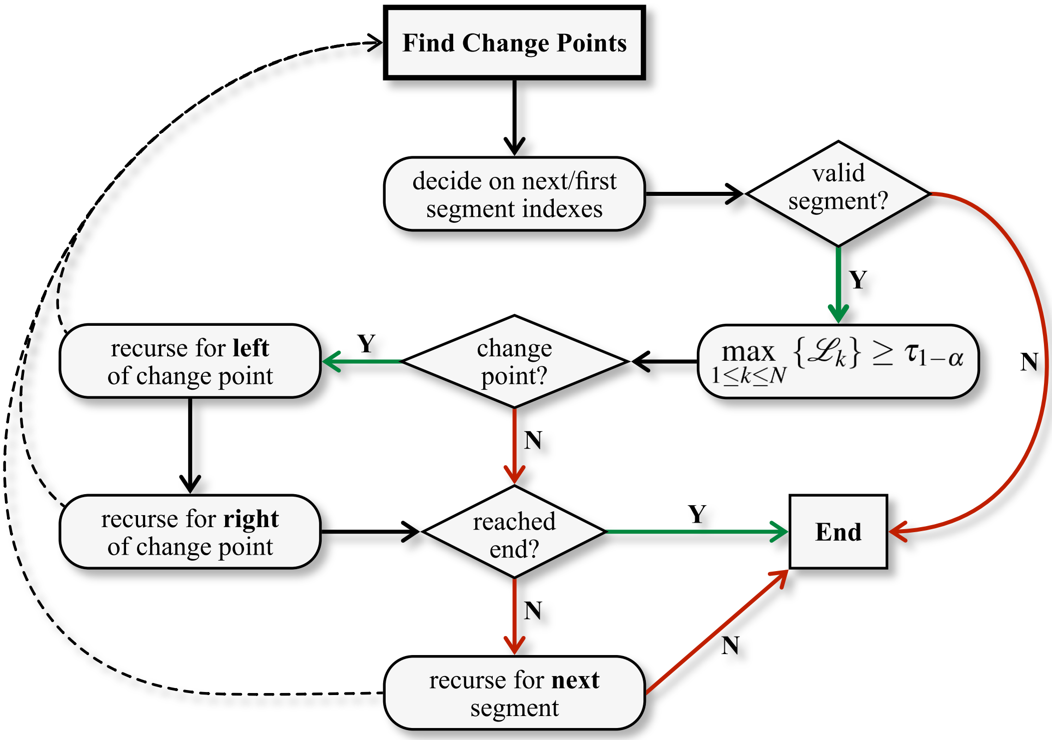


FIGURE 4.1: Flowchart depicting the recursive algorithm for change-point detection. The process begins by identifying potential change points in the data sequence and proceeds through several decision nodes. After defining the initial segment indices, the algorithm checks if the segment is valid. If valid, it searches for a change point within the segment; if no change point is detected, the algorithm recurses to the next segment. If a change point is identified, the algorithm splits the data at that point and recursively analyses the left and right sub-segments. This process continues until the end of the data sequence is reached, concluding the analysis. The important code for this can be seen in Sections A.1, A.2 and A.4.

arrival times. Working in the intensity regime (the first approach) is often done to simplify the analysis work flow, and suitable CPA algorithms have been developed [56, 71, 72]. The choice of bin size is crucial in this case, as it determines the temporal resolution with which dynamics can be resolved. However, data binning inevitably introduces bias in the analysis [55, 69, 73, 74]. This bias is especially relevant when the dynamics of the intensity fluctuations occur at a rate comparable to the chosen bin size.

Change-point detection in the temporal regime avoids the bias introduced by data

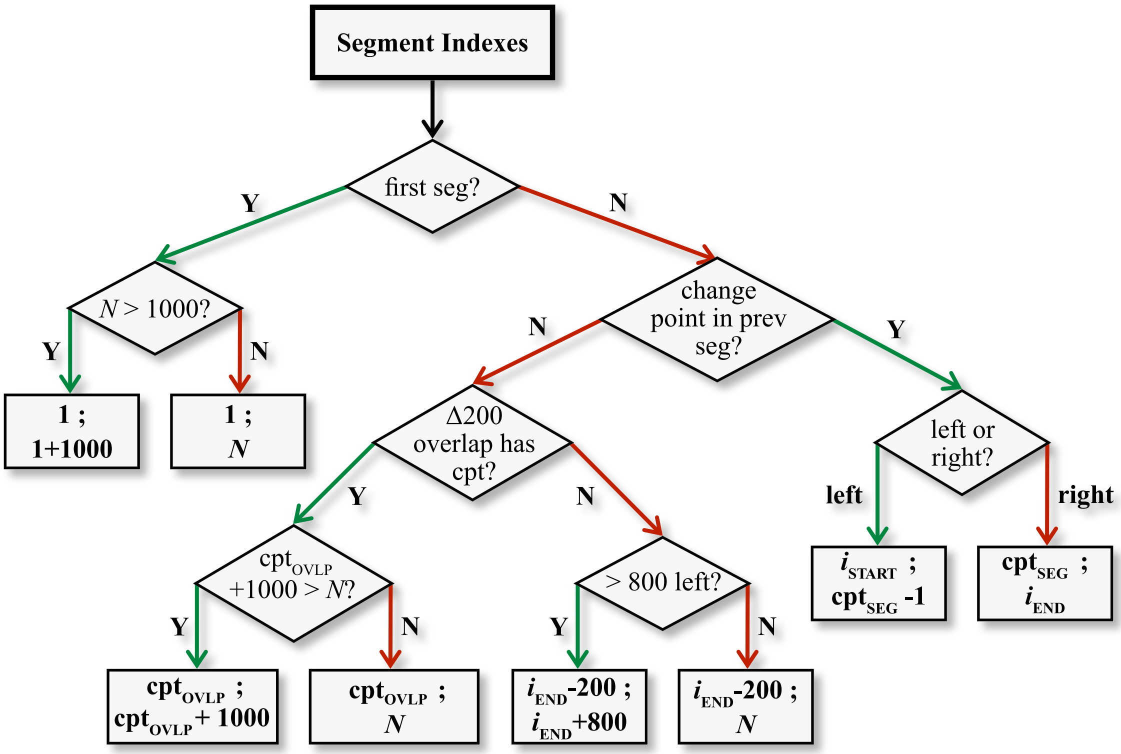
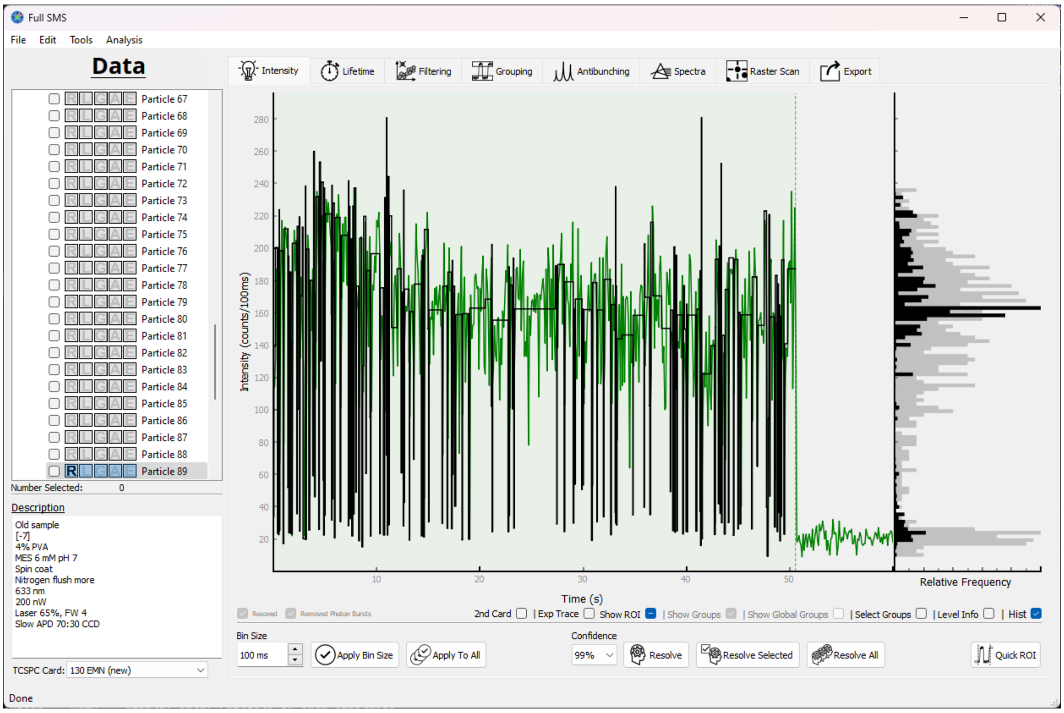


FIGURE 4.2: Segmentation is critical for the effective use of the change-point detection algorithm (see Section 3.2.2). This figure illustrates the decision tree for determining segment indexes. Starting with the segment index decision, the flow splits based on whether it is the first segment and various conditions on segment length N, overlap, and presence of previous change points. For the first segment, the decision branches based on whether $N > 1000$, with adjustments made to the segment range accordingly. For subsequent segments, the flow depends on whether a change point was detected in the preceding segment and on specific overlap conditions (e.g., Δ200 overlap). Conditions like the presence of overlap with change points or availability of points on the left (> 800) guide index adjustments for overlapping regions. The code for this can be seen in Section A.3.

binning by working with the individual photon arrival times. This enables the optimal amount of information to be retrieved from a photon stream, limited only by the shot noise. We opt for the CPA method introduced in Ref. [55] because of the independence of a physical model underlying the photon statistics. This method compares the duration between each pair of consecutive photon detections to the durations between neighbouring pairs of photon detections and assigns a change



Number Selected: 0

Description

Old sample
[-7]
4% PVA
MES 6 mM pH 7
Spin coat
Nitrogen flush more
633 nm
200 nW
Laser 65%, FW 4
Slow APD 70:30 CCD

TCSPC Card: 130 EMN (new)

Bin Size: 100 ms

Confidence: 99%

Done

FIGURE 4.3: Screenshot of a resolved fluorescence brightness trace of a single Alexa Fluor 647 dye molecule with trimmed region of interest (ROI) shown in green. See text as well as the user guide for details.

point to the instant when a sustained abrupt change in these durations occurs. A chosen confidence value determines a statistical criterion for each potential change point that must be met before acceptance. The higher the confidence value, the more stringent the criteria that need to be met and the less sensitive the detection of the change point. In addition, each change point has an associated temporal error region, defined by the first and last detected photons that pass the statistical test. The resulting change points separate the measurement into segments.

After loading a measured data set, an intensity trace is displayed in the first tab of *Full SMS* with a corresponding intensity histogram on the right (see Figure B.1). In this tab, CPA can be performed on the data. The screenshot in Figure 4.3 is an example of a measurement of individual Alexa Fluor 647 (hereafter called Alexa) dye molecules—a common dye used for microscopic chromophore labelling—that was subjected to CPA. The green trace represents the binned intensities calculated

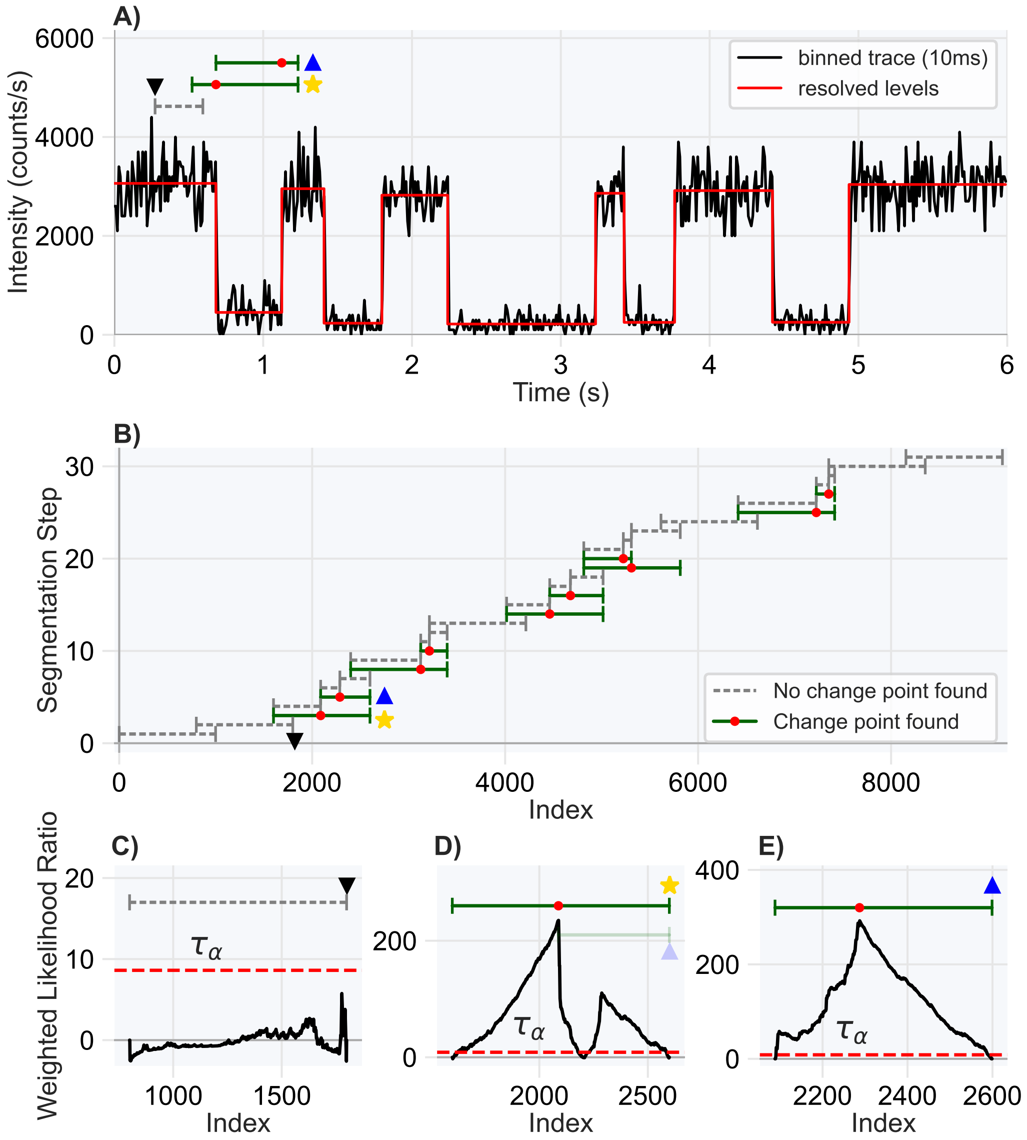


FIGURE 4.4: Illustration of the change point analysis process applied to our SMS data, which is the initial step in our measurement pipeline.

from the measured time series, whereas the black trace represents the resolved levels. Note that the CPA intensity level resolution is independent of the bin size used for the intensity representation since the CPA operates on the underlying time series. An additional functionality is the display of two-channel data (Figure B.2). Currently, the two channels are analysed separately, except for the second-order correlation analysis (see Section 4.2.4).

The general algorithm for CPA is illustrated in the flow diagram in Figure 4.1, and the important code sections can be seen in Appendix A. As explained in Section 3.2, the CPA algorithm is most effective in segments of 1000 data points or less. Figure 4.2 illustrates the algorithm used to achieve proper segmentation during the CPA.

Figure 4.4 acts as an example of the CPA process. The analysis aims to identify statistical changes in the intensity signal, which are subsequently grouped into defined levels, forming the foundation for further analysis. **Panel A)** (top) presents the final result of this change point analysis. The black line represents the raw, binned trace of intensity counts over time, while the red segments denote the resolved levels. These levels are statistically defined based on detected change points in the intensity signal, highlighting the transitions between different states. **Panel B)** (middle) demonstrates the segmentation process. Here, the analysis starts by searching for change points within an initial segment (bottom left of the panel) and progresses sequentially. If no change point is identified within a segment, the analysis shifts to the right by some predefined amount, currently 200 data points, and repeats the search with an overlapping window. Once a change point is detected, marked by a red dot on the solid line, the segment is divided into left and right sections, which are independently analysed for further change points. This segmentation and iterative search continues, producing a stepwise trace that corresponds to the detected change points. **Panels C)–E)** (bottom) provide examples of the weighted log-likelihood ratio test calculated with Equation (3.8) for each data point in three segments. The y-axis represents the weighted log-likelihood ratio value and the x-axis represents the data point indexes. The three examples of these tests are marked by symbols consistent across all the panels (black inverted triangle, yellow star, and blue triangle). These tests assess the statistical likelihood of a change point within each segment. In **Panel C)**, the test does not exceed the red threshold line, indicated by τ_α (indicating a 5% false positive rate), meaning no change point was detected. In **Panel D)**, the likelihood ratio exceeds the threshold at two peaks, indicating the presence of a change point. The maximum peak defines the change point, and subsequent searches are conducted on both sides of this segment. **Panel E)** shows a scenario where only one additional change point is found to the right of the initial detected change point. This figure thus provides a comprehensive overview of the change point detection process, from initial signal segmentation to statistical verification of detected points, which is essential for accurately defining state transitions in our data.

4.2.2 Intensity-Level Grouping

The information gained from CPA can be significantly increased by statistically associating the resolved intensity levels with each other. A grouping or clustering of this kind has multiple uses. In the first place, it serves as a tool to resolve the underlying states in the system being studied [55]. An added advantage is performing fluorescence decay fitting of levels with low numbers of detected photons. The efficacy of fluorescence decay fitting (discussed in Section 4.2.3) depends, among other factors, on the number of data points used. Investigating states that emit very few photons, therefore, often proves to be problematic. By grouping several of these dark or dim states, a successful fit is possible where it would otherwise involve a large uncertainty.

The clustering method implemented in *Full SMS* is a mixed model of an agglomerative hierarchical clustering (AHC) method and expectation maximisation clustering (EMC) method [55, 72, 75]. Figure 4.5 illustrates the algorithm used for the grouping analysis. AHC starts by considering each intensity level as a group of its own. It then merges the pairs of groups that maximise a log-likelihood ratio merit function calculated using the number of photons within and the total duration of the two candidate groups (see Equation (3.12)). In other words, pairs of groups with the smallest intensity difference will be combined. This merging of group pairs is repeated until all groups have merged into a single group comprising all levels.

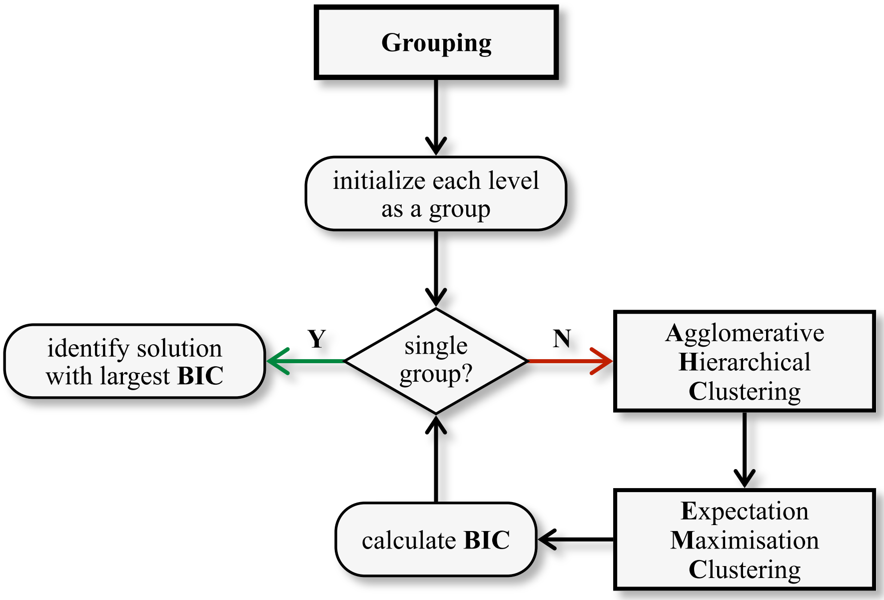


FIGURE 4.5: Flowchart illustrating the clustering process followed by the application of Bayesian Information Criterion (BIC) to identify the most representative grouping solution. The process begins with grouping, where each level is initialised as a separate group. A check is performed to determine if there is only a single group. If there is, the solution with the largest BIC is identified as the optimal grouping. If multiple groups remain, the algorithm proceeds to Agglomerative Hierarchical Clustering and then Expectation Maximisation Clustering to refine groupings. The BIC value is then calculated for each solution, guiding the decision toward the grouping configuration that maximises the BIC score. The code for this can be seen in Sections A.5 and A.8.

One of the limitations of AHC is that the outcome is highly dependent on the initial state and outliers, and once two groups have merged, the decision is not challenged. For this reason, the outcome of the AHC—the combined levels corresponding to each number of possible groups—is used as an initial state for more advanced EMC, which serves to optimise the intensity-level groupings. In the EMC algorithm, the log-likelihood that each group would have a certain resulting intensity is calculated given the number of photons in the group and the group's total duration, as defined by the Poisson probability density function. After the total probability is calculated, new probabilities for each level related to each group are calculated. Figure 4.6 is a detailed flow diagram of the AHC-EMC algorithm.

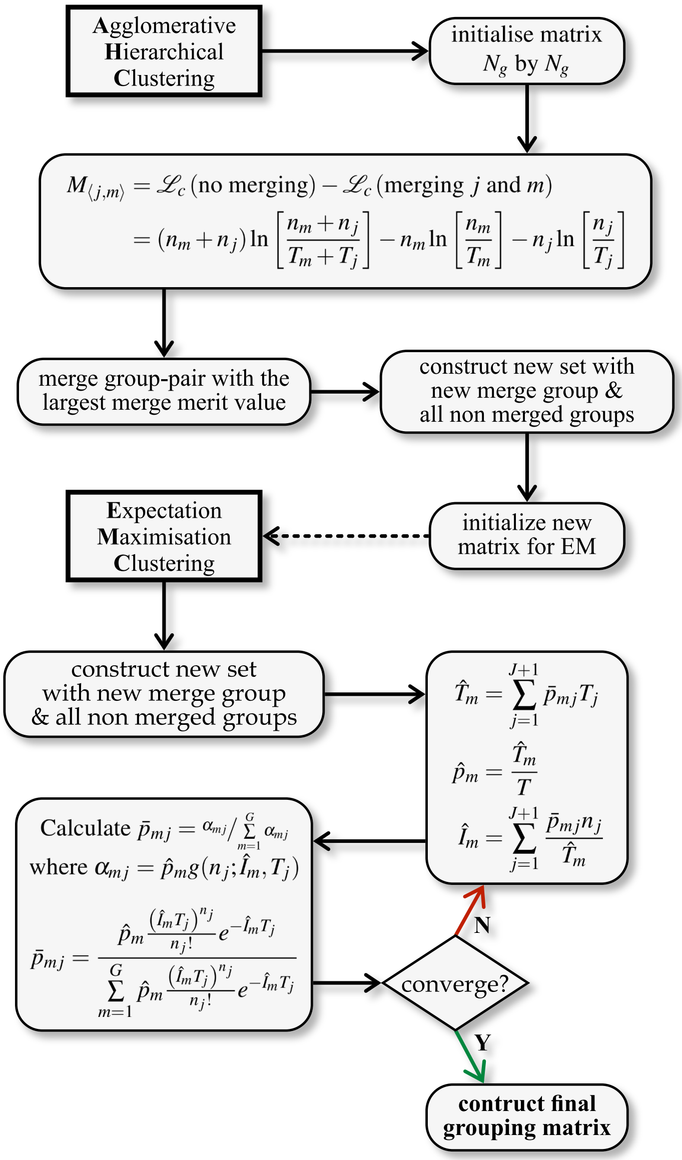
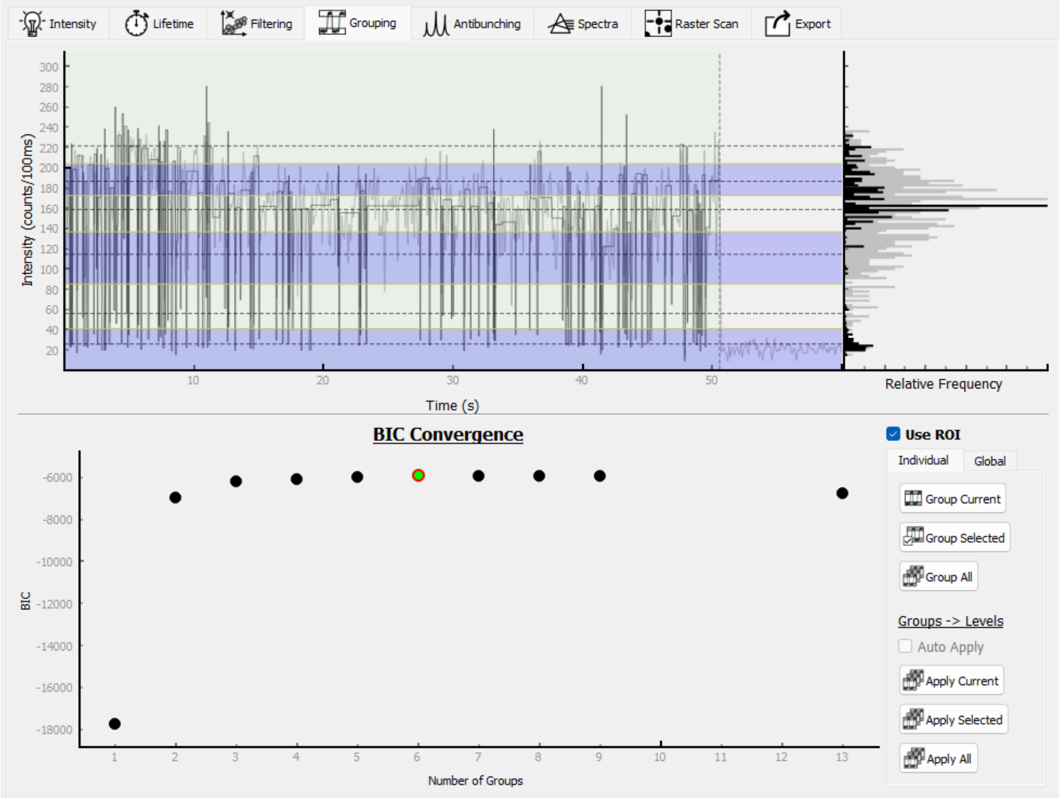


FIGURE 4.6: Flowchart of the Agglomerative Hierarchical Clustering and Expectation Maximisation Clustering integration process for constructing the final grouping matrix. The algorithm begins with the initialisation of an $N_g \times N_g$ matrix, where N_g is the number of groups. The merit value $M_{(j,m)}$ is calculated for each group pair (j,m) with Equation (3.12), based on the difference in log-likelihood $\mathscr{L}_c$ between scenarios with and without merging groups j and m. The group-pair with the largest merit value is merged, and a new set is constructed with the merged group alongside all non-merged groups. The process proceeds to Expectation Maximisation Clustering (EMC), where a new matrix is initialised. In this step, the parameters $\hat{T}_m$, $\hat{p}_m$, and $\hat{I}_m$ are iteratively estimated, based on the expressions provided (see Equation (3.15)), to optimise the groupings. Specifically, $\bar{p}_{mj}$ is calculated by iteratively adjusting the probability weights α_{mj} for each group m relative to all possible groups G, considering the likelihood and observed metrics n_j, T_j, and others (see Equation (3.14)). The convergence criterion is then assessed: if the parameters converge, the final grouping matrix is constructed; if not, the algorithm loops back to further iterations in EMC with updated parameters. The code for the AHC can be seen in Section A.6, and the code for the EMC can be seen in Section A.7.

Neither the AHC nor the EMC method predicts the most likely number of groups. The latter is done using the Bayesian Information Criterion (BIC) [55]. *Full SMS* provides an interactive tool to visualise the groups identified in each AHC–EMC grouping scheme along with the corresponding BIC value.

The screenshot in Figure 4.7 is an example of the intensity-level grouping of data from an Alexa measurement. The upper pane shows the group intensity as dotted lines and the group boundaries as yellow lines. The resulting intensity “bands" are indicated by alternating light-blue and white bands. The BIC values for each grouping scheme are shown in the bottom panel where the *best* grouping scheme is indicated as a green circle, representing the statistically most likely number of states, which corresponds to the largest BIC value. The option that is currently applied is indicated by the red outline, but the user can apply alternative solutions by *clicking* on any of the other circles. In this example, the best solution is also applied. After performing the grouping, the groups can also be viewed in the *Intensity* tab (Figure B.4).

To demonstrate the advantage of grouping levels when fitting fluorescence lifetimes, the fitting results of three data sets are compared, as shown in Figure 5.3. (Note that this figure is not a direct output from *Full SMS* but was created by using exported data). The three samples for this case study are Alexa (see above), Qdot 605, and the light-harvesting complex II (LHCII). Qdot 605 is a commercially available type of CdSe/ZnS core-shell quantum dot and LHCII is the main light-harvesting complex of plants, a pigment-protein complex containing several energetically strongly connected chromophores that are responsible for absorbing photons that provide the necessary energy to drive the initial photochemical processes in the photosynthetic apparatus of higher plants. The first column of these panels (A, D, and G) shows the distributions of the numbers of photons used in each fitted level for both the non-grouped and grouped cases. As the total number of levels between the non-grouped and grouped cases differs, the Y-axis was normalised to facilitate comparison. The



☑ Use ROI

☐ Auto Apply

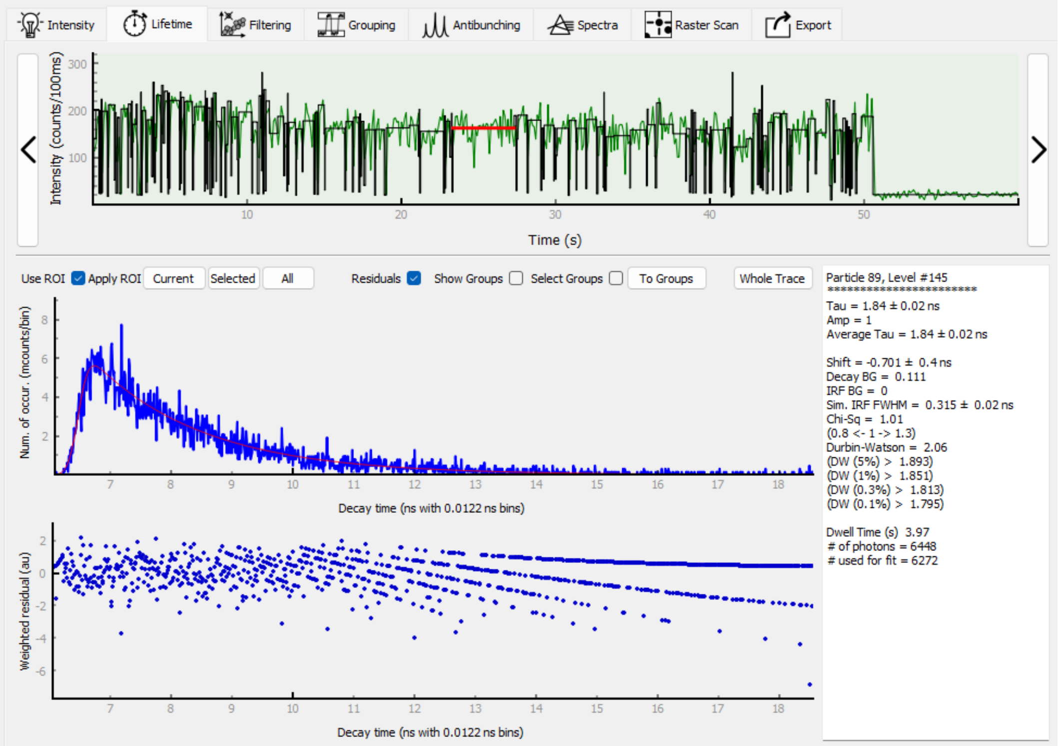
FIGURE 4.7: Screenshot of an intensity-level grouping of the Alexa intensity trace in Figure 4.3. See text and user guide for details.

second and third columns show two-dimensional distributions of the weighted average of the fitted lifetimes and the level intensity for each level, with the second column showing non-grouped levels and the third column showing grouped levels. It is clear from panels A, D, and G that the number of photons used to fit the lifetimes in the grouped levels is significantly larger than that when no grouping was done. The advantage of performing lifetime fitting on grouped levels is particularly evident for the Qdot 605 and LHCII data, which show the expected predominantly linear relationships between lifetime and intensity (panels F and I). A linear relationship is expected when the change in intensity is due to a change in the non-radiative decay rate [47, 76, 77]. This linear relationship is significantly broadened in the Alexa data (panel C), which can be explained by the heterogeneous distribution of the molecular orientations relative to the elliptically polarised excitation light resulting in varying absorption cross sections and, consequently, altered fluorescence intensities.

Although this grouping method already offers several advantages, the resolved intensity levels of each trace are grouped in isolation from other traces, which could hamper the successful resolution of the underlying states. For example, not all states are likely to be accessed in each measurement. In *Full SMS*, a global analysis can be performed where the region of interest (ROI) (i.e., trimmed trace—see below) is added to a single data set and then grouped, as shown in Figure B.4. Systematic intensity variations that may have resulted, for example, from focal drift during measurements will negatively affect the efficacy of a global grouping analysis. In a case where the lifetime and intensity are linearly correlated (like for Qdot 605 and LHCII in Figure 5.3) the user has the option to perform an intensity normalisation, which will greatly improve the result of such an analysis. A sufficiently large data set is likely to resolve the true states, but this operation is computationally expensive.

4.2.3 Fluorescence Decay Fitting

To extract fluorescence lifetime information, the photon arrival times relative to the excitation laser pulse are histogrammed and fitted in an iterative reconvolution minimisation procedure, which is based on either the least-squares (LS) [78] or maximum likelihood (ML) [79] approaches. The approaches give identical results for high photon counts, and in such cases LS is preferred since it is more computationally efficient and robust. For low photon counts, ML should be used since it correctly accounts for Poisson noise [79]. Lifetime fitting is performed within the *Lifetime* tab of the software (Figure 4.8). A fit can be performed on an entire experimental trace or individual resolved levels using only the photons from the level. After the grouping has been performed, a fit can also be done for each level group, as each intensity state is typically associated with a specific lifetime. This is especially powerful for dim levels or data with very fast switching, where the individually fitted levels do not contain many photons from which to extract a lifetime.



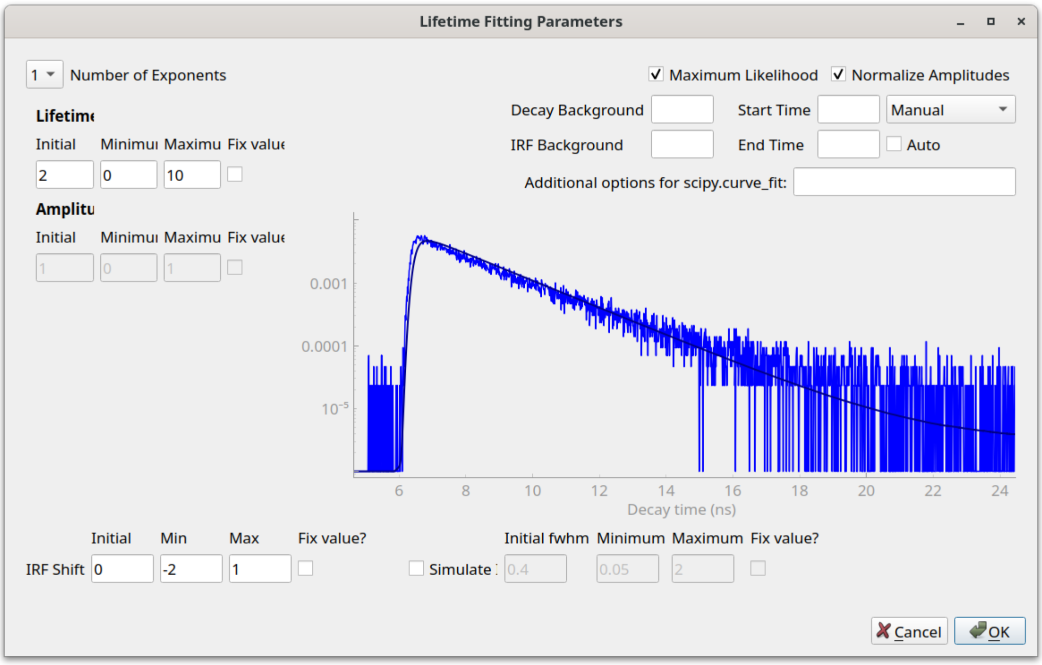
```
Particle 89, Level #145
***************************
Tau = 1.84 ± 0.02 ns
Amp = 1
Average Tau = 1.84 ± 0.02 ns

Shift = -0.701 ±  0.4 ns
Decay BG =  0.111
IRF BG =  0
Sim. IRF FWHM =  0.315 ±  0.02 ns
Chi-Sq =  1.01
(0.8 <- 1 -> 1.3)
Durbin-Watson =  2.06
(DW (5%) >  1.893)
(DW (1%) >  1.851)
(DW (0.3%) >  1.813)
(DW (0.1%) >  1.795)

Dwell Time (s)  3.97
# of photons = 6448
# used for fit = 6272
```

FIGURE 4.8: Screenshot of an example lifetime fit corresponding to a single resolved brightness level of the Alexa trace shown in Figure 4.3. Fitting residuals are shown in the lowest pane and fitting results are on the right. See text and user guide for details.

An interactive dialog is used to choose the parameters for the fitting (Figure 4.9). Multi-exponential fitting with up to three components is supported, and either a measured or simulated Gaussian IRF can be used, and decay and IRF background can be manually specified or determined automatically. After fitting(s), the results are shown in the *Lifetime* tab (Figure 4.8). The fits are shown along with the residuals, if enabled, and the fitting parameters and goodness-of-fit statistics are shown in the right pane. To evaluate the goodness-of-fit, we use the reduced χ^2 value (only for LS) and the Durbin–Watson (DW) parameter [80] (for both LS and ML). The latter measures autocorrelation in the residuals, which is more sensitive to minor fitting errors than χ^2. Our code also provides tools to automatically identify suitable boundaries for curve fitting, which is important for evaluating the fit using the DW parameter.

4.2.4 Second-Order Correlation Function

Single quantum emitters exhibit so-called antibunching of photons, as seen in the second-order correlation function ($g^{(2)}(\tau)$) of photon arrival times [81]. This is calculated as a cross-correlation between two individual photon channels from detectors in a Hanbury Brown–Twiss (HBT) configuration, since the dead time of a detector precludes the measurement of coincident photons by a single detector.


Lifetime Fitting Parameters

1 Number of Exponents
☑ Maximum Likelihood ☑ Normalize Amplitudes
Decay Background:
Start Time: Manual
IRF Background:
End Time: ☐ Auto
Additional options for scipy.curve_fit:

Lifetim
Initial: 2 · Minimu: 0 · Maximu: 10 · Fix valu ☐

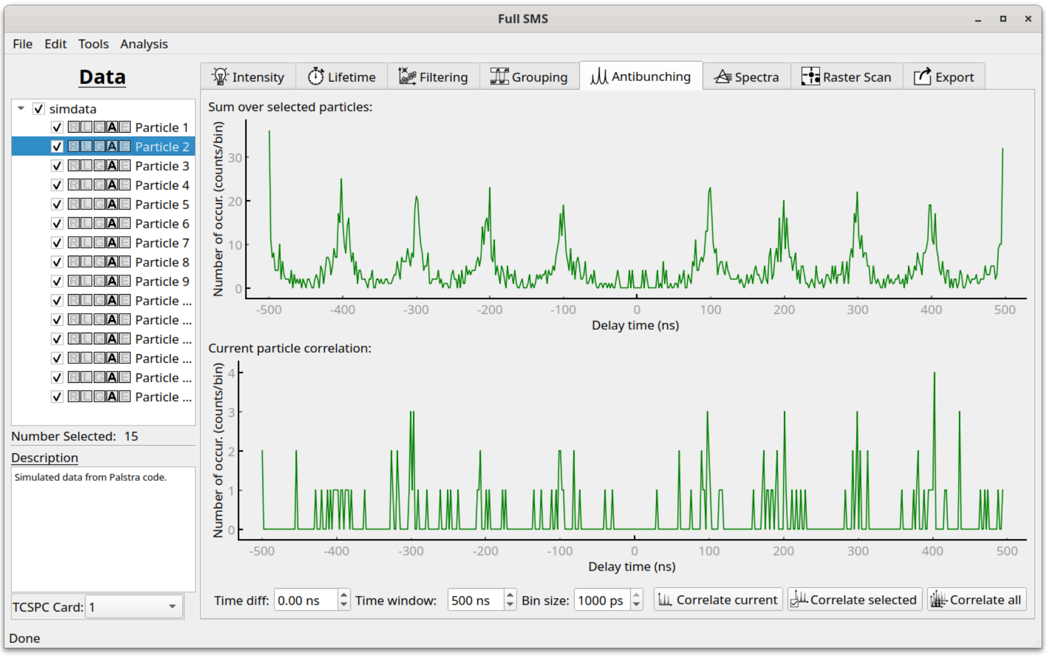
Amplitu
Initial: 1 · Minimu: 0 · Maximu: 1 · Fix valu ☐

IRF Shift: Initial: 0 · Min: -2 · Max: 1 · Fix value? ☐
☐ Simulate: Initial fwhm: 0.4 · Minimum: 0.05 · Maximum: 2 · Fix value? ☐

FIGURE 4.9: Screenshot of the user-definable parameters for the lifetime fitting, showing the decay histogram for the Alexa trace in Figure 4.3. At the top left is the number of lifetime exponents that should be fit. Below that, initial values and boundaries for the lifetimes can be provided. On the bottom left, the initial value and boundaries for the IRF shift can be defined. The user has the option for an IRF to be simulated, which can often result in better fits. Additional parameters can be defined on the top right, namely background values for the decay and IRF and the start and end times of the fit window to be used. These values can also be set automatically by the software. Lastly, at the top right are options to use maximum-likelihood fitting and to normalise the lifetime amplitudes.

The HBT configuration consists of a beam splitter dividing a stream of photons between two detectors connected to a coincidence unit and was its original application to measure the sizes of stars [82] before finding use in quantum optics. A perfect single emitter will never emit two photons at the same time, giving a value of $g^{(2)}(0) = 0$. The number of independent emitters, $1/(1 - g^{(2)}(0))$, is the reciprocal of the value of the "antibunching dip", $1 - g^{(2)}(0)$. *Full SMS* can calculate $g^{(2)}(\tau)$for two-channel data with an adjustable time window and bin size.

We demonstrate the analysis of photon antibunching using simulated data obtained from the code of Palstra and Koenderink [63] and converted to the *Full SMS* HDF5 format (Figure 4.10). In the *Antibunching* tab of the software, $g^{(2)}(\tau)$can be calculated over an adjustable time window and bin size. The sum of all selected particles' $g^{(2)}(\tau)$is shown in the top pane, which is useful in the case of low-intensity or short



Number Selected: 15

Description

Simulated data from Palstra code.

TCSPC Card: 1

Time diff: 0.00 ns

Time window: 500 ns

Bin size: 1000 ps

FIGURE 4.10: Antibunching analysis using second-order photon correlation histogram ($g^{(2)}(\tau)$). The bottom pane shows the current particle's correlation histogram, while the top pane shows the sum of correlations over all the selected particles, indicating clear evidence of photon antibunching through the absence of a zero-delay peak.

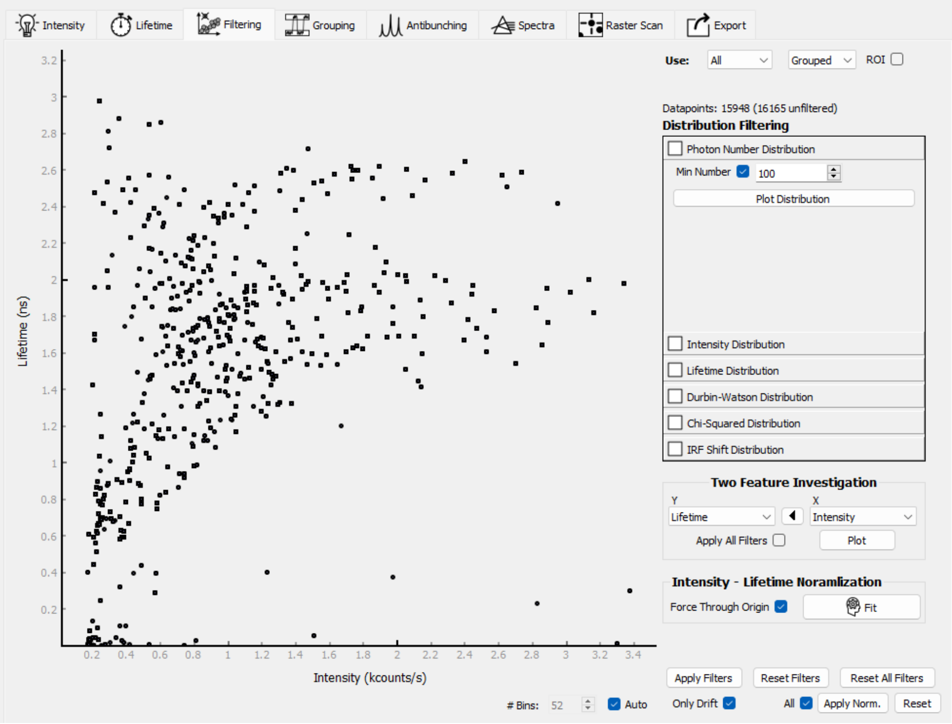
measurements where antibunching is not easily apparent in a single particle's correlation histogram. Note that while the demonstration shows a simulation for pulsed excitation, the calculation of $g^{(2)}(\tau)$following continuous-wave excitation is identical and can, therefore, also be done by *Full SMS*.

4.2.5 Data Filtering

In the *Filtering* tab of the software (Figure 4.11), the data can be filtered after level resolution, grouping, and lifetime fitting. The parameters used for filtering are the number of photons, the intensity, the average fitted lifetime, the DW parameter of the fit, χ^2 of the fit, and the fitted IRF shift for each level or group. A histogram of each parameter or a 2D scatter plot of two filter values can be plotted to aid in determining the appropriate filter parameters (as shown in Figure 4.11). Examples include filtering out levels with less than a certain number of photons (e.g. 100) or with lifetime or intensity values that are outliers. Filtering can be performed for individual particles, a selection, or the entire data set.

4.2.6 Display of Spectra

In the *Spectra* tab of the software (Figure 4.12), spectral time traces can be viewed, which are displayed as a 2D colour map of the photon counts (intensity) as a function of the bin time and wavelength on the X and Y axes, respectively. The colour scale can be adjusted for optimal contrast by setting the upper and lower thresholds, as well as the relative intensities of set colours along the scale. A histogram of pixel intensities is shown to aid in this regard. An ROI (separate from the 'trace ROI'—see below) can be defined, allowing a plot of intensity as a function of time or wavelength to be displayed.



Use: All Grouped ROI ☐

Datapoints: 15948 (16165 unfiltered)

☐ Photon Number Distribution

Min Number ☑ 100

☐ Intensity Distribution

☐ Lifetime Distribution

☐ Durbin-Watson Distribution

☐ Chi-Squared Distribution

☐ IRF Shift Distribution

Y Lifetime X Intensity

Apply All Filters ☐

Force Through Origin ☑

Bins: 52 ☑ Auto

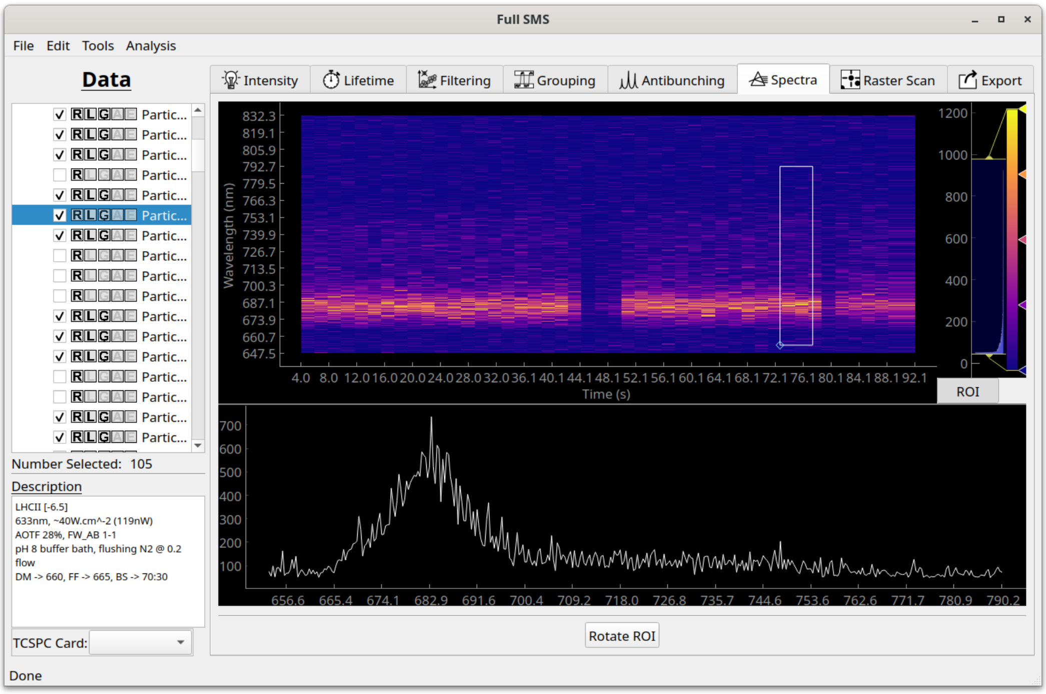
Only Drift ☑

All ☑

FIGURE 4.11: Data filtering example for Alexa, showing a scatter plot of intensity and lifetime values for grouped levels. The data is filtered to include only groups with more than 100 photons. The various filtering criteria are mentioned in the main text. The number of histogram bins is set automatically but can also be set manually under *#Bins* by first unselecting *Auto*, and options for intensity–lifetime normalisation. If *Only Drift* is checked, only intensities below the regression line will be normalised, else all intensities are normalised. Selecting *All* applies the normalisation to all levels and groups instead of only the currently displayed data.

4.2.7 Display of Raster Scan Images

The *Raster Scan* tab of the software (Figure B.5) displays raster-scan images. In SMS experiments, these images are mostly used to locate particles to be measured. Therefore, the functionality of this tab is currently limited to displaying images and indicating the position of the currently selected particle. However, the images can be easily exported if further image analysis or processing is required.



Number Selected: 105

Description

LHCII [-6.5]
633nm, ~40W.cm^-2 (119nW)
AOTF 28%, FW_AB 1-1
pH 8 buffer bath, flushing N2 @ 0.2 flow
DM -> 660, FF -> 665, BS -> 70:30

FIGURE 4.12: Example of a spectral measurement of a single LHCII pigment-protein complex, showing a full spectral time trace (top) and temporally summed spectrum of the selected ROI (bottom).

4.3 Additional Tools

In addition to the previously mentioned advantages of *Full SMS*, some additional *tools* have been developed. These tools add significant functionality to the analysis or enhance the user experience by automating common tasks.

4.3.1 Measurement Data Conversion

The native format of measurement data for which this software suite has been designed (HDF5) is custom, in that the particular structure and naming used were arbitrary choices that needed to be made. As mentioned above, *Full SMS* uses a custom HDF5 file format—custom in the sense that the specific grouping and naming of data inside the hierarchical format are customised. Therefore, if measurements are made in a format that differs from this, either in file format or structure, it will be necessary to convert the original data to the format that *Full SMS* can read.

This can be done *ad hoc* or built into the software. As an example of the latter, the software includes built-in conversion tools to convert measurements (with the appropriate structure) in either PicoQuant's '.pt3' format or plain text '.csv' format to the required '.h5' format. As an example of the former, we include two scripts to convert publicly available data from the Koenderink and Schlau-Cohen groups [83, 84], available on the GitHub repository under *src/conversion scripts*. These also act as examples for future contributors of how such a conversion can be done. Our software can also be easily extended to allow the analysis of Photon-HDF5 data.

4.3.2 Trace Region of Interest (ROI)

When SMS measurements are performed, depending on the nature of the sample being studied, it is almost certain that a single particle cannot be measured for an indefinitely long time. This is usually due to photodamage, as the excitation intensities are typically large. Particles often bleach during a measurement run and therefore it is often valuable to select a subsection of the measurement for analysis, excluding the data originating from a bleached particle (i.e. trimming the end of the intensity trace). However, in cases where data sets contain many measurements, applying some user-defined criteria to determine the ROI for each measurement individually is often a tedious task. To aid in this task, a tool has been included to automate the application of an ROI over the entire dataset (Figure B.6). The user defines a lower intensity threshold and a lower duration threshold. Resolved levels are considered until the last one that satisfies both thresholds. Each data set is analysed from back to front, and the point (if any) at which both of these thresholds are broken is set to be the end of the ROI. An example is shown in Figure 4.3, where the ROI is shown by the green shaded area, excluding the presumably bleached state at the end of the trace.

4.3.3 Photon Burst Detection

In certain types of SMS measurements, very short-lived spikes in intensity are recorded, a phenomenon that may be caused by cosmic rays or unwanted fluorescent particles diffusing through the focal volume. In most cases, such spikes skew the analysis and are therefore better to remove. *Full SMS* provides the user with two methods of identifying and removing photon bursts (Figure B.7). The first is based on the standard deviation (σ) of the mean intensity of each resolved level. A multiple of σ, as chosen by the user, above the mean intensity is used as a threshold definition. Alternatively, the user can define a manual intensity threshold. In either case, if any levels exceed the threshold, they are identified as potential photon bursts and can be removed.

4.4 Application Architecture

4.4.1 General Architecture

The program was written entirely in Python, using a package called PyQt5 to interface with the Qt GUI framework. The Qt GUI framework was chosen because it has rich features, works on most platforms (Windows/Mac/Linux), and allows rapid development of GUI windows using a GUI builder (Qt Designer). However, the primary complexity of adding a GUI is not in the design of the windows, but in the underlying framework that supports it.

Full SMS was designed using object-orientated programming and makes use of a *controller* architecture, in which several class instances are responsible for separate aspects of the program. The interaction with the GUI, as well as the functional implementations of the analysis, is handled by several different controller classes. For example, the intensity controller is responsible for converting time series to binned intensity traces and interacting with the GUI to display the selected intensity trace. However, much of the advanced functionality that is used within the controllers is implemented in separate modules. For example, the CPA code is found within its own module, but is called from within the intensity controller. In this manner, the code is organised into logical sections.

4.4.2 Multiprocessing

One of the critical limitations of the Python language[2], as implemented by CPython[3], is that it makes use of a global interpreter lock (GIL). The GIL is a necessary restriction that allows only one thread to execute code for a single Python process, effectively limiting the CPU core usage to a single core. Multi-threading is still possible by allowing a single process to switch between threads rapidly. In the case of a GUI, this allows the interface to remain responsive to the user, even if a separate thread is running to, e.g., perform an analysis. However, the GIL prevents effective usage of modern computing power when performing analyses that allow for parallelisation and would benefit from multiple CPU cores.

Several methods have been developed to achieve multicore processing, which, in most cases, involve starting multiple Python processes. The primary complexity then is the communication between processes.

The approach that the *Full SMS* software uses is as follows. Once the user interacts with the GUI, a thread is started to perform the analysis, while allowing the GUI to remain reactive. The analysis thread then uses Python's built-in `multiprocessing` module to start several Python processes, which will be referred to as worker processes. The number of worker processes created is automatically scaled to the number of CPU cores the system has available. The use of queues achieves communication between processes. Queues are populated by the analysis thread with pairings of copies of a function with serialised versions of the data to which the function should be applied. The functions and serialised data are loaded from the queue by

---

[2] This limitation is still the case as of Python version 3.12, but is planned to be removed in some future release.

[3] CPython is the most common implementation of Python written in C.

the worker processes and executed. Afterwards, the results are added to another queue in the form of serialised objects. The unfortunate consequence is that the parent process must consolidate the returned results into the original data structures. This can be done easily for simple objects, but for complex objects, as is the case for most of the objects in this software, great care needs to be taken to integrate all the relevant parts of the returned object. On completion of all the parts of the analysis, the parent process terminates the worker processes and the thread finishes.

4.4.3 Distribution and Compatibility

The repository for *Full SMS* is publicly hosted on GitHub and can be accessed at `http://github.com/BioPhysicsUP/Full_SMS`. The easiest way to run the software is via a Windows installer that can be downloaded from the repository releases page. This installer does not have any prerequisites, but is currently only available for Windows. Instructions for running from source can be found in the GitHub README. Detailed documentation of the software, including how to contribute, is available at `https://up-biophysics-sms.readthedocs.io/en/latest/index.html`. The data used for this paper can be downloaded at `https://doi.org/10.5281/zenodo.12150465`.

4.4.4 Measurement File Structure

Full SMS is designed to analyse data sets stored in HDF5 format, with a specific custom structure. This means that any data not in this format has to be converted before it can be opened. Currently, the software has built-in conversion tools for converting PicoQuant's '.pt3' format, as well as plain text '.csv' files, containing time-tagged time-resolved (TTTR) photon data, along with plain text spectral time traces. In principle, a wide range of data formats can be converted; however, a fundamental requirement is that the intensity data consist of photon arrival times and not pre-binned traces or traces obtained by gated detection (see Section 4.2.1).

Currently, up to two photon channels are supported. Our software provides for the display of raster scan images and spectra; however, these are not required to be present in the dataset for *Full SMS* to open it.

HDF5 Format

HDF5 (Hierarchical Data Format version 5) is a file format and set of tools designed to store and manage complex data. It is widely used in scientific computing and engineering due to its ability to efficiently handle large datasets, its flexibility to organise different types of data, and its compatibility with multiple platforms and programming languages. This format is particularly suitable for storing structured data, including multidimensional arrays, images, and time series data, along with the corresponding metadata.

One of the key strengths of HDF5 is its ability to handle very large datasets, even those that exceed the available memory of the system. This is achieved through fragmentation and compression, which allows data to be stored in smaller and more manageable pieces and reduces the overall file size. HDF5 also provides direct access capabilities, which means that users can access a specific portion of the data without having to load the entire dataset into memory. This feature is particularly valuable in big data applications, where processing efficiency is crucial.

Another important aspect of HDF5 is its support for different types of data, including integers, floating point numbers, and complex data structures. The flexibility to store a wide variety of data types, combined with the hierarchical organisation, makes HDF5 an excellent choice for applications that involve multiple related datasets. For example, in scientific experiments, HDF5 can be used to store raw sensor data, processed data, and metadata that describe the conditions under which the experiment was carried out, all within a single file.

HDF5 is also designed with parallel processing in mind, which makes it suitable for high-performance computing environments. It provides mechanisms for multiple processes to read and write to the same file concurrently, enabling efficient data management in large-scale simulations and experiments. The HDF5 API is available in many programming languages, such as Python, C, and MATLAB, further enhancing its versatility and ease of integration into existing workflows.

Groups in HDF5 are essential for structuring data in a hierarchical way. They act as containers that hold both datasets and other groups, allowing users to logically organise related data. This nested structure can be used to represent relationships between different parts of a dataset, making it easier to navigate and manage complex data. For example, a group might represent an entire experiment, with subgroups for individual trials, each containing relevant datasets. This approach helps maintain a clear and organised representation of the data, facilitating both data management and analysis.

Both groups and data sets in HDF5 can have attributes. Attributes are small pieces of metadata that provide additional information about a group or dataset. They are similar to annotations and can store descriptive information such as units of measurement, experimental conditions, or any other relevant context. Attributes are particularly useful because they allow users to attach meaningful metadata directly to the data, ensuring that important contextual information is readily available. This capability enhances the usability and interpretability of the data, especially in collaborative environments or when revisiting data sets after some time. Attributes in HDF5 can be of various data types, similar to datasets, including integers, floating-point numbers, and strings, providing flexibility in how metadata are represented.

In general, HDF5 is a powerful and flexible tool for managing complex data, offering a combination of efficient storage, rich metadata support, and scalability. Its design makes it ideal for use in data-intensive fields such as physics, astronomy, bioinformatics, and engineering, where managing large, heterogeneous datasets is a common requirement.

4.4.5 HDF5 for SMS

A common challenge faced by research groups that develop their own experimental systems is that the choice of any non-standard file format used to store measurements hampers sharing data between groups. Therefore, some time and effort was spent trying to choose a format that exactly addresses this barrier. One of the other formats that was considered, Photon-HDF5, is based on HDF5, and seems to have been more readily adopted by research groups. However, the primary issue with this format is that it was designed exclusively for time-based measurements. This meant that we would not be able to use the same format to include measurements of other kinds, such as raster scans (images), spectral measurements, etc.

Ultimately HDF5 was chosen for the following reasons: While it may not be that well adopted, it is far more flexible, and accounts for most experimental measurement types one could need. It also has better organisational features which are beneficial when performing multi-modal measurements. Additionally, it is relatively easy to understand and has SDKs for most programming languages, which are open-sourced, and therefore does not lock the data into the format exclusively. Lastly, it is not a custom in-house format but supported by a third-party group that works to maintain and improve on the format.

The hierarchical structure of an HDF5 file is used to organise the experimental data. This structure provides an effective way to store complex data in a systematic and easily navigable manner.

In this HDF5 structure, the top-level group is labelled "Particle". This group contains the Spectra dataset and a linked dataset called Raster Scans, which are associated with the Particle group. Each of these subgroups contains specific datasets and attributes that describe different aspects of the experimental data.

The Spectra dataset is represented as a two-dimensional array, where one dimension is Wavelength and the other is Time Series. This data set stores the spectral measurements over time. Additionally, there are attributes associated with the Spectra dataset, namely Wavelengths and Time Series, which provide contextual information such as the specific wavelengths measured and the timing of these measurements.

The Raster Scans dataset is linked to the Particle group and is represented as a two-dimensional array called Intensities, which represents measurements taken at different spatial positions. Each particle is associated with a raster scan, but a raster scan can be linked to multiple particles. By using linked datasets for raster scans, the need for data duplication is eliminated, as the raster scan data do not need to be replicated for each particle. This structure ensures efficient storage, reduces redundancy, and maintains data consistency by allowing multiple particles to reference the same raster scan dataset. The attributes for this data set are the X-axis values and the Y-axis values, which define the spatial coordinates where the measurements were recorded, enabling precise localisation within the raster scan.

The Particle group is more complex, containing multiple datasets that represent pulse relative/micro-time series and absolute/micro-time series. These data sets capture different temporal aspects of the particle measurements, providing both time-relative and absolute timing information. The attributes within the Particle group are extensive, including Measurement Description, User, Date, and Position in Raster Scan. These attributes provide critical contextual details, such as a description of the measurements, the identity of the user who performed the experiment, the date of the measurement, and the specific location within the raster scan where each particle was detected. Additionally, the relationship attribute links the particles to a specific raster scan, providing further insight into the spatial context of the particle data.

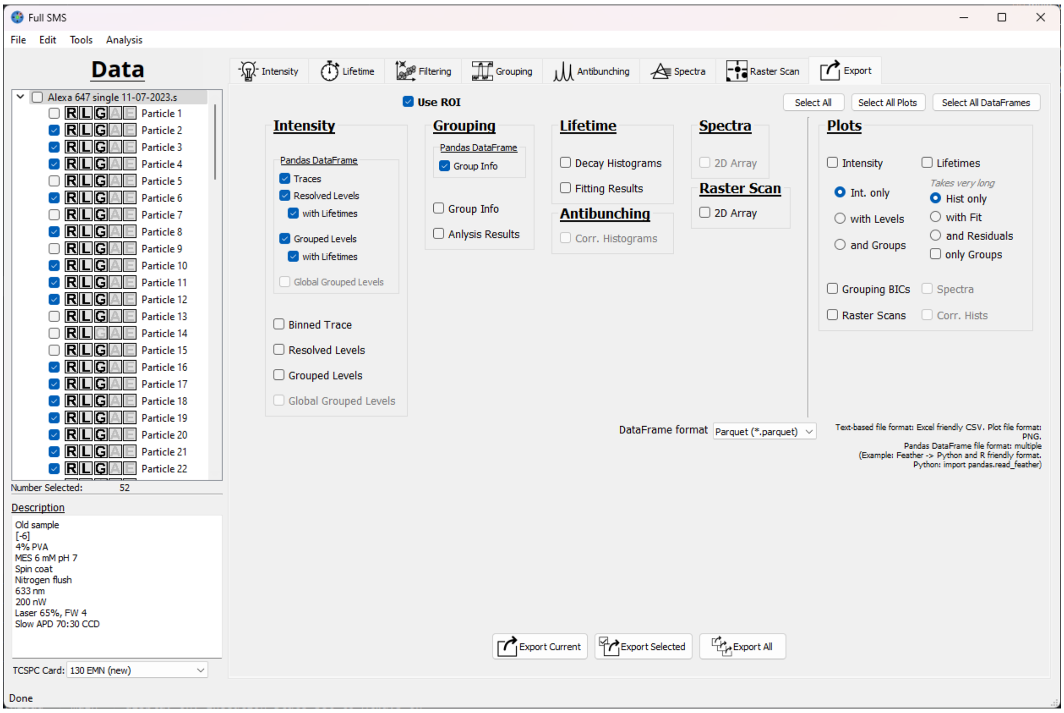
This hierarchical structure, using groups, datasets, and attributes, allows for an organised and detailed representation of the experimental data. Groups help to logically separate different types of measurements, while datasets store the core measurement data. Attributes enrich the datasets by providing additional contextual information, ensuring that the data are well documented and easily interpretable.

4.5 Data Export

In the *Export* tab of the software (Figure 4.13), the analysed data can be exported for further analysis. Binned intensity traces, resolved and grouped levels, and lifetime fitting results can all be exported as plain text, or in the form of a parquet file (a commonly used open-source column-orientated binary data storage format maintained by Apache), which enables more convenient downstream processing for users familiar with Pandas, a Python library for data manipulation and analysis, which can easily read and load parquet files. Apart from this final export, it is also possible to save the current state of the analysis to continue working on it at a later time. This includes the binned intensity, resolved, grouped, and fitted levels, as well as the current particle selection.

The data export functionality includes the ability to export both visual representations and raw data in multiple formats, facilitating further analysis and sharing of results. The exported formats include images (in PNG format) for visual representation of plots, as well as data files in CSV and Parquet formats for detailed numerical analysis. This flexibility in data export ensures compatibility with a wide range of software tools and workflows.

Export features are structured around two main types of data outputs: **Traces** and **Decay Curves**. Each type has multiple sub-options that define how the data can be exported based on specific groupings or features.



☐ Alexa 647 single 11-07-2023.s
☐ R L G A E Particle 1
☑ R L G A E Particle 2
☑ R L G A E Particle 3
☑ R L G A E Particle 4
☐ R L G A E Particle 5
☑ R L G A E Particle 6
☐ R L G A E Particle 7
☑ R L G A E Particle 8
☐ R L G A E Particle 9
☑ R L G A E Particle 10
☑ R L G A E Particle 11
☑ R L G A E Particle 12
☐ R L G A E Particle 13
☐ R L G A E Particle 14
☐ R L G A E Particle 15
☑ R L G A E Particle 16
☑ R L G A E Particle 17
☑ R L G A E Particle 18
☑ R L G A E Particle 19
☑ R L G A E Particle 20
☑ R L G A E Particle 21
☑ R L G A E Particle 22

Number Selected: 52

Description
Old sample
[-6]
4% PVA
MES 6 mM pH 7
Spin coat
Nitrogen flush
633 nm
200 nW
Laser 65%, FW 4
Slow APD 70:30 CCD

TCSPC Card: 130 EMN (new)

☑ Use ROI

Intensity
Pandas DataFrame
☑ Traces
☑ Resolved Levels
☑ with Lifetimes
☑ Grouped Levels
☑ with Lifetimes
☐ Global Grouped Levels
☐ Binned Trace
☐ Resolved Levels
☐ Grouped Levels
☐ Global Grouped Levels

Grouping
Pandas DataFrame
☑ Group Info
☐ Group Info
☐ Anlysis Results

Lifetime
☐ Decay Histograms
☐ Fitting Results

Antibunching
☐ Corr. Histograms

Spectra
☐ 2D Array

Raster Scan
☐ 2D Array

Plots
☐ Intensity
● Int. only
○ with Levels
○ and Groups
☐ Lifetimes
Takes very long
● Hist only
○ with Fit
○ and Residuals
☐ only Groups
☐ Grouping BICs
☐ Spectra
☐ Raster Scans
☐ Corr. Hists

DataFrame format Parquet (*.parquet)

Text-based file format: Excel friendly CSV. Plot file format: PNG.
Pandas DataFrame file format: multiple (Example: Feather -> Python and R friendly format. Python: import pandas.read_feather)

FIGURE 4.13: Exporting tab of the software. The needed data can be selected, which can then be exported for the current, selected, or all particles.

Traces Export

Traces can be exported with different levels of detail:

- **With Levels**: This option includes the trace data along with associated levels.
- **With Groups**: This option includes the definitions of the groups identified in the trace.

For the Traces export, the output can be provided as a PNG image showing the plot of the traces, which is useful for visual inspection and presentation. Additionally, the trace data itself can be exported in CSV or Parquet format, which can then be further processed or analysed using statistical or data analysis software.

Decay Curves Export

Decay Curves can also be exported in a variety of ways, each providing distinct information:

- **For Traces**: This option exports the decay curves for each whole trace.
- **For Levels**: This option exports the decay curves for each fitted level.
- **For Groups**: This option exports the decay curves for each fitted group.
- **With Fitting**: This option includes fitted decay curves alongside the raw decay data.
- **With Residuals**: Exporting decay curves with residuals provides a means to evaluate the differences between the fitted model and the actual data, providing insight into the accuracy of the fitting process.

Similarly to traces, decay curves can be exported both as PNG images for visual inspection and as CSV or parquet format for numerical analysis. The inclusion of fitted curves and residuals helps users assess the quality of their data models and make informed decisions about the next steps in their analysis.

4.6 Conclusion

We have described a new GUI-based application for the advanced analysis of single-molecule spectroscopic data. It uses a custom HDF5 file format that is uniquely suited to multi-parameter measurements on single particles. The software is user-friendly, requires no programming knowledge to use, while also being fully open source and easily extendable. It allows analysis of photon-by-photon data to extract fluorescence intensity change points and groups the resulting intensity levels using hierarchical agglomeration. It also allows for fitting fluorescence lifetimes of individual and grouped levels, the calculation of the second-order intensity correlation function, data filtering options, and a few additional features.

Full SMS is suited for further extensions. An extension of multichannel functionality to include multichannel change point detection [85] would be useful for analysing smFRET data. The change-point analysis functionality could be further extended to allow for analysis of binned data. For the grouping analysis, the relative likelihood values can be added as calculated from the BIC [86], for easier interpretation. The lifetime functionality could be extended to allow fitting with more than three exponential components and with decay models other than a multi-exponential, such as lifetime distributions. Functionality could be added to analyse spectra (using fitting or phasors) and to analyse fluorescence correlation spectroscopy (FCS) and real-time feedback-driven single-particle tracking [27] data. This will allow for correlation of lifetime, spectral, and diffusion properties. Adding the ability to interface directly with other data formats, such as Photon-HDF5 [64], would allow more groups to start using the software immediately.

Chapter 5

Case Studies

What follows is a description of the samples used and then a series of comparative case studies that are meant to serve as illustrations of the efficacy and benefits that *Full SMS* offers as an analysis tool. It is important to note that the analyses here are not direct outputs of the software, but are products of secondary analyses that utilise the unique outputs of *Full SMS*. The experimental setup used for these case studies is described in Section 2.2 and the preparation of the samples are discussed in Section 2.4.

5.1 Analysis of Light Harvesting Complex II

The study of LHCII isolated from spinach provides further insights into the dynamics of fluorescence at different quenching states. In Figure 5.1, a histogram of photon counts in lifetime fits is presented. The histogram compares resolved and grouped data, showing the distribution of photon counts across various intensity levels on a logarithmic scale. The comparison shows a clear improvement in the number of photons that were used for the lifetime analysis; this serves as one of the significant advantages that level grouping offers.

Further analysis of LHCII fluorescence lifetime and intensity distribution is shown in Figure 5.2. Panel A shows the distribution of lifetime versus intensity for the resolved data, where a wider spread across both axes suggests more variability in

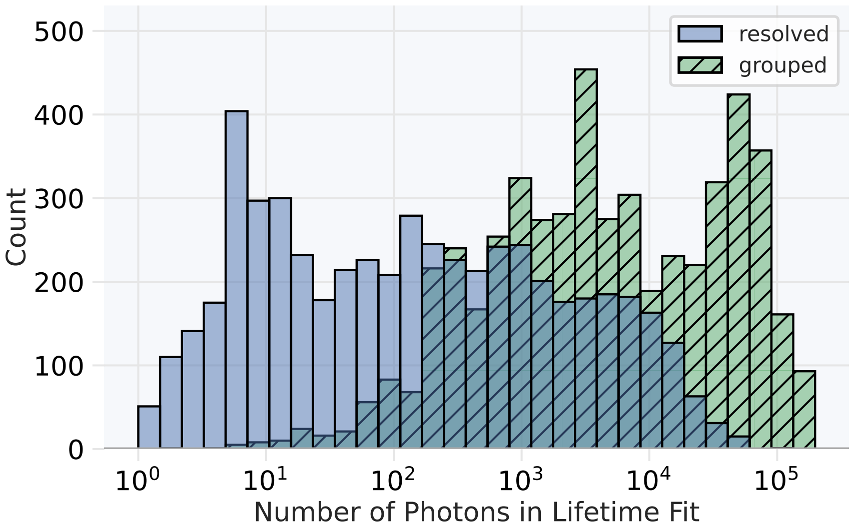


FIGURE 5.1: Histogram showing the distribution of photon counts in lifetime fits, comparing resolved and grouped data. The x-axis represents the number of photons on a logarithmic scale, while the y-axis shows the count frequency. The resolved data (blue) and grouped data (green with hatches) indicate distinct distributions across photon count levels.

emission. In panel B, the grouped data show a tighter distribution, with a regression line fit indicating a more constrained relationship between the lifetime and intensity. This reduced spread upon grouping suggests that grouping may improve consistency in observed emission characteristics. It also allows for a reasonable linear regression fitting of the relationship between lifetime and fluorescence intensity in LHCII, the nature of which has already been shown to be linear when emulating non-saturated conditions [87].

5.2 Comparative Study

To further demonstrate the advantage of grouping levels when fitting fluorescence lifetimes, the fitting results of three datasets are compared, as shown in Figure 5.3. The first column of these panels (A, D, and G) shows the distributions of the numbers of photons used in each fitted level for both the non-grouped and grouped cases. As the total number of levels between the non-grouped and grouped cases differs,

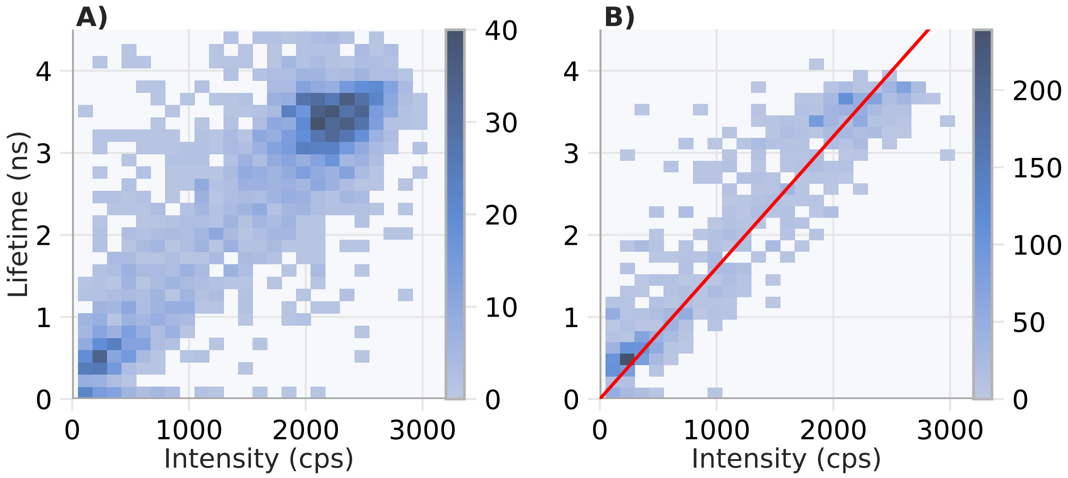


FIGURE 5.2: Comparison of LHCII Lifetime and Intensity Distributions for Resolved and Grouped Levels. **Panel A)** shows the distribution of lifetime (ns) versus intensity (cps) for the resolved level, exhibiting a broader spread across both axes. **Panel B)** displays the distribution for the resolved and then grouped levels, with a fitted regression line (red) highlighting the relationship between intensity and lifetime. The grouping reduces the distribution's spread, indicating a more constrained relationship between the two parameters at the grouped level. The colour intensity in each panel represents density, as indicated by the colour bar on the right.

the Y-axis was normalised to facilitate comparison. The second and third columns show two-dimensional distributions of the weighted average of the fitted lifetimes and the level intensity for each level, with the second column showing non-grouped levels and the third column showing grouped levels. It is clear from panels A, D and G that the number of photons used for fitting the lifetimes in the grouped levels is significantly larger than when no grouping was done. The advantage of performing lifetime fitting on grouped levels is particularly evident for the Qdot 605 and LHCII data, which shows the expected predominantly linear relationships between lifetime and intensity (panels F and I). This linear relationship is significantly broadened in the Alexa data (panel C), which can be explained by the heterogeneous distribution of the molecular orientations relative to the elliptically polarised excitation light resulting in varying absorption cross sections and, consequently, altered fluorescence intensities.

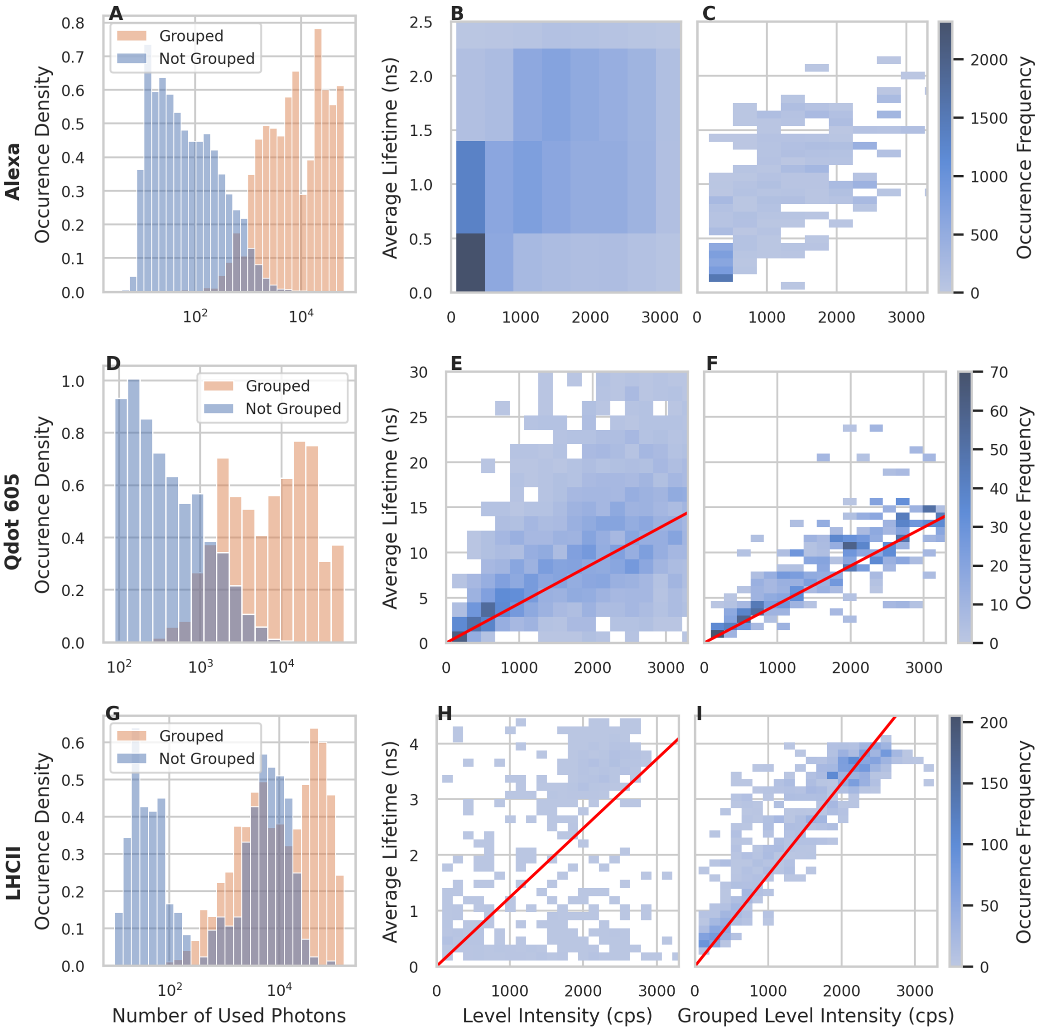


FIGURE 5.3: Lifetime fitting results of three datasets: Alexa Fluor 647 dye, Qdot 605 (CdSe/ZnS quantum dots), and LHCII. The first column shows the normalised distributions of the numbers of photons used in each fitted level for both the non-grouped and grouped cases. The second and third columns depict the relationship between the weighted average of the fitted lifetimes for each level and the corresponding level intensity, where the second column is for the non-grouped levels and the third column is for the grouped levels. The red lines denote linear fits through the origin.

In all three samples, the combination of change point analysis with grouping provides a distinct advantage, predominantly in terms of lifetime fitting.

Furthermore, the results show that the lifetimes for Qdot 605 vary between ~ 4 ns and ~ 14 ns. The quenched lifetime corresponds well with the literature, whereas the unquenched lifetime is somewhat lower than the reported value, but in the same order of magnitude [88]. This could indicate that the sample was quenched, which would likely be due to environmental effects. Similarly, Ref. [89] reports a long lifetime of Alexa 647 to be ~1.2 ns, and a fast lifetime component of ~0.4 ns, which corresponds with our results. Lastly, in the case of LHCII, both the linear relationship between fluorescence and lifetime, as well as the lifetimes of the quenched and unquenched states agree with literature [56].

5.3 Fluorescence Intermittency Models

To understand the transition dynamics between different quenching states, we compared the distributions of intensity, dwell time, and photon counts in different states (see Figure 5.4). Most previous studies investigate fluorescence intermittency in LHCII with a two-state model: switching between an *on* and *off* state [34, 56, 90–92]. In these studies, thresholds are used to define the *on* and *off* states, an example of which can be seen in Figure 5.3A.

The grouping analysis, followed by intensity normalisation due to the linear relationship between fluorescence lifetime and intensity in LHCII, enabled us to perform a Gaussian fit of the *on* and *off* population distributions. In turn, we could consequently define the *on* and *off* states with two thresholds given by a 2σ deviation from the maximum—as indicated by the red long-dashed lines instead of the black short-dashed line for the midpoint-based threshold in Figure 5.5C. This resulted in the identification of an *intermediate* population between the *on* and *off* states. This then allowed for a three-state model to be used: *on*, *intermediate* and

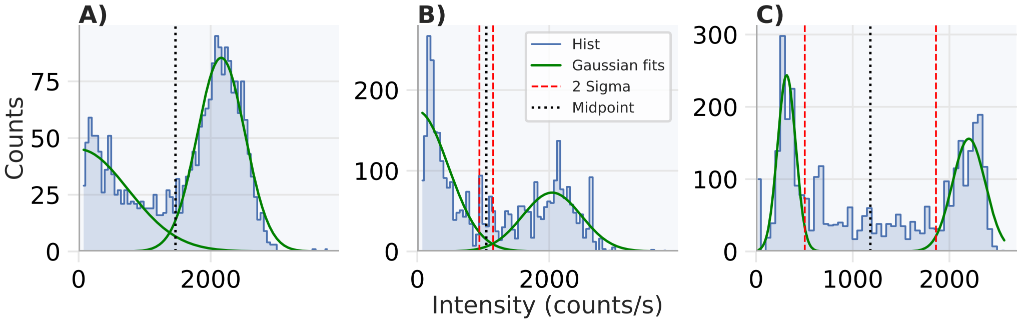


FIGURE 5.4: Histogram distributions of approximately 100 LHCII fluorescence intensity measurements, illustrating the effects of change-point and grouping analyses on intensity distribution. **Panel A)** Distribution of resolved levels obtained solely from change-point analysis, showing two distinct peaks corresponding to different intensity states. **Panel B)** Distribution of grouped levels after applying grouping analysis, which enhances the resolution between the two primary states (on and off). **Panel C)** Distribution of normalised intensities, where normalisation is performed based on a linear relationship between intensity and fluorescence lifetime (see Figure 5.3) to correct for inconsistencies in measurement conditions affecting the maximum measurable intensity. The progression from **Panels A) to C)** demonstrates that grouping and normalisation improve the separation between the on and off states, revealing three distinct populations: on, off, and intermediate levels. The green line represents Gaussian fits to the peaks, the black dotted line indicates the midpoint between the peaks, and the red dashed lines mark a 2-sigma deviation from each peak.

off. Lastly, this can be compared with a multi-state model, in which switches to a higher state and switches to lower states form two populations. Due to the relatively low signal-to-noise ratio (SNR) of the measurement, it was necessary to perform intensity normalisation to resolve the *intermediate* state. However, with a higher SNR, the grouping alone should be able to improve the distribution sufficiently for a three-state model to be used.

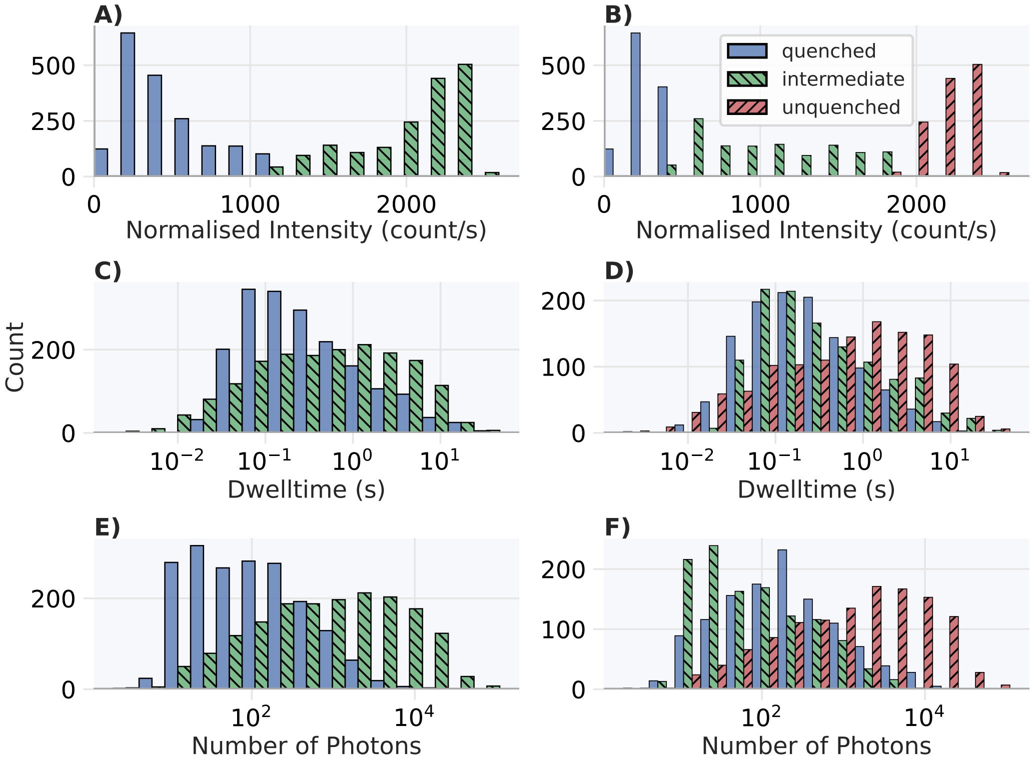


FIGURE 5.5: Distribution Analysis of LHCII Intensity, Dwell Time, and Photon Counts for Different Quenching States. **Panels A and B** display histograms of normalised intensity (in counts/s) for quenched, intermediate, and unquenched states, illustrating differences in intensity distributions across states. **Panels C and D** show the distribution of dwell times (in seconds) for each quenching state, highlighting variations in dwell time profiles. **Panels E and F** present photon count distributions, with different quenching states demonstrating unique patterns in photon accumulation. Each histogram provides insight into the characteristics of the respective quenching states across different parameters.

Figure 5.5 presents histograms that display the distribution of intensity, dwell time, and photon counts across different quenching states. Panels A and B show the normalised intensity distributions, while Panels C and D reveal dwell time distributions across quenched, intermediate, and unquenched states. Finally, panels E and F illustrate the patterns in photon accumulation, which vary significantly between quenching states. This analysis helps to clarify the photophysical differences between states and highlights the effect of quenching on fluorescence behaviour. We see that the dwell time and photon number distributions of the *intermediate* states follow closely those of the *off* states. It is therefore possible that the *intermediate* states are more related to the *off* states than the *on* states.

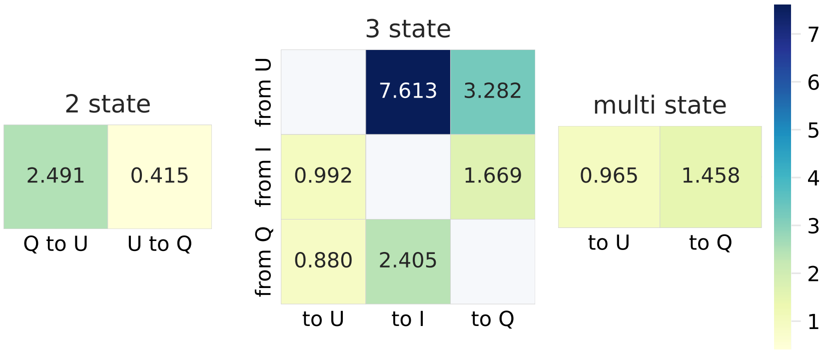


FIGURE 5.6: Transition Rates Between Quenching States in Two-State, Three-State, and Multi-State Models. Heatmaps display the estimated transition rates between quenching states: quenched (Q), unquenched (U), and intermediate (I). Each sub-panel represents a different model: the two-state model (left) shows transitions between Q and U; the three-state model (centre) includes an intermediate state (I) with transitions among Q, U, and I; and the multi-state model (right) generalises transitions without explicit intermediate states. Colour intensity reflects the magnitude of transition rates, as indicated by the colour bar, with higher values shown in darker shades.

The transition rates between different quenching states are further analysed in Figure 5.6. The heat maps depict estimated transition rates in two-state, three-state, and multi-state models. This comparative analysis across models provides insights into the dynamic behaviour of fluorescence transitions.

We see from the three-state model that not only are the intermediate states more often accessed from the unquenched state than the quenched state, but also the switching from the unquenched state to the intermediate state is the most common switch by some margin. The high unquenched-to-intermediate switching rate could be indicative of instability of the unquenched state. This could be due to slight sample degradation that results in pigments briefly decoupling from the excitonic bath, and therefore effectively lowering the absorption cross-section, which would also explain why this deviation from the expected two-state model occurs.

The three-state model does, however, show us that the more accurate quenched-to-unquenched and unquenched-to-quenched rates differ considerably from the simpler two-state model. We see, with the intermediate state being resolved, that the quenched state is more often populated by the unquenched state than the unquenched state is populated by the quenched state, which is the opposite of what the two-state model indicates. This is also more in line with the multi-state model, which due to its design would provide a more reliable indication of the overall switching direction, but tends to oversimplify the true dynamics. This shows that unless it is incorporated in the analysis, the presence of the intermediate states could lead to faulty interpretations, a result that ultimately would not have been possible without level grouping and the improved lifetime fitting enabled.

Both the multi-state model and the preferential population of states observed here have also been observed for bacterial antenna complexes (LH2) from purple bacteria [93, 94].

5.4 Conclusion

What the case studies reveal is that the analysis results of *Full SMS* enable users to derive insights from SMS measurements. What is more, a fraction of the time and significantly less effort are required to analyse SMS data, due to the ease-of-use that *Full SMS* provides. Additionally, the grouping analysis can lead to novel findings by improving the lifetime fittings.

Chapter 6

Summary

Most of the energy that sustains life on earth can be traced back to photosynthesis, and so its importance is difficult to overstate. Our understanding of this crucial process is still growing as we develop ever-advanced techniques to investigate the world of proteins and their interactions with light. To this end, this thesis provides a thorough introduction of the *Full SMS* software suite, which is the main contribution this work has made to the field through its utility in investigating photosynthetic systems in single molecule spectroscopy studies. It presents the foundation that the analysis is built on, as well as develops the technical solutions that the software had to solve. These include the framework that organises and orchestrates the analysis, the development of a multiprocess/parallel processing strategy, as well as data filtering and other crucial tools. Essentially, this allows investigators to achieve meaningful analyses in hours instead of days, and also lowers the technical ability users require to analyse SMS measurements. The thesis also showcases how the analysis is applied to two common single-photon-emission fluorescent molecules, as well as a complex organic pigment complex that plays a critical role in photosynthesis. In all case studies, the advantages gained from analysis are shown, and the observed fluorescent properties broadly agree with literature. Lastly, the thesis also demonstrates the development and prototyping of a novel experimental device used to expand on the capabilities of the in-house single molecule spectroscopy setup. In this way, this thesis makes a significant contribution to the field of single-molecule

biophysics, presenting a tool to further the investigation of the light-dependent processes in photosynthesis which take place at the nanometre scale.

The *Full SMS* suite, as of the time of writing, is in a constant state of improvement as a code base which can be expanded to include additional features. The following possible features have been identified, to name a few: an improved ability for datasets measured on other setups to be converted or read natively—work on which is currently ongoing; an expansion on the two time channels currently provisioned for to allow any number of channels, thereby enabling multichannel measurements; and an introduction of more change point analysis and grouping methods which can be selected for use by the user. To this point, it is worth investigating Photon-HDF5 as an alternative or supplementary file format, due to its broad adoption. It would also be useful to include an embedded Jupyter notebook platform to facilitate ad-hoc analyses. There are also several optimisations that could be made to improve the performance of analysis, such as just-in-time (JIT) compiler frameworks for Python such as Numba or Python experimental support for JIT compilation. Finally, the addition of other grouping analysis methods would provide the user with several options.

Appendix A

Code Sections

A.1 Run Change Point Analysis

```python
def run_cpa(self, all_sums: CPSums, confidence=None, end_time_s=None):
    """
    Runs the change point analysis.

    If the ChangePointAnalysis wasn't initialised with a confidence interval, or if
    the analysis is to be rerun with a new confidence interval, this method starts said analysis.

    Parameters
    ----------
    end_time_s: float
        Time at which to end analysis. If not provided the whole trace will be used.
    confidence: float
        Confidence interval. Valid values are 0.99, 0.95, 0.90 and 0.69.

    Returns
    -------
    num_cpts: int
        Number of change points detected
    cpt_inds: ndarray
        Indexes of change points
    conf_regions: list(tuple(int, int))
        Index region corresponding to confidence interval
    dt_uncertainty: ndarray
        Array of uncertainty in time corresonding to confidence interval
    """
    self.load_settings()
    if confidence is not None:
        assert confidence in [
            0.99,
            0.95,
            0.90,
            0.69,
        ], "ChangePointAnalysis:\tConfidence value given not valid."
        self.confidence = confidence
        self._tau = TauData(confidence)
    else:
        assert (
            self.confidence is not None
        ), "ChangePointAnalysis:\tNo confidence value provided."

```

```
    if end_time_s is not None:
        if self._abstimes.size != 0:
            self.end_at_photon = np.argmax(self._abstimes[:] > (end_time_s * 1e9))
        else:
            self.end_at_photon = 0
        if self.end_at_photon == 0:
            self.end_at_photon = self.num_photons
    self._find_all_cpts(all_sums=all_sums)
    self.has_run = True
```

A.2 Find Change Points

```
def _find_all_cpts(
    self,
    all_sums: CPSums,
    _seg_inds: Tuple[int, int] = None,
    _side: str = None,
    _prev_cpt: int = None,
):
    """
    Find all change points in particle.

    Recursive function that finds all change points that meets the confidence criteria.

    .. note::
        The first call doesn't need to be called with any parameters.

    .. note::
        The top level assigns the number of detected change points to
        the .num_cpts attribute of this instance of ChangePointAnalysis.

    Parameters
    ----------
    _prev_cpt : int, Optional
        The index of the change point for the right leg (the next_seg_start if side =
        ↪  'left').
    _seg_inds : (int, int), Optional
        The index of the segment that is to be searched. Calculated by _next_seg_ind
        ↪  method.
    _side : str, Optional
        Determines current segment is left or right of a previously
        detected change point. Valid values are 'left' or 'right'.
    """
    is_top_level = False

    if self._finding is False:
        is_top_level = True
        self._finding = True
        assert _seg_inds is None, (
            "ChangePointAnalysis:\tDo not provide seg_inds when calling, it's used for
            ↪   "
            "recursive calling only. "
        )

    if is_top_level:
        _seg_inds = self._next_seg_ind()
        self._i = 0
    else:
        _seg_inds = self._next_seg_ind(
            prev_seg_inds=_seg_inds, prev_change_point=_prev_cpt, _side=_side
        )
```

```python
    self._i += 1

    if _seg_inds != (None, None):
        cpt_found, cpt = self.__weighted_likelihood_ratio(all_sums, _seg_inds)

        if cpt_found:
            # Left side of change point
            self._find_all_cpts(all_sums, _seg_inds, _side="left", _prev_cpt=cpt)
            # Right side of change point
            self._find_all_cpts(all_sums, _seg_inds, _side="right", _prev_cpt=cpt)
            pass  # Exits if recursive

        if self.end_at_photon is not None:
            end_ind = self.end_at_photon
        else:
            end_ind = self.num_photons

        if _seg_inds[1] + 9 <= end_ind and _side is None:
            self._find_all_cpts(all_sums, _seg_inds, _prev_cpt=cpt)
```

A.3 Find Segment

```python
def _next_seg_ind(self, prev_seg_inds: Tuple[int, int] = None, prev_change_point=None, _side=None):
    """
    Determines the next segment indices for change point detection.

    Parameters:
    - prev_start (int, optional): Start index of the previous segment.
    - prev_end (int, optional): End index of the previous segment.
    - change_point (int, optional): Index of the detected change point in the previous segment.
    - _side (str, optional): Indicates where the change point was detected ('left', 'right', or None).
    - change_points (list, optional): List of all detected change points.

    Returns:
    - (int, int): Start and end indices of the next segment.
    """

    if prev_seg_inds is None:
        # First call
        start = 0
        end = min(self.num_photons - 1, 999)
    else:
        prev_start, prev_end = prev_seg_inds
        if _side == 'left':
            # Change point detected on the left side
            start = prev_start
            end = prev_change_point - 1
        elif _side == 'right':
            # Change point detected on the right side
            start = prev_change_point
            end = prev_end
        else:
            # No change detected, overlap by 200 data points
            start = prev_end - 200
            # Adjust start if there's a change point in the overlap region
            cpts_in_overlap = [cp for cp in self.cpt_inds if start <= cp <= prev_end]
            if len(cpts_in_overlap) > 0:
                start = max(cpts_in_overlap)
```

```
            end = start + 1000
            end = min(end, self.num_photons - 1)  # Ensure end index is within dataset bounds
    return start, end

    return next_start_ind, next_end_ind
```

A.4 Weighted Likelihood Ratio Test

```
def __weighted_likelihood_ratio(
    self, all_sums: CPSums, seg_inds=None
) -> Tuple[bool, Optional[int]]:
    """
    Calculates the Weighted & Standardised Likelihood ratio and detects the possible change point.

    Based on 'Detection of Intensity Change Points in Time-Resolved Single-Molecule Measurements'
    from Watkins nad Yang, J. Phys. Chem. B 2005, 109, 617-628 (http://pubs.acs.org/doi/abs/10.1021/jp0467548)

    If the possible change point is greater than the tau_a value for the corresponding
    confidence interval and number of data points the detected change points, it's
    confidence region (as defined by tau_b), and the corresponding uncertainty in time
    is added to this instance of ChangePointAnalysis.

    Parameters
    ----------
    seg_inds : (int, int), optional
        Segment indexes (start, end).

    Returns
    -------
    cpt_found : bool
        True if a change point was detected.
    cpt : int, Optional
        The index of the change point, if one was detected.
    """

    # settings
    min_num_photons = self.settings.cpa_min_num_photons
    min_boundary_offset = self.settings.cpa_min_boundary_offset

    assert (
        type(seg_inds) is tuple
    ), "ChangePointAnalysis:\tSegment index's not given."
    start_ind, end_ind = seg_inds
    n = end_ind - start_ind
    assert (
        n <= 1000
    ), "ChangePointAnalysis:\tIndex's given result in more than a segment of more than 1000 points."
    if n < min_num_photons:
        cpt_found = False
        return cpt_found, None
    time_data = self._abstimes[start_ind:end_ind]

    ini_time = time_data[0]
    period = time_data[-1] - ini_time

    wlr = np.zeros(n, float)
```

```

    # sig_e = np.pi ** 2 / 6 - sum(1 / j ** 2 for j in range(1, (n - 1) + 1))
    sig_e = all_sums.get_sig_e(n)

    for k in range(2, (n - 2) + 1):  # Remember!!!! range(1, N) = [1, ... , N-1]
        sum_set = all_sums.get_set(n, k)

        cap_v_k = (time_data[k] - ini_time) / period  # Just after eq. 4

        # u_k = -sum(1 / j for j in range(k, (n - 1) + 1))  # Just after eq. 6
        u_k = sum_set["u_k"]

        # u_n_k = -sum(1 / j for j in range(n - k, (n - 1) + 1))  # Just after eq. 6
        u_n_k = sum_set["u_n_k"]

        l0_minus_expec_l0 = (
            -2 * k * np.log(cap_v_k)
            + 2 * k * u_k
            - 2 * (n - k) * np.log(1 - cap_v_k)
            + 2 * (n - k) * u_n_k
        )  # Just after eq. 6

        # v_k2 = sum(1 / j ** 2 for j in range(k, (n - 1) + 1))  # Just before eq. 7
        v2_k = sum_set["v2_k"]

        # v2_n_k = sum(1 / j ** 2 for j in range(n - k, (n - 1) + 1))  # Just before
        ↪  eq. 7
        v2_n_k = sum_set["v2_n_k"]

        sigma_k = np.sqrt(
            4 * (k**2) * v2_k
            + 4 * ((n - k) ** 2) * v2_n_k
            - 8 * k * (n - k) * sig_e
        )  # Just before eq. 7, and note errata
        w_k = (1 / 2) * np.log((4 * k * (n - k)) / n**2)  # Just after eq. 6

        wlr.itemset(
            k, l0_minus_expec_l0 / sigma_k + w_k
        )  # Eq. 6 and just after eq. 6

    max_ind_local = int(wlr.argmax())

    cpt = None
    if (
        max_ind_local >= min_boundary_offset
        and n - max_ind_local >= min_boundary_offset
        and wlr[max_ind_local] >= self._tau.get_tau_a(n)
    ):
        cpt = max_ind_local + start_ind
        self.cpt_inds = np.append(self.cpt_inds, max_ind_local + start_ind)
        tau_b_inv = wlr.max() - self._tau.get_tau_b(n)
        region_all_local = np.where(wlr >= tau_b_inv)[0]
        region = (region_all_local[0] + start_ind, region_all_local[-1] + start_ind)
        dt = self._abstimes[region[1]] - self._abstimes[region[0]]
        self.conf_regions.append(region)  # list, not ndarray
        # self.dt_uncertainty = np.append(self.dt_uncertainty, dt)
        cpt_found = True
    else:
        cpt_found = False

    return cpt_found, cpt
```

A.5 Run Grouping

```python
def run_grouping(self, feedback_queue: JoinableQueue = None):
    """
    Run grouping

    Returns
    -------

    """

    try:
        if self.has_groups:
            self.clear_and_backup_results()
        if self._particle.has_levels:
            steps = []
            if self._particle.num_levels == 1:
                self.steps = [ClusteringStep(self._particle, single_level=True)]
                self.num_steps = 1
                self.best_step_ind = 0
                self.selected_step_ind = 0
                self.has_groups = True
                logger.info(f"{self._particle.name} has only one level")
                logger.info(self.steps)
            else:
                c_step = ClusteringStep(self._particle, first=True)
                if not self.use_roi_for_grouping:
                    current_num_groups = self._particle.num_levels
                else:
                    current_num_groups = self._particle.num_levels_roi
                if feedback_queue is not None:
                    feedback_queue.put(
                        ProcessProgressTask(
                            task_cmd=ProcessProgressCmd.Start,
                            args=(current_num_groups,),
                        )
                    )
                inital_num_groups = current_num_groups
                while current_num_groups != 1:
                    c_step.ahc()
                    c_step.emc()
                    c_step.calc_bic()
                    c_step.group_2_levels()
                    # print([lvl.int_p_s for lvl in c_step.group_levels])
                    steps.append(c_step)
                    current_num_groups = c_step.num_groups
                    if current_num_groups != 1:
                        # print(current_num_groups)
                        c_step = c_step.setup_next_step()
                        if feedback_queue is not None:
                            feedback_queue.put(
                                ProcessProgressTask(
                                    task_cmd=ProcessProgressCmd.SetValue,
                                    args=(inital_num_groups - current_num_groups,),
                                )
                            )

                if feedback_queue is not None:
                    feedback_queue.put(
                        ProcessProgressTask(task_cmd=ProcessProgressCmd.Complete)
                    )
                self.steps = steps
                self.num_steps = len(steps)
                self.bics = [step.bic for step in steps]
```

```python
                self.best_step_ind = np.argmax(self.bics)
                self.selected_step_ind = self.best_step_ind
                self.has_groups = True
                self.grouped_with_roi = self.use_roi_for_grouping
                if self.backup is not None:
                    self.backup = None
                    self.plots_need_to_be_updated = True
                logger.info(f"{self._particle.name} levels grouped")
        else:
            logger.info(f"{self._particle.name} has no levels to group")
    except Exception as e:
        self.restore_and_delete_backup()
        logger.error(e)
        pass
```

A.6 Agglomerative Hierarchical Clustering

```python
def ahc(self):
    """Agglomerative Hierarchical Clustering"""

    # Calculate merge merit value for each possible merging situation
    merge_merit = np.full(
        shape=(self._num_prev_groups, self._num_prev_groups), fill_value=-np.inf
    )
    for j, group_j in enumerate(self._seed_groups):  # Row
        for m, group_m in enumerate(self._seed_groups):  # Column
            if j < m:
                n_m = group_m.num_photons  # Number of photons in group m
                n_j = group_j.num_photons  # Number of photons in group j
                t_m = group_m.dwell_time_s  # Dwell time of group m in seconds
                t_j = group_j.dwell_time_s  # Dwell time of group j in seconds

                merge_merit[
                    j, m
                ] = (  # Log-Likelihood Merge merit function for joining groups j and m, eq. 11
                    (n_m + n_j)
                    * np.log(
                        (n_m + n_j) / (t_m + t_j)
                    )  # (n_m - n_j)\ln[\frac{n_m + n_j}{T_m + T_j}]
                    - n_m * np.log(n_m / t_m)  # - n_m\ln[\frac{n_m}{T_m}]
                    - n_j * np.log(n_j / t_j)  # - n_j\ln[\frac{n_j}{T_j}]
                )

    max_j, max_m = max_jm(merge_merit)  # Find most probably merge

    # Perform merging of groups max_j and max_m
    new_groups = []
    p_mj = np.zeros(shape=(self._num_prev_groups - 1, self._num_levels))
    group_num = -1
    for i, group_i in enumerate(self._seed_groups):
        if i != max_m:
            group_num += 1
            if i == max_j:  # Then merge
                merge_levels = group_i.lvls_inds.copy()
                merge_levels.extend(self._seed_groups[max_m].lvls_inds.copy())
                merge_levels.sort()
                # New group made up of all the levels in groups max_j and max_m
                new_groups.append(
                    Group(lvls_inds=merge_levels, particle=self._particle, group_ind=group_num)
```

```python
                )
            else:
                new_groups.append(group_i)

            # Set up new assignment matrix, to be as initial start for EM clustering
            for ind in new_groups[-1].lvls_inds:
                p_mj[group_num, ind] = 1

    self._ahc_p_mj = p_mj
    self._ahc_groups = new_groups
```

A.7 Expectation Maximisation

```python
def emc(self):
    """Expectation Maximisation clustering"""

    p_mj = self._ahc_p_mj.copy()  # Initial state, as provided by AHC

    prev_p_mj = p_mj.copy()
    if not self._particle.use_roi_for_grouping:
        levels = self._particle.cpts.levels
    else:
        levels = self._particle.levels_roi

    # M-Step
    ######################################################################
    i = 0
    diff_p_mj = 1
    p_hat_g = None
    while diff_p_mj > 1e-5 and i < 50:
        i += 1
        if not self._particle.use_roi_for_grouping:
            cap_t = self._particle.dwell_time_s
        else:
            cap_t = self._particle.dwell_time_roi

        t_hat = np.zeros(shape=(self._num_prev_groups - 1,))
        n_hat = np.zeros_like(t_hat)
        p_hat = np.zeros_like(t_hat)
        i_hat = np.zeros_like(t_hat)

        p_hat_g = np.zeros(shape=(self._num_prev_groups - 1, self._num_levels))
        denom = np.zeros(shape=(self._num_levels,))

        for m, group in enumerate(self._ahc_groups):  # As in Fig. 9
            # \hat{T}_m = \sum\limits_{j=1}^{J+1}\bar{p}_{mj}T_j
            t_hat[m] = np.sum(
                [p_mj[m, j] * l.dwell_time_s for j, l in enumerate(levels)]
            )

            # \hat{n}_m = \sum\limits_{j=1}^{J+1}\bar{p}_{mj}n_j
            n_hat[m] = np.sum(
                [p_mj[m, j] * l.num_photons for j, l in enumerate(levels)]
            )

            # \hat{p}_m = \frac{\hat{T}_m}{T}
            p_hat[m] = t_hat[m] / cap_t

            # \hat{I}_m = \sum\limits_{j=1}^{J+1}\frac{\bar{p}_{mj}n_j}{\hat{T}_m}
            i_hat[m] = np.sum(
                [
```

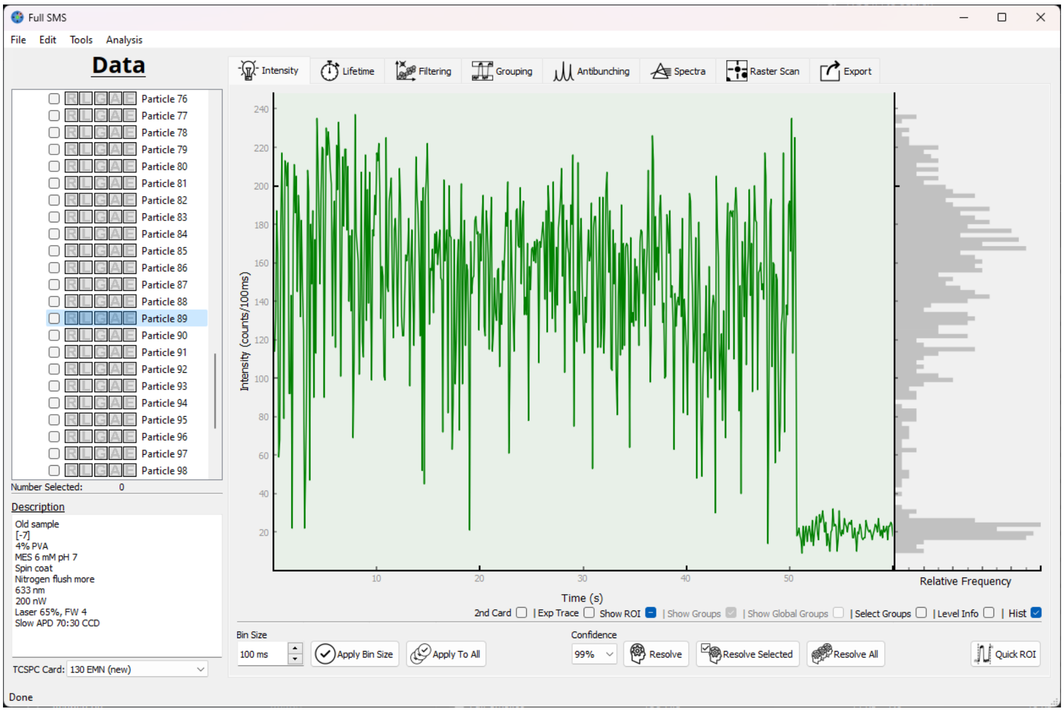
```
                    p_mj[m, j] * l.num_photons / t_hat[m]
                    for j, l in enumerate(levels)
                ]
            )

            # Let \bar{p}_{mj} = \frac{\alpha_{mj}}{\sum_{m=1}^G\alpha_{mj}}
            # where \alpha_{mj} = \hat{p}_mg(n_j;\hat{I}_m,T_j)
            for j, l in enumerate(levels):
                p_hat_g[m, j] = p_hat[m] * poisson.pmf(
                    # l.num_photons, i_hat[m] * l.dwell_time_s
                    n_hat[m], i_hat[m] * l.dwell_time_s
                )  # \alpha_{mj}

        # \sum_{m=1}^G\alpha_{mj} = \sum_{m=1}^G\hat{p}_mg(n_j;\hat{I}_m,T_j)
        for j in range(self._num_levels):
            denom[j] = np.sum(p_hat_g[:, j])

            for m in range(self._num_prev_groups - 1):
                try:
                    p_mj[m, j] = (
                        p_hat_g[m, j] / denom[j]
                    )  # \bar{p}_{mj}=\frac{\alpha_{mj}}{\sum_{m=1}^G\alpha_{mj}}
                except:
                    # print('here')  # TODO: Fix div by zero
                    pass

        diff_p_mj = np.sum(np.abs(prev_p_mj - p_mj))
        prev_p_mj = p_mj.copy()

    level_p_max = np.argmax(p_mj, 0)
    eff_p_mj = np.zeros_like(p_mj)
    for j in range(self._num_levels):
        eff_p_mj[level_p_max[j], j] = 1
    self._em_p_mj = eff_p_mj

    # E-Step
    ######################################################################
    log_l = 0
    for m in range(self._num_prev_groups - 1):
        for j in range(self._num_levels):
            if (
                p_hat_g[m, j] > 1e-200 and eff_p_mj[m, j] != 0
            ):  # Close to smallest value that doesn't result in -inf
                log_l += eff_p_mj[m, j] * np.log(p_hat_g[m, j])
    self._em_log_l = log_l

    new_groups = []
    for m in range(self._num_prev_groups - 1):
        g_m_levels = list(np.nonzero(self._em_p_mj[m, :])[0])
        if len(g_m_levels):
            new_groups.append(Group(lvls_inds=g_m_levels, particle=self._particle, group_ind=m))
    new_groups.sort(key=lambda group: group.int_p_s)
    self.groups = new_groups
    self.num_groups = len(new_groups)
    if self.num_groups == 1:
        self.last = True

    level_group_ind = [None] * self._num_levels
    for group_num, group in enumerate(self.groups):
        for level in group.lvls_inds:
            level_group_ind[level] = group_num
    self.level_group_ind = level_group_ind
```

A.8 BIC

```python
def calc_bic(self):
    num_cp = self._particle.cpts.num_cpts
    num_g = self.num_groups

    self.bic = (
        2 * self._em_log_l
        - (2 * num_g - 1) * np.log(num_cp)
        - num_cp * np.log(self._particle.num_photons)
    )
```

Appendix B

Graphical User Interface



Number Selected: 0

Description

Old sample
[-7]
4% PVA
MES 6 mM pH 7
Spin coat
Nitrogen flush more
633 nm
200 nW
Laser 65%, FW 4
Slow APD 70:30 CCD

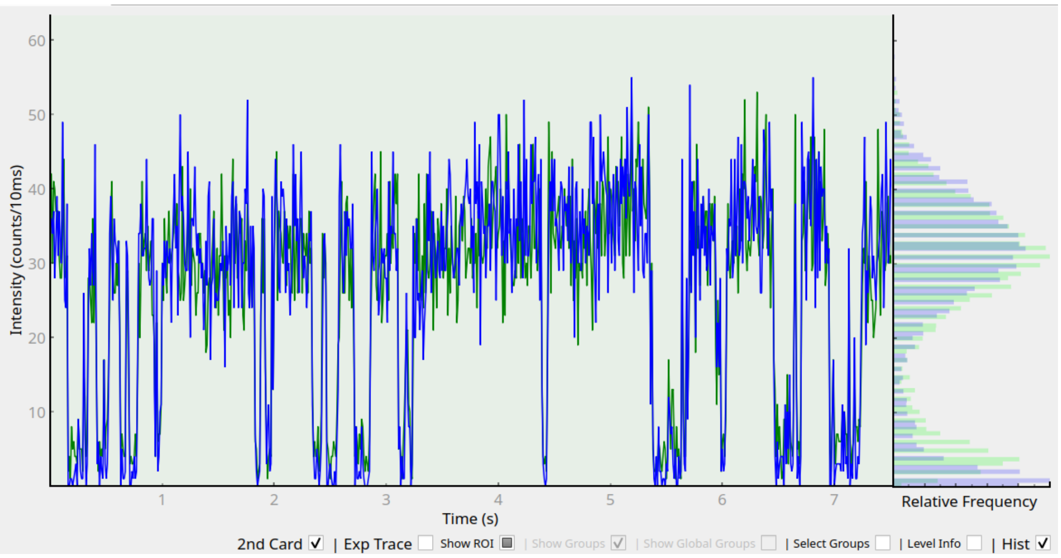
TCSPC Card: 130 EMN (new)

2nd Card ☐ | Exp Trace ☐ Show ROI ☑ | Show Groups ☑ | Show Global Groups ☐ | Select Groups ☐ | Level Info ☐ | Hist ☑

Bin Size: 100 ms

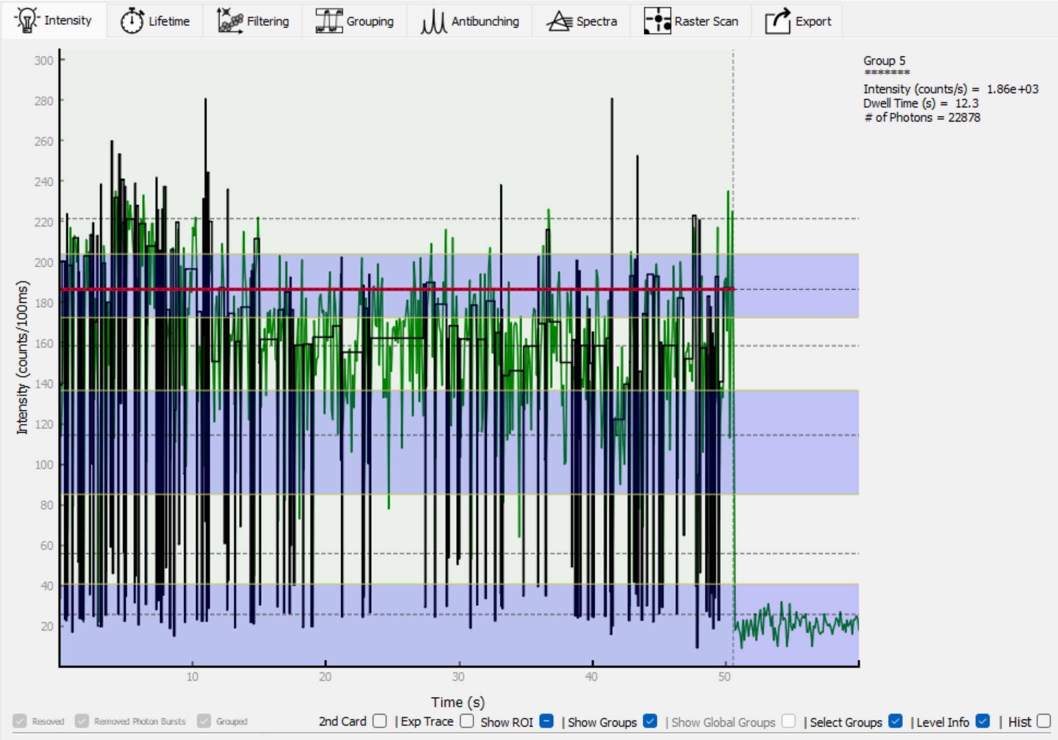
Confidence: 99%

FIGURE B.1: Intensity trace example of Alexa as shown in the *Intensity* tab, showing a binned intensity trace before the resolving of the intensity levels is actioned. The same trace is shown in Fig. 4.3 with the levels resolved.



2nd Card ☑ | Exp Trace ☐ Show ROI ■ | Show Groups ☑ | Show Global Groups ☐ | Select Groups ☐ | Level Info ☐ | Hist ☑

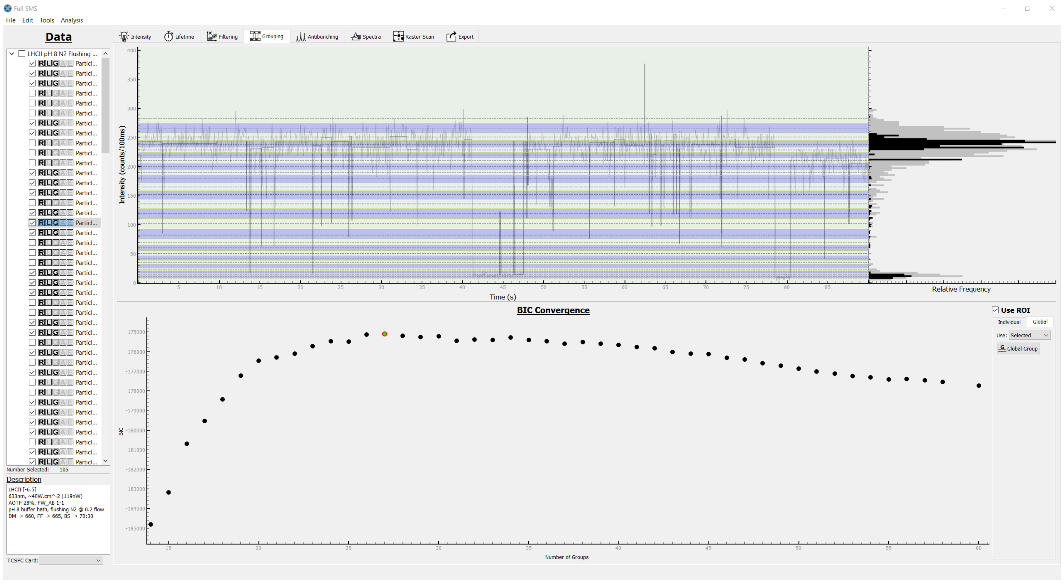
FIGURE B.2: Two-channel intensity trace example, showing the binned intensity traces from each channel as a separate colour.



Group 5
****
Intensity (counts/s) = 1.86e+03
Dwell Time (s) = 12.3
# of Photons = 22878

☑ Resolved ☑ Removed Photon Bursts ☑ Grouped 2nd Card ☐ | Exp Trace ☐ Show ROI ■ | Show Groups ☑ | Show Global Groups ☐ | Select Groups ☑ | Level Info ☑ | Hist ☐

FIGURE B.3: Group selection example using the groups shown in Fig. 4.7. The group's intensity, dwell time, and number of photons are displayed.

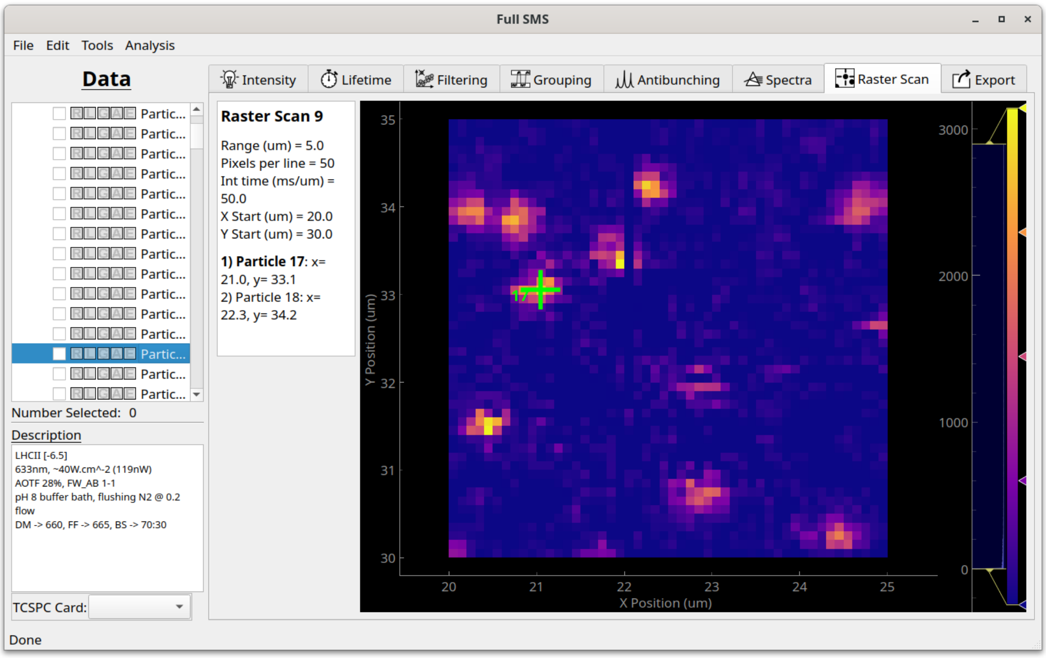


Number Selected: 105

Description

LHCII [-6.5]
633nm, ~40W.cm^-2 (119nW)
AOTF 28%, FW_AB 1-1
pH 8 buffer bath, flushing N2 @ 0.2 flow
DM -> 660, FF -> 665, BS -> 70:30

FIGURE B.4: Global grouping example of the data of 105 LHCII complexes. The bottom pane shows the progression of the grouping of the appended data, which originally comprised $\sim 10^5$ individual intensity levels. The largest BIC value corresponds to 27 distinct states (down from $\sim 10^5$ states), some of which are not accessed in the intensity trace in the upper pane.



Number Selected: 0

Description

LHCII [-6.5]
633nm, ~40W.cm^-2 (119nW)
AOTF 28%, FW_AB 1-1
pH 8 buffer bath, flushing N2 @ 0.2 flow
DM -> 660, FF -> 665, BS -> 70:30

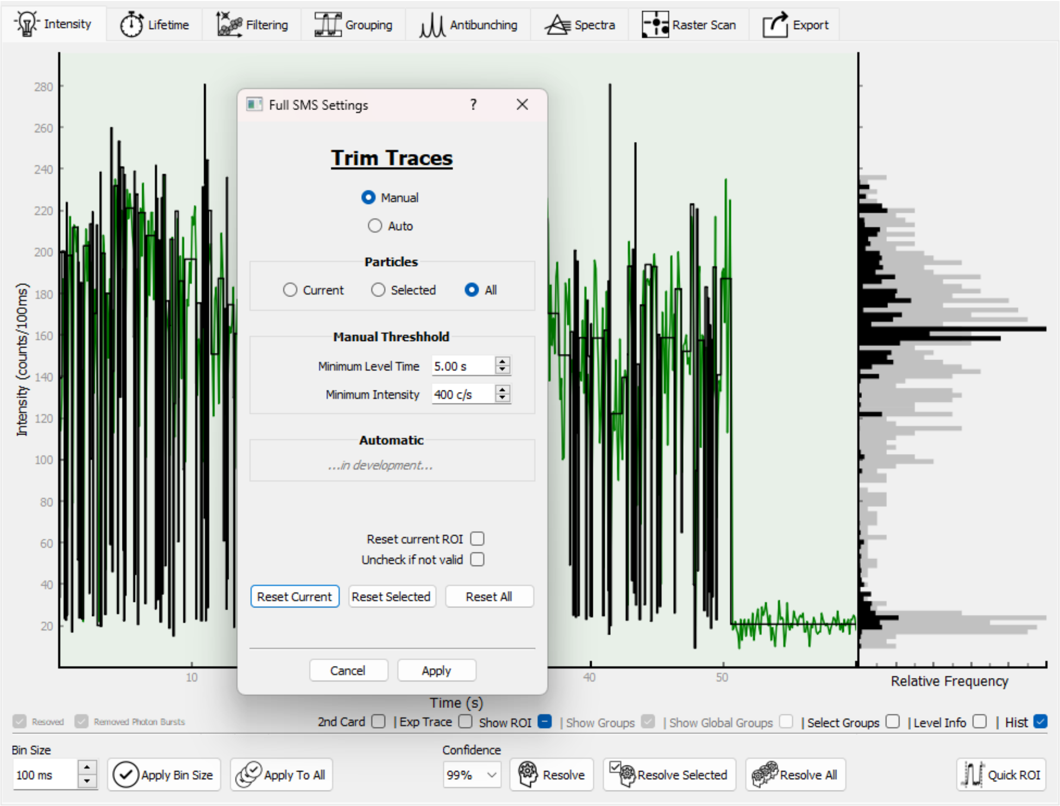
Raster Scan 9

Range (um) = 5.0
Pixels per line = 50
Int time (ms/um) = 50.0
X Start (um) = 20.0
Y Start (um) = 30.0

1) Particle 17: x= 21.0, y= 33.1
2) Particle 18: x= 22.3, y= 34.2

Done

FIGURE B.5: Example of a raster-scan image of a few LHCII complexes immobilised on a glass coverslip via poly-L-lysine, indicating the position of the currently selected particle with a green plus. The raster-scan details and coordinates of selected particles are indicated on the left of the image.



Trim Traces

● Manual
○ Auto

Particles

○ Current ○ Selected ● All

Manual Threshhold

Minimum Level Time 5.00 s
Minimum Intensity 400 c/s

Automatic

...in development...

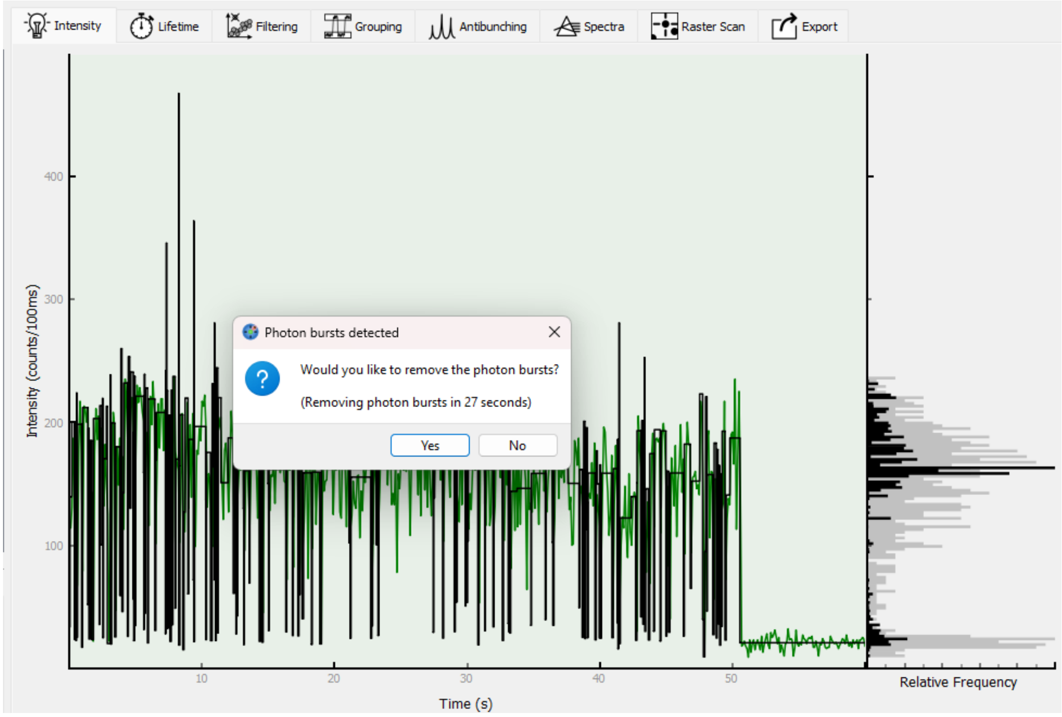
Reset current ROI ☐
Uncheck if not valid ☐

2nd Card ☐ | Exp Trace ☐ Show ROI ☑ | Show Groups ☑ | Show Global Groups ☐ | Select Groups ☐ | Level Info ☐ | Hist ☑

Bin Size 100 ms

Confidence 99%

FIGURE B.6: Trace trimming tool dialog with the Alexa intensity trace shown in Fig. 4.3. The end of a trace is trimmed if it is below the given minimum intensity and longer than the given minimum dwell time. An automated choice of these parameters is a future extension that is currently under development.



Photon bursts detected

Would you like to remove the photon bursts?

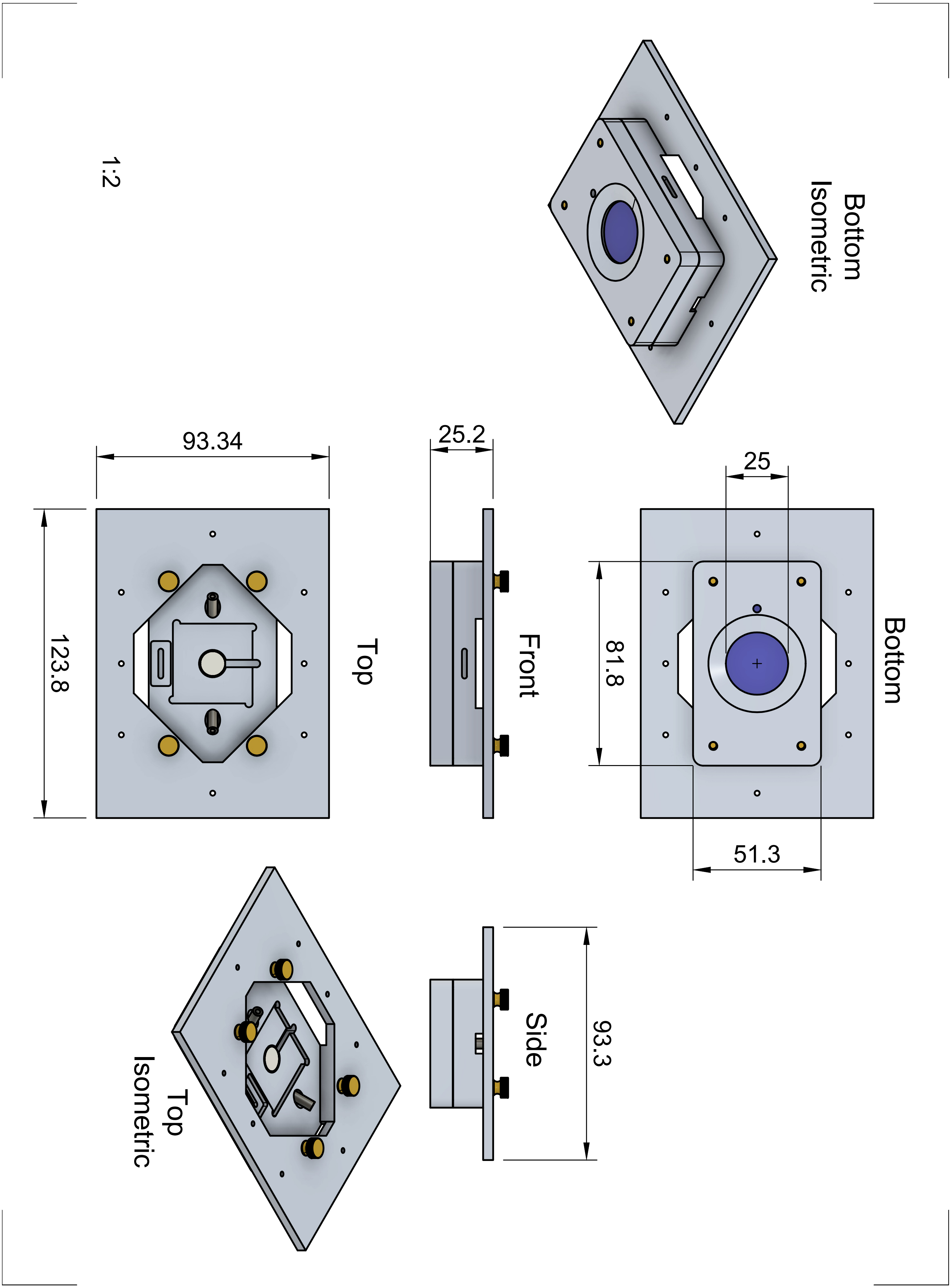
(Removing photon bursts in 27 seconds)

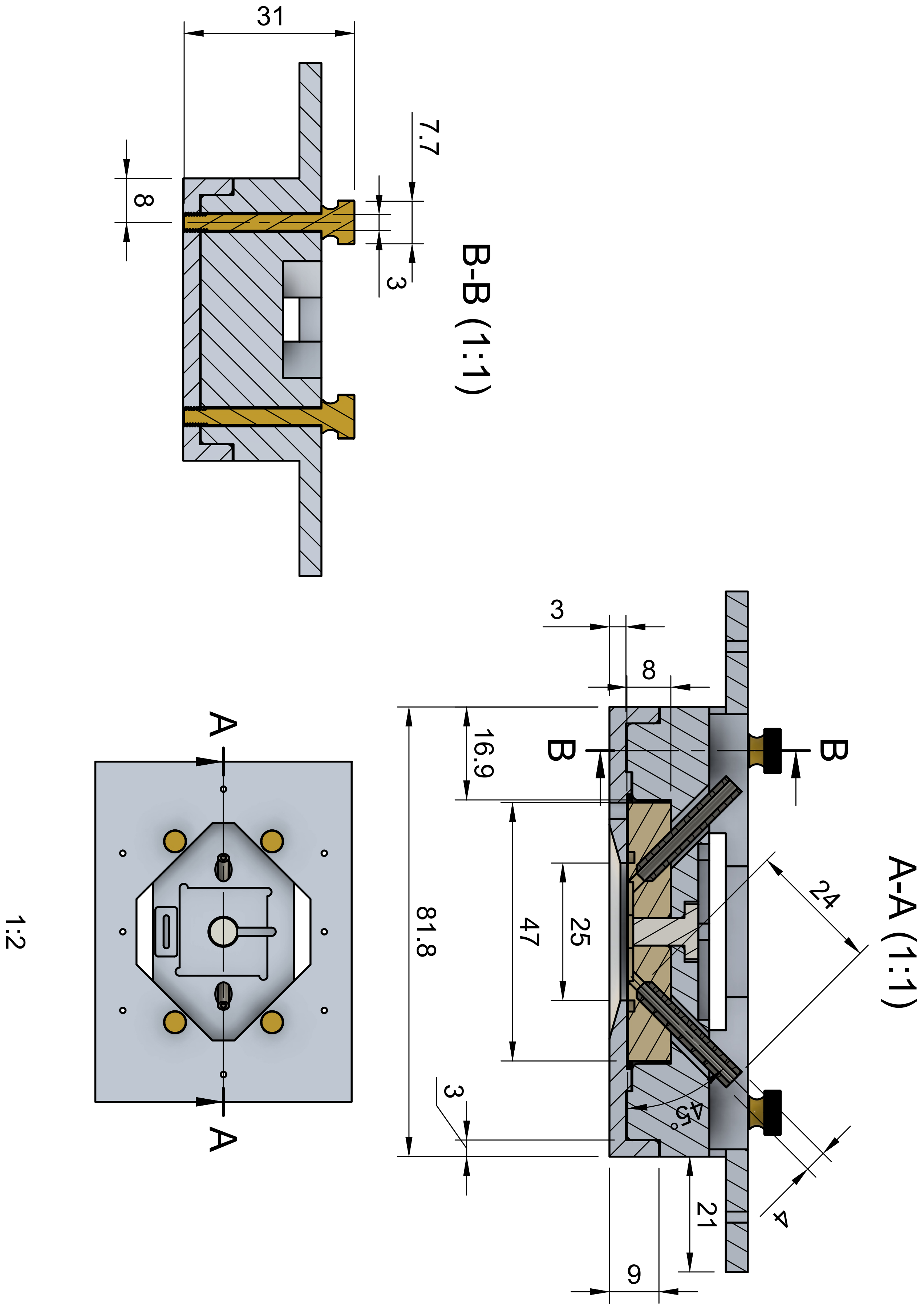
FIGURE B.7: photon burst dialog example with the Alexa intensity trace shown in Fig. 4.3. This dialog is presented automatically after resolving levels, if photon bursts are detected.

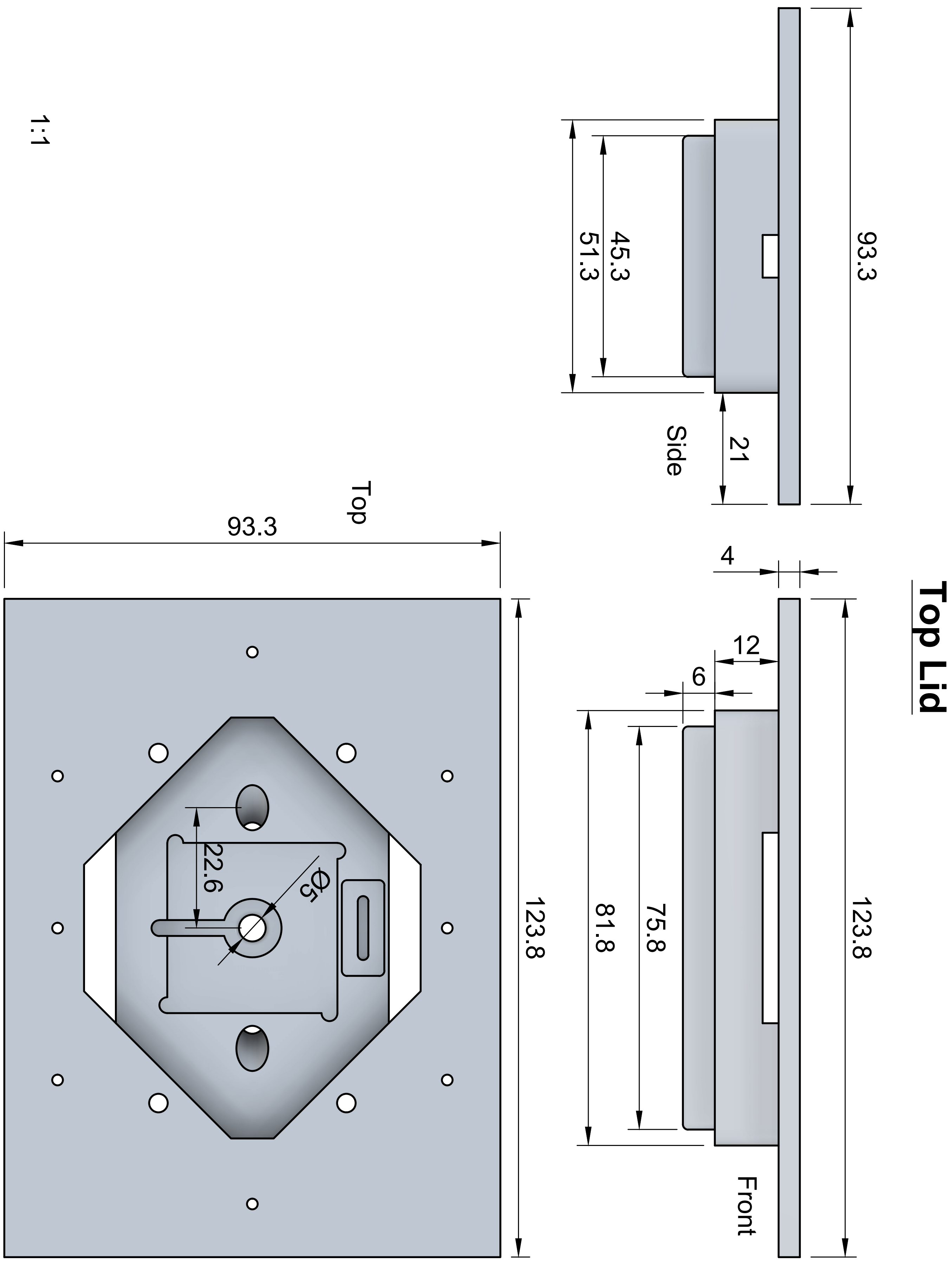
Appendix C

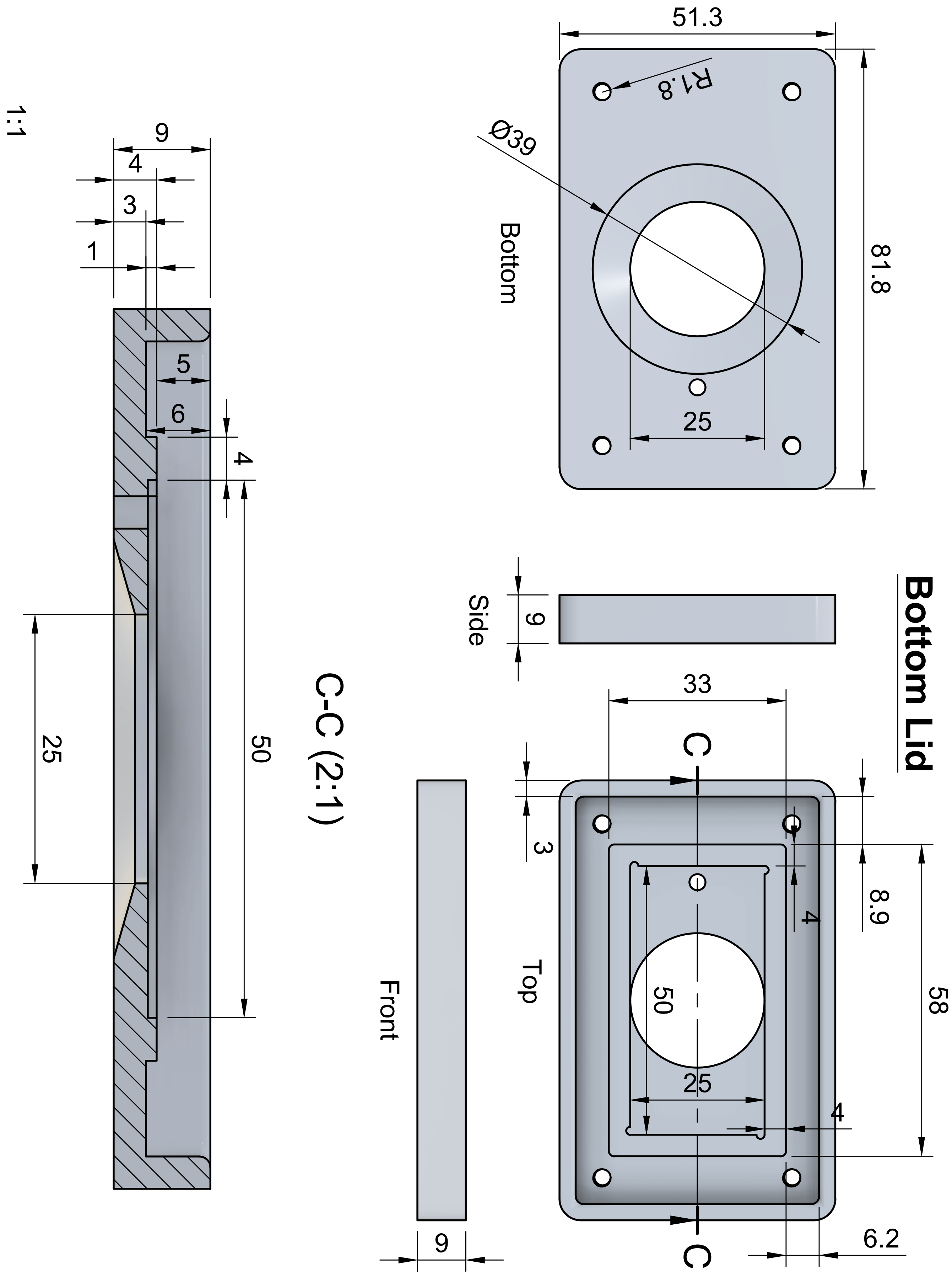
Sample Cell Holder

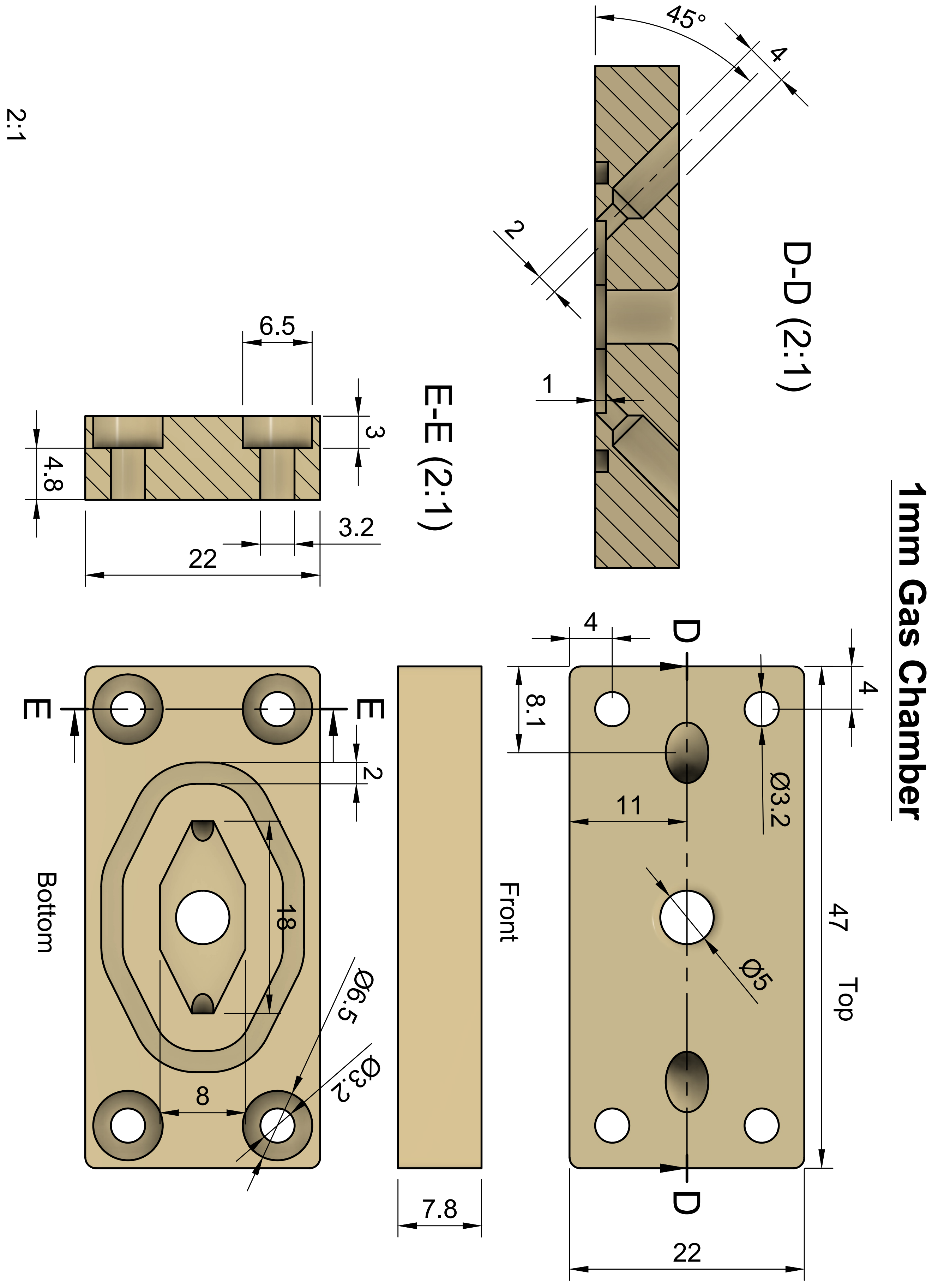
The pages that follow contain CAD drawings the Sample Cell Holder design and show some of the important dimensions. All the units are in mm, and the page scale is indicated on each page.


1:2
Bottom Isometric
93.34
25.2
25
123.8
Top
Front
81.8
Bottom
51.3
Top Isometric
Side
93.3


31
8
7.7
3
B-B (1:1)
3
8
16.9
B
B
81.8
47
25
24
A-A (1:1)
3
45°
4
21
9
A
A
1:2


Top Lid
1:1
Top
Side
Front
93.3
93.3
123.8
123.8
45.3
51.3
21
4
12
6
81.8
75.8
22.6
Ø5


Bottom Lid
1:1
Bottom
Side
Top
Front
C-C (2:1)
C
C
51.3
81.8
R1.8
Ø39
25
9
33
3
8.9
4
58
50
25
4
6.2
9
9
4
3
1
5
6
4
50
25


1mm Gas Chamber
2:1
D-D (2:1)
45°
4
2
1
E-E (2:1)
6.5
3
4.8
3.2
22
D
D
4
8.1
4
Ø3.2
11
Ø5
47
Top
22
Front
7.8
E
E
2
18
Ø6.5
Ø3.2
8
Bottom

Bibliography

[1] R. Hooke 1635-1703, *Micrographia: Or Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses. With Observations and Inquiries Thereupon.* London: J. Martyn, and J. Allestry, 1665.

[2] C. Huygens, *Traité de La Lumière*. A Leide: Chez Pierre vander Aa, marchand libraire, 1690. DOI: 10.5479/sil.294285.39088000545160.

[3] J. C. Maxwell, "VIII. A dynamical theory of the electromagnetic field," *Philos. Trans. R. Soc. Lond.*, vol. 155, pp. 459–512, 1865. DOI: 10.1098/rstl.1865.0008.

[4] G. Kirchhoff and R. Bunsen, "Chemical analysis by observation of spectra," *Ann. Phys. Chem.*, vol. 110, pp. 161–189, 1860.

[5] A. Einstein, *Zur Quantentheorie Der Strahlung*. Verlag u. Druck Gebr. Leemann, 1916.

[6] T. H. Maiman, "Stimulated Optical Radiation in Ruby," *Nature*, vol. 187m, no. 4736, pp. 493–494, Aug. 1960. DOI: 10.1038/187493a0.

[7] W. E. Moerner and L. Kador, "Optical detection and spectroscopy of single molecules in a solid," *Phys. Rev. Lett.*, vol. 62, no. 21, pp. 2535–2538, May 1989. DOI: 10.1103/PhysRevLett.62.2535.

[8] M. Orrit and J. Bernard, "Single pentacene molecules detected by fluorescence excitation in a p-terphenyl crystal," *Phys. Rev. Lett.*, vol. 65, no. 21, pp. 2716–2719, Nov. 1990. DOI: 10.1103/PhysRevLett.65.2716.

[9] E. Brooks Shera et al., "Detection of single fluorescent molecules," *Chem. Phys. Lett.*, vol. 174, no. 6, pp. 553–557, 1990. DOI: 10.1016/0009-2614(90)85485-U.

[10] S. A. Soper et al., "Single-molecule detection of Rhodamine 6G in ethanolic solutions using continuous wave laser excitation," *Anal. Chem.*, vol. 63, no. 5, pp. 432–437, Mar. 1991. DOI: 10.1021/ac00005a009.

[11] S. A. Soper, L. M. Davis, and E. B. Shera, "Detection and identification of single molecules in solution," *J. Opt. Soc. Am. B*, vol. 9, no. 10, pp. 1761–1769, Oct. 1992. DOI: 10.1364/JOSAB.9.001761.

[12] D. C. Nguyen, R. A. Keller, J. H. Jett, and J. C. Martin, "Detection of single molecules of phycoerythrin in hydrodynamically focused flows by laser-induced fluorescence," *Anal. Chem.*, vol. 59, no. 17, pp. 2158–2161, Sep. 1987. DOI: 10.1021/ac00144a032.

[13] W. E. Moerner, "A dozen years of single-molecule spectroscopy in physics, chemistry, and biophysics," *J. Phys. Chem. B*, vol. 106, no. 5, pp. 910–927, 2002. DOI: 10.1021/jp012992g.

[14] B. Van Heerden, N. A. Vickers, T. P. Krüger, and S. B. Andersson, "Real-time feedback-driven single-particle tracking: A survey and perspective," *Small*, vol. 18, no. 29, p. 2 107 024, 2022.

[15] R. Roy, S. Hohng, and T. Ha, "A practical guide to single-molecule FRET," *Nat. Methods*, vol. 5, no. 6, pp. 507–516, 2008.

[16] H. Mazal and G. Haran, "Single-molecule FRET methods to study the dynamics of proteins at work," *Curr. Opin. Biomed. Eng.*, vol. 12, pp. 8–17, Dec. 2019. DOI: `10.1016/j.cobme.2019.08.007`.

[17] T. Ha et al., "Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism," *Proc. Natl. Acad. Sci. USA*, vol. 96, no. 3, pp. 893–898, Feb. 1999. DOI: `10.1073/pnas.96.3.893`.

[18] Y. Jiang et al., "Sensing cooperativity in ATP hydrolysis for single multisubunit enzymes in solution," *Proc. Natl. Acad. Sci. USA*, vol. 108, no. 41, pp. 16 962–16 967, Sep. 2011. DOI: `10.1073/pnas.1112244108`.

[19] A. N. Kapanidis et al., "Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism," *Science*, vol. 314, no. 5802, pp. 1144–1147, Nov. 2006. DOI: `10.1126/science.1131399`.

[20] S. Hou, J. Exell, and K. Welsher, "Real-time 3D single molecule tracking," *Nat. Commun.*, vol. 11, no. 1, p. 3607, 1 2020. DOI: `10.1038/s41467-020-17444-6`.

[21] T. P. J. Krüger et al., "How reduced excitonic coupling enhances light harvesting in the main photosynthetic antennae of diatoms," *Proc. Natl. Acad. Sci. USA*, vol. 114, no. 52, E11063–E11071, Dec. 2017. DOI: `10.1073/pnas.1714656115`.

[22] T. Kondo, W. J. Chen, and G. S. Schlau-Cohen, "Single-molecule fluorescence spectroscopy of photosynthetic systems," *Chem. Rev.*, vol. 117, no. 2, pp. 860–898, 2017. DOI: `10.1021/acs.chemrev.6b00195`.

[23] M. Gwizdala et al., "Switching an individual phycobilisome off and on," *J. Phys. Chem. Lett.*, vol. 9, no. 9, pp. 2426–2432, 2018. DOI: `10.1021/acs.jpclett.8b00767`.

[24] J. M. Gruber, P. Malý, T. P. J. Krüger, and R. van Grondelle, "From isolated light-harvesting complexes to the thylakoid membrane: A single-molecule perspective," *Nanophotonics*, vol. 7, no. 1, pp. 81–92, 2018. DOI: `10.1515/nanoph-2017-0014`.

[25] H. H. Tuson and J. S. Biteen, "Unveiling the inner workings of live bacteria using super-resolution microscopy," *Anal. Chem.*, vol. 87, no. 1, pp. 42–63, Jan. 2015. DOI: `10.1021/ac5041346`.

[26] K. Welsher and H. Yang, "Multi-resolution 3D visualization of the early stages of cellular uptake of peptide-coated nanoparticles," *Nat. Nanotechnol.*, vol. 9, no. 3, pp. 198–203, Mar. 2014. DOI: `10.1038/nnano.2014.12`.

[27] B. van Heerden and T. P. J. Krüger, "Theoretical comparison of real-time feedback-driven single-particle tracking techniques," *J. Chem. Phys.*, vol. 157, no. 8, p. 084 111, Aug. 2022. DOI: `10.1063/5.0096729`.

[28] F. P. García de Arquer et al., "Semiconductor quantum dots: Technological progress and future challenges," *Science*, vol. 373, no. 6555, eaaz8541, Aug. 2021. DOI: `10.1126/science.aaz8541`.

[29] C. Toninelli et al., "Single organic molecules for photonic quantum technologies," *Nat. Mater.*, vol. 20, no. 12, pp. 1615–1628, Dec. 2021. DOI: `10.1038/s41563-021-00987-4`.

[30] T. Zhang et al., "Toward quantitative bio-sensing with nitrogen–vacancy center in diamond," *ACS Sens.*, vol. 6, no. 6, pp. 2077–2107, Jun. 2021. DOI: `10.1021/acssensors.1c00415`.

[31] G. Lubin et al., "Photon correlations in spectroscopy and microscopy," *ACS Photonics*, vol. 9, no. 9, pp. 2891–2904, Sep. 2022. DOI: `10.1021/acsphotonics.2c00817`.

[32] M. Heilemann et al., "Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes," *Angew. Chem. Int. Ed.*, vol. 47, no. 33, pp. 6172–6176, 2008. DOI: `10.1002/anie.200802376`.

[33] S. van de Linde, S. Wolter, and M. Sauer, "Single-molecule photoswitching and localization," *Aust. J. Chem.*, vol. 64, no. 5, pp. 503–511, May 2011. DOI: `10.1071/CH10284`.

[34] T. P. J. Krüger et al., "Controlled disorder in plant light-harvesting complex II explains its photoprotective role," *Biophys. J.*, vol. 102, no. 11, pp. 2669–2676, 2012. DOI: `10.1016/j.bpj.2012.04.044`.

[35] M. Gwizdala et al., "Controlling light harvesting with light," *J. Am. Chem. Soc.*, vol. 138, no. 36, pp. 11 616–11 622, 2016. DOI: `10.1021/jacs.6b04811`.

[36] L. Mustárdy and G. Garab, "Granum revisited. A three-dimensional model - Where things fall into place," *Trends Plant Sci.*, vol. 8, no. 3, pp. 117–122, 2003. DOI: `10.1016/S1360-1385(03)00015-3`.

[37] C. Laloi and M. Havaux, "Key players of singlet oxygen-induced cell death in plants," *Front. Plant Sci.*, vol. 6, p. 39, February 2015. DOI: `10.3389/fpls.2015.00039`.

[38] P. Horton, A. V. Ruban, and R. G. Walters, "Regulation of Light Harvesting in Green Plants," *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, vol. 47, no. 1, pp. 655–684, 1996. DOI: `10.1146/annurev.arplant.47.1.655`.

[39] Z. Liu et al., "Crystal structure of spinach major light-harvesting complex at 2.72 A resolution," *Nature*, vol. 428, no. 6980, pp. 287–292, 2004. DOI: `10.5940/jcrsj.46.Supplement_19`.

[40] V. I. Novoderezhkin, M. A. Palacios, H. Van Amerongen, and R. Van Grondelle, "Excitation dynamics in the LHCII complex of higher plants: Modeling based on the 2.72 Å crystal structure," *J. Phys. Chem. B*, vol. 109, no. 20, pp. 10 493–10 504, 2005. DOI: `10.1021/jp044082f`.

[41] R. van Grondelle and V. I. Novoderezhkin, "Energy transfer in photosynthesis: Experimental insights and quantitative models," *Phys. Chem. Chem. Phys.*, vol. 8, no. 7, pp. 793–807, 2006. DOI: `10.1039/b514032c`.

[42] T. Basché, S. Kummer, and C. Bräuchle, "Direct spectroscopic observation of quantum jumps of a single molecule," *Nature*, vol. 373, no. 6510, pp. 132–134, 1995. DOI: `10.1038/373132a0`.

[43] M. Nirmal et al., "Fluorescence intermittency in single cadmium selenide nanocrystals," *Nature*, vol. 383, no. 6603, pp. 802–804, 1996. DOI: `10.1038/383802a0`.

[44] A. A. Cordones and S. R. Leone, "Mechanisms for charge trapping in single semiconductor nanocrystals probed by fluorescence blinking," *Chemical Society Reviews*, vol. 42, no. 8, pp. 3209–3221, 2013. DOI: `10.1039/C2CS35452G`.

[45] K. Kawai, M. Fujitsuka, and A. Maruyama, "Single-molecule study of redox reaction kinetics by observing fluorescence blinking," *Acc. Chem. Res.*, vol. 54, no. 4, pp. 1001–1010, 2021. DOI: `10.1021/acs.accounts.0c00754`.

[46] J. Tang and R. A. Marcus, "Mechanisms of fluorescence blinking in semiconductor nanocrystal quantum dots," *J. Chem. Phys.*, vol. 123, no. 5, p. 054 704, Aug. 2005. DOI: `10.1063/1.1993567`.

[47] G. S. Schlau-Cohen et al., "Single-Molecule Identification of Quenched and Unquenched States of LHCII," *J. Phys. Chem. Lett.*, vol. 6, no. 5, pp. 860–867, Mar. 2015. DOI: `10.1021/acs.jpclett.5b00034`.

[48] R. M. Dickson, A. B. Cubitt, R. Y. Tsien, and W. E. Moerner, "On/off blinking and switching behaviour of single molecules of green fluorescent protein," *Nature*, vol. 388, no. 6640, pp. 355–358, 1997. DOI: `10.1038/41048`.

[49] E. Abbe, "Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung," *Arch. Für Mikrosk. Anat.*, vol. 9, no. 1, pp. 413–468, 1873.

[50] P. Xu, L. Tian, M. Kloz, and R. Croce, "Molecular insights into Zeaxanthin-dependent quenching in higher plants," *Sci. Rep.*, 2015. DOI: `10.1038/srep13679`.

[51] J. L. Botha, B. van Heerden, and T. P. Krüger, "Advanced analysis of single-molecule spectroscopic data," *Biophys. Rep.*, vol. 4, no. 3, Sep. 2024. DOI: `10.1016/j.bpr.2024.100173`.

[52] M. C. A. S. Hadzic et al., "Reliable State Identification and State Transition Detection in Fluorescence Intensity-Based Single-Molecule Förster Resonance Energy-Transfer Data," *J. Phys. Chem. B*, vol. 122, no. 23, pp. 6134–6147, Jun. 2018. DOI: `10.1021/acs.jpcb.7b12483`.

[53] S. Schmid, M. Götz, and T. Hugel, "Single-Molecule Analysis beyond Dwell Times: Demonstration and Assessment in and out of Equilibrium," *Biophys. J.*, vol. 111, no. 7, pp. 1375–1384, Oct. 2016. DOI: `10.1016/j.bpj.2016.08.023`.

[54] S. Pressé et al., "Single Molecule Conformational Memory Extraction: P5ab RNA Hairpin," *J. Phys. Chem. B*, vol. 118, no. 24, pp. 6597–6603, Jun. 2014. DOI: `10.1021/jp500611f`.

[55] L. P. Watkins and H. Yang, "Detection of intensity change Points in time-resolved single-molecule measurements," *J. Phys. Chem. B*, vol. 109, no. 1, pp. 617–628, Jan. 2005. DOI: `10.1021/jp0467548`.

[56] T. P. J. Krüger, C. Ilioaia, L. Valkunas, and R. Van Grondelle, "Fluorescence intermittency from the main plant light-harvesting complex: Sensitivity to the local environment," *J. Phys. Chem. B*, vol. 115, no. 18, pp. 5083–5095, 2011. DOI: `10.1021/jp109833x`.

[57] R. Henderson, "A problem with the likelihood ratio test for a change-point hazard rate model," *Biometrika*, vol. 77, no. 4, pp. 835–843, 1990.

[58] M. Noé, "The calculation of distributions of two-sided Kolmogorov-Smirnov type statistics," *Ann. Math. Stat.*, pp. 58–64, 1972.

[59] C. A. Seidel. "Universität Düsseldorf: Software Package for MFD, FCS and MFIS," Software | Package for MFD, FCS and MFIS. (), [Online]. Available: `https://www.mpc.hhu.de/en/software/mfd-fcs-and-mfis` (visited on 03/05/2024).

[60] W. Schrimpf, A. Barth, J. Hendrix, and D. C. Lamb, "PAM: A framework for integrated analysis of imaging, single-molecule, and ensemble fluorescence data," *Biophys. J.*, vol. 114, no. 7, pp. 1518–1528, Apr. 2018. DOI: `10.1016/j.bpj.2018.02.035`.

[61] E. Gratton. "Globals Software - Laboratory for Fluorescence Dynamics." (), [Online]. Available: `https://www.lfd.uci.edu/globals/` (visited on 06/06/2024).

[62] S. Preus et al., "iSMS: Single-molecule FRET microscopy software," *Nat Methods*, vol. 12, no. 7, pp. 593–594, Jul. 2015. DOI: `10.1038/nmeth.3435`.

[63] I. M. Palstra and A. F. Koenderink, "A Python toolbox for unbiased statistical analysis of fluorescence intermittency of multilevel emitters," *J. Phys. Chem. C*, vol. 125, no. 22, 2021. DOI: `10.1021/acs.jpcc.1c01670`.

[64] A. Ingargiola et al., "Photon-HDF5: An open file format for timestamp-based single-molecule fluorescence experiments," *Biophys. J.*, vol. 110, no. 1, pp. 26–33, Jan. 2016. DOI: `10.1016/j.bpj.2015.11.013`.

[65] J. J. Snellenburg et al., "Glotaran: A Java-based graphical user interface for the R package TIMP," *J. Stat. Softw.*, vol. 49, pp. 1–22, Jun. 2012. DOI: `10.18637/JSS.V049.I03`.

[66] I. H. M. van Stokkum, J. Weißenborn, S. Weigand, and J. J. Snellenburg, "Pyglotaran: A lego-like Python framework for global and target analysis of time-resolved spectra," *Photochem. Photobiol. Sci.*, vol. 22, no. 10, pp. 2413–2431, Oct. 2023. DOI: `10.1007/s43630-023-00460-y`.

[67] T. Basché, "Fluorescence intensity fluctuations of single atoms, molecules and nanoparticles," *J. Lumin.*, Proceedings of the Eleventh International Conference on Dynamical Processes in Excited States of Solids, vol. 76–77, pp. 263–269, Feb. 1998. DOI: `10.1016/S0022-2313(97)00284-6`.

[68] P. Frantsuzov, M. Kuno, B. Jankó, and R. A. Marcus, "Universal emission intermittency in quantum dots, nanorods and nanowires," *Nat. Phys.*, vol. 4, no. 7, pp. 519–522, 7 Jul. 2008. DOI: `10.1038/nphys1001`.

[69] D. L. Ensign and V. S. Pande, "Bayesian detection of intensity changes in single molecule and molecular dynamics trajectories," *J. Phys. Chem. B*, vol. 114, no. 1, pp. 280–292, Jan. 2010. DOI: `10.1021/jp906786b`.

[70] S. Yin, N. Song, and H. Yang, "Detection of Velocity and Diffusion Coefficient Change Points in Single-Particle Trajectories," *Biophysical Journal*, vol. 115, no. 2, pp. 217–229, Jul. 2018. DOI: `10.1016/j.bpj.2017.11.008`.

[71] N. Song and H. Yang, "Parallelization of change point detection," *J. Phys. Chem. A*, vol. 121, no. 27, pp. 5100–5109, Jul. 2017. DOI: `10.1021/acs.jpca.7b04378`.

[72] H. Li and H. Yang, "Statistical learning of discrete states in time series," *J. Phys. Chem. B*, vol. 123, no. 3, pp. 689–701, Jan. 2019. DOI: `10.1021/acs.jpcb.8b10561`.

[73] J. P. Hoogenboom, W. K. den Otter, and H. L. Offerhaus, "Accurate and unbiased estimation of power-law exponents from single-emitter blinking data," *J. Chem. Phys.*, vol. 125, no. 20, p. 204 713, Nov. 2006. DOI: `10.1063/1.2387165`.

[74] J. Houel et al., "Autocorrelation analysis for the unbiased determination of power-law exponents in single-quantum-dot blinking," *ACS Nano*, vol. 9, no. 1, pp. 886–893, Jan. 2015. DOI: `10.1021/nn506598t`.

[75] J. H. Ward, "Hierarchical grouping to optimize an objective function," *J. Am. Stat. Assoc.*, vol. 58, no. 301, 1963. DOI: `10.1080/01621459.1963.10500845`.

[76] P. Tinnefeld, D.-P. Herten, and M. Sauer, "Photophysical Dynamics of Single Molecules Studied by Spectrally-Resolved Fluorescence Lifetime Imaging Microscopy (SFLIM)," *J. Phys. Chem. A*, vol. 105, no. 34, pp. 7989–8003, Aug. 2001. DOI: `10.1021/jp010365l`.

[77] K. D. Weston et al., "Measuring the Number of Independent Emitters in Single-Molecule Fluorescence Images and Trajectories Using Coincident Photons," *Anal. Chem.*, vol. 74, no. 20, pp. 5342–5349, Oct. 2002. DOI: `10.1021/ac025730z`.

[78] A. Grinvald and I. Z. Steinberg, "On the analysis of fluorescence decay kinetics by the method of least-squares," *Anal. Biochem.*, vol. 59, no. 2, pp. 583–598, 1974. DOI: `10.1016/0003-2697(74)90312-1`.

[79] Ž. Bajzer, T. M. Therneau, J. C. Sharp, and F. G. Prendergast, "Maximum likelihood method for the analysis of time-resolved fluorescence decay curves," *Eur. Biophys. J.*, vol. 20, no. 5, pp. 247–262, Dec. 1991. DOI: `10.1007/BF00450560`.

[80] J. Durbin and G. S. Watson, "Testing for serial correlation in least squares regression. II," *Biometrika*, vol. 38, no. 1/2, 1951. DOI: `10.2307/2332325`.

[81] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*. Cambridge: Cambridge University Press, 1995. DOI: `10.1017/cbo9781139644105`.

[82] R. Hanbury Brown and R. Q. Twiss, "Correlation between Photons in two Coherent Beams of Light," *Nature*, vol. 177, no. 4497, pp. 27–29, Jan. 1956. DOI: `10.1038/177027a0`.

[83] I. M. Palstra et al., "Intermittency of CsPbBr3 Perovskite Quantum Dots Analyzed by an Unbiased Statistical Analysis," *J. Phys. Chem. C*, vol. 125, no. 22, pp. 12 061–12 072, Jun. 2021. DOI: `10.1021/acs.jpcc.1c01671`.

[84] J. M. Troiano et al., "Identification of distinct pH- and zeaxanthin-dependent quenching in LHCSR3 from Chlamydomonas reinhardtii," *eLife*, vol. 10, D. M. Kramer and C. S. Hardtke, Eds., e60383, Jan. 2021. DOI: `10.7554/eLife.60383`.

[85] H. Wilson and Q. Wang, "ABEL-FRET: Tether-free single-molecule FRET with hydrodynamic profiling," *Nat. Methods*, vol. 18, no. 7, pp. 816–820, 7 Jul. 2021. DOI: `10.1038/s41592-021-01173-9`.

[86] K. P. Burnham, D. R. Anderson, and K. P. Burnham, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. New York: Springer, 2002, ISBN: 978-0-387-95364-9.

[87] J. M. Gruber et al., "Photophysics in single light-harvesting complexes II: From micelle to native nanodisks," in *Single Mol. Spectrosc. Superresolution Imaging IX*, vol. 9714, 2016, pp. 45–54.

[88] M. Belitsch et al., "Gray State Dynamics in the Blinking of Single Type I Colloidal Quantum Dots," *Nano*, vol. 13, no. 04, p. 1 850 039, Apr. 2018. DOI: `10.1142/S179329201850039X`.

[89] J. K. Karlsson, A. Laude, M. J. Hall, and A. Harriman, “Photo-isomerization of the cyanine dye alexa-fluor 647 (af-647) in the context of dstorm super-resolution microscopy,” *Chem. Eur. J.*, vol. 25, no. 65, pp. 14 983–14 998, 2019.

[90] L. Valkunas et al., “How photosynthetic proteins switch,” *J. Phys. Chem. Lett.*, vol. 3, no. 19, pp. 2779–2784, 2012. DOI: `10.1021/jz300983r`.

[91] J. Chmeliov et al., “Fluorescence blinking of single major light-harvesting complexes,” *New J. Phys.*, vol. 15, 2013. DOI: `10.1088/1367-2630/15/8/085007`.

[92] T. P. J. Krüger and R. Van Grondelle, “Design principles of natural light-harvesting as revealed by single molecule spectroscopy,” *Phys. B Condens. Matter*, vol. 480, pp. 7–13, 2016. DOI: `10.1016/j.physb.2015.08.005`.

[93] M. Schörner et al., “Multi-Level, Multi Time-Scale Fluorescence Intermittency of Photosynthetic LH2 Complexes: A Precursor of Non-Photochemical Quenching?” *J. Phys. Chem. B*, vol. 119, no. 44, pp. 13 958–13 963, 2015. DOI: `10.1021/acs.jpcb.5b06979`.

[94] M. Schörner et al., “Conformational Memory of a Protein Revealed by Single-Molecule Spectroscopy,” *J. Phys. Chem. B*, vol. 119, no. 44, pp. 13 964–13 970, 2015. DOI: `10.1021/acs.jpcb.5b07494`.