

Algorithms for the Correction of Photobleaching

Rory Nolan

University of Oxford

Contents

Preface	7
Ways to read	7
Apportion of credit	7
Publications	8
Acronyms used in this thesis	11
Mathematical symbols used in this thesis	12
Abstract	13
1 Introduction	15
1.1 Fluorescence	15
1.2 Confocal light microscopy	17
1.3 Widefield light microscopy	24
1.4 Intensity traces	29
1.5 Diffusion	30

1.6	Fluorescence fluctuation spectroscopy	32
1.7	Fluorescence correlation spectroscopy	37
1.8	Applications of FCS and FFS	44
2	Instrumentation and Software	47
2.1	Instrumentation	47
2.2	Software programs, languages and tools	48
2.3	My software packages	50
3	Photobleaching Correction	55
3.1	Introduction to photobleaching	55
3.2	The effects of bleaching in FCS and FFS	57
3.3	Exponential fitting detrending	60
3.4	Boxcar smoothing detrending	66
3.5	Choosing the correct smoothing parameter for detrending	69
3.6	Exponential smoothing detrending	71
3.7	Correcting for non-stationary variance	74
3.8	Caveats of fitting and smoothing approaches to detrending	75
3.9	Robin Hood detrending	77
3.10	A comparison of detrending methods	80
4	Applications	85

CONTENTS	5
4.1 Dimerization of FKBP12	85
4.2 <i>In vitro</i> number and brightness	90
4.3 HIV-1 receptor stoichiometry	92
5 Discussion	103
5.1 Fluorescence fluctuation spectroscopy	103
5.2 The evolution of detrending algorithms	104
5.3 Applications of the new detrending techniques	106
6 Future plans	109
6.1 Robin Hood publication	109
6.2 Translate software to <i>ImageJ</i>	109
6.3 Study real data bleaching profiles with simulations	110
6.4 Compare FRET with FCS	110

Preface

Ways to read

This thesis may be read on the web at <https://rorynolan.github.io/phdthesis/>. If you are reading this on the web now but would like a PDF version, click on the download symbol at the top left of the page (to the right of the **A**) and select PDF.

Apportion of credit

For clarity I include the following reliable rules of thumb:

- Molecular biology was done by Maro Iliopoulou and Luis Alvarez.
- Imaging was done by Maro Iliopoulou, Luis Alvarez and Sergi Padilla-Parra.

- The idea that correction for bleaching was the crucial step for FFS analysis was formulated by Luis Alvarez, Sergi Padilla-Parra and I.
- I formulated the solutions for how to correctly correct for bleaching, i.e. the automatic parameter choice and the Robin Hood algorithm.
- I wrote all of the software and maintain all of it.
- All FFS analysis was performed using my software. The software was used to analyse data by Maro Iliopoulou, Luis Alvarez, Sergi Padilla Parra and I.
- Structural modelling was done by Thomas Bowden and Yasunori Watanabe.
- On all papers where I am the first listed author, I wrote the paper, taking suggested amendments and augmentation from other listed authors. The NSMB paper ([Iliopoulou et al., 2018](#)) was written by Sergi; I also made significant contributions to the writing of that manuscript but my main role in that project was in data analysis.

Publications

I hereby list my publications. These can all be downloaded from <https://github.com/rorynolan/phdthesis/tree/master/papers>, where the naming convention is JournalYEAR.pdf.

First author

- R. Nolan and S. Padilla-Parra. “filestrings: An R package for file and string manipulation”. In: *Journal of Open Source Software* 2.14 (2017).
- R. Nolan and S. Padilla-Parra. “exampletestr—An easy start to unit testing R packages”. In: *Wellcome Open Research* 2 (2017).
- R. Nolan, L. Alvarez, J. Elegheert, et al. “nandb—number and brightness in R with a novel automatic detrending algorithm”. In: *Bioinformatics* 33.21 (2017).
- R. Nolan, M. Iliopoulou, L. Alvarez, et al. “Detecting protein aggregation and interactions in live cells: a guide to Number and Brightness”. In: *Methods* (2017).
- R. Nolan and S. Padilla-Parra. “ijtiff: An R package providing TIFF I/O for ImageJ users”. In: *Journal of Open Source Software* 3.23 (2018).
- R. Nolan, L. Alvarez, S. C. Griffiths, et al. “Calibration-Free In-Vitro Quantification of Protein Homo-Oligomerization Using Commercial Instrumentation and Free, Open Source Brightness Analysis Software”. In: *Journal of Visualized Experiments* 0.0 (2018).

Co-first author

- M. Iliopoulou, R. Nolan, et al. “A dynamic three step mechanism drives the HIV-1 pre-fusion reaction”. In: *Nat. Struct. Mol. Biol.* 0.0 (2018).

Other

- D. M. Jones, L. A. Alvarez, R. Nolan, et al. “Dynamin-2 stabilizes the HIV-1 fusion pore with a low oligomeric state”. In: *Cell reports* 18.2 (2017).
- G. M. Jakobsdottir, M. Iliopoulou, R. Nolan, et al. “On the whereabouts of HIV-1 cellular entry and its fusion ports”. In: *Trends in molecular medicine* (2017).
- Q. F. Wills, E. Mellado-Gomez, R. Nolan, et al. “The nature and nurture of cell heterogeneity: accounting for macrophage gene-environment interactions with single-cell RNA-Seq”. In: *BMC genomics* 18.1 (2017).

In preparation

- The *Robin Hood* algorithm (described in section 3.9) will be submitted to a journal after the submission of this thesis but before the end of my PhD contract.

Acronyms used in this thesis

Acronym	Term
ACF	AutoCorrelation Function
ALEX	Alternating Laser Excitation
APD	Avalanche Photo-Diode
BMC	BioMed Central
CCD	Charge-Coupled Device
ccN&B	cross-correlated Number and Brightness
CCR	C-Chemokyne Receptor 5
CD4	Cluster of Differentiation 4
CMOS	Complementary Metal–Oxide Semiconductor
CXCR4	C-X-C motif Chemokyne Receptor 4
DNA	Deoxyribonucleic Acid
dSTORM	direct Stochastic Optical Reconstruction Microscopy
EM	Electron Microscopy
EMCCD	Electron-Multiplied Charge-Coupled Device
FCS	Fluorescence Correlation Spectroscopy
FFS	Fluorescence Fluctuation Spectroscopy
FIJI	FIJI Is Just ImageJ
FKBP	FK Binding Protein
FRET	Forster Resonance Energy Transfer
GaAsP	Gallium arsenide phosphide
GFP	Green Fluorescent Protein

Mathematical symbols used in this thesis

Symbol	Meaning
θ	angle of incident light ray
B	apparent brightness of entities
N	apparent number of entities
G	autocorrelation
φ	concentration
corr	correlation
θ_c	critical angle
crosscor	cross-correlation
B_{cc}	cross-correlated brightness
σ^2_{cc}	cross-variance
\tilde{x}	deviation of x from the mean μ of X
D	diffusion constant
J	diffusive flux
ϵ	entity brightness
E	expectation operator
τ	exponential smoothing parameter
j	frame number
S_0	highest occupied molecular orbital
I	intensity
l	length parameter of boxcar
T_1	lowest energy triplet state

Abstract

The measured intensity (ideally in units of photon counts) of a fluorescent sample over time constitutes a time-series called an *intensity trace*. The idea of *fluorescence fluctuation spectroscopy* (FFS) is to extract information from intensity traces.

Photobleaching is the phenomenon of the breaking of fluorophores (light emitters) over time. Photobleaching causes fluorescent signal to diminish over time. This changes the intensity trace, introducing a downward trend.

Many quantitative methods in FFS implicitly assume that there is no bleaching in the data. Hence, data with significant levels of photobleaching must be corrected prior to the application of equations and algorithms in these fields. This correction is often termed *detrending*, since its aim is to remove the downward trend in the data introduced by photobleaching.

Previous detrending methods have two main caveats.

1. They rely on either fitting or smoothing, both of which approximate data as continuous. This is inappropriate for fluorescence intensity data, which is count data (i.e. discrete, not continuous).
2. They require the user to choose a detrending parameter. The choice of this parameter is crucial to the success or failure of the detrending routine, yet instructions on how to choose it did not exist.

The work in this thesis solves problems 1 and 2 above by means of an automatic (no user input required) parameter finding routine and a new detrending algorithm which treats the data as discrete and avoids fitting and smoothing, thereby avoiding the approximation of non-continuous data as continuous.

These advancements are then used in a study investigating the stoichiometry of the interaction of the HIV-1 virus' envelope with its cellular receptors and co-receptors over time. This is the first study of its kind in live cells and it was facilitated by the advances in detrending presented in this thesis.

Chapter 1

Introduction

1.1 Fluorescence

Definition 1.1. *Fluorescence* is the emission of photons related to the process of *relaxation* for a given molecule from an electronically *excited* state (S_1) to the *ground* state (S_0).

These two electronic states (S_0 and S_1) of a molecule are defined as the highest occupied molecular orbital (S_0 , HOMO) and the lowest unoccupied molecular orbital (S_1 , LUMO). Each electronic state has various associated *vibrational* states between which *non-radiative* transitions can occur. An electron can be *excited*

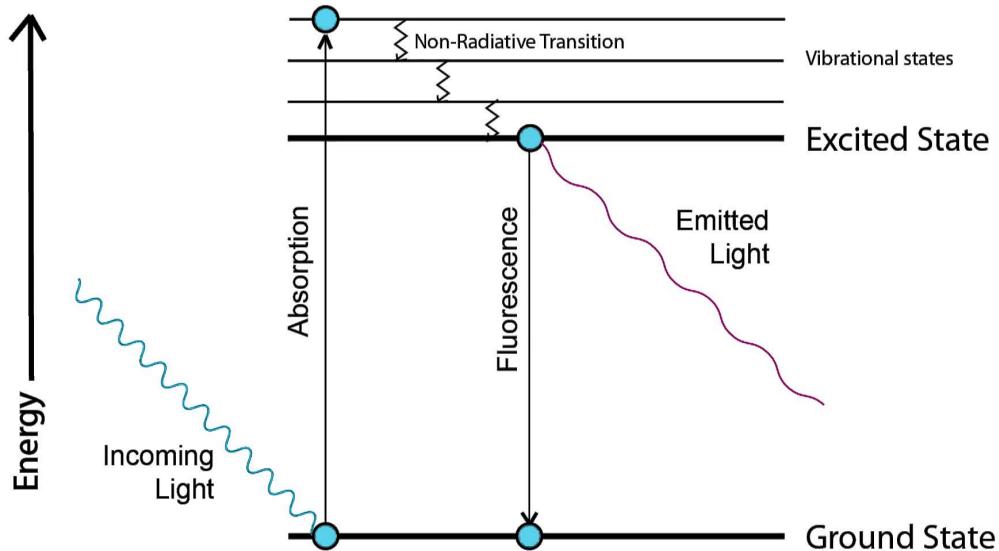


Figure 1.1: Jablonski diagram of the process of fluorescence. Absorption causes excitation, relaxation causes the emission of light.

from the ground state S_0 into the higher energy excited state S_1 by absorption of a photon of the appropriate energy. The electron will remain in the excited state (possibly undergoing non-radiative transitions through vibrational states) for a period of nanoseconds before *relaxing* to the lower energy ground state, losing energy by means of emitting a photon of that energy. This is shown by means of a *Jablonski* diagram (figure 1.1).

Definition 1.2. The *fluorescence lifetime* of an electron is the amount of time it remains in the excited state before returning to the ground state.

Definition 1.3. A *fluorophore* is a fluorescent chemical compound that can re-emit light upon light excitation.

Fluorophores under constant excitation emit light one photon at a time according to Poisson statistics.

1.1.1 Phosphorescence

Definition 1.4. Excited electrons can undergo a *forbidden* transition into the *triplet* state T_1 , where they remain for a period of milliseconds before relaxing to the ground state with the emission of a photon. This is *phosphorescence*.

1.1.2 Photobleaching

In reality, an incident photon can *break* the fluorophore such that it is subsequently unable to emit light. A fluorophore to which this has happened is said to be *photobleached*.

1.2 Confocal light microscopy

All of the images used in my PhD were collected on a confocal microscope. This type of microscope guarantees that only in-focus light is collected at the detector. See figure 1.2.

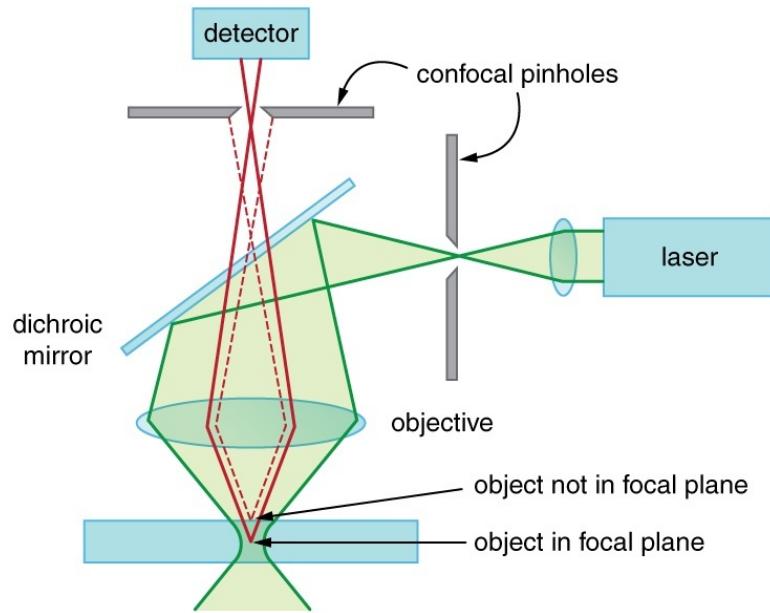


Figure 1.2: Confocal microscope light path showing how out of focus light does not make it to the detector.²

Definition 1.5. The *confocal volume* is the *in-focus* volume within a sample that is efficiently detected using a system designed with confocal optics ([FCSXpert, 2018](#)).

An image is acquired on a confocal microscope by scanning this apparatus across a sample, collecting one pixel at a time.

1.2.1 Detectors

There are many types of confocal microscope detector. I will discuss the most common ones here. Most rely on the photoelectric effect ([Einstein, 1905a](#)), a

phenomenon whereby incident light upon a certain material causes electrons to dissociate from that material.

Definition 1.6. The *quantum efficiency* (QE) of a detector is the fraction of fluorescent signal that is reported by the detector.

The quantum nature of the photoelectric effect makes it practically impossible for detectors to reach 100% efficiency.

1.2.1.1 Photon multiplier tubes

Photon multiplier tubes (PMTs) use the photoelectric effect. Dissociated electrons are accelerated through a potential difference towards a cathode where their accumulated kinetic energy is used to release more electrons. These are then accelerated in vacuum towards another cathode and so on for a predefined number of these accelerations until at last the electrons are discharged into an anode where the current is measured ([Hammamatsu, 2007](#)). See figure 1.3. This current corresponds to the incident light intensity.

The QE of standard PMTs is approximately 25% for blue-green light. This is because many incident photons on the multi-alkali material of regular PMT photocathodes fail to free any electrons. This problem gets worse at higher wavelengths where photons have lower energy. PMTs also suffer significantly from thermal

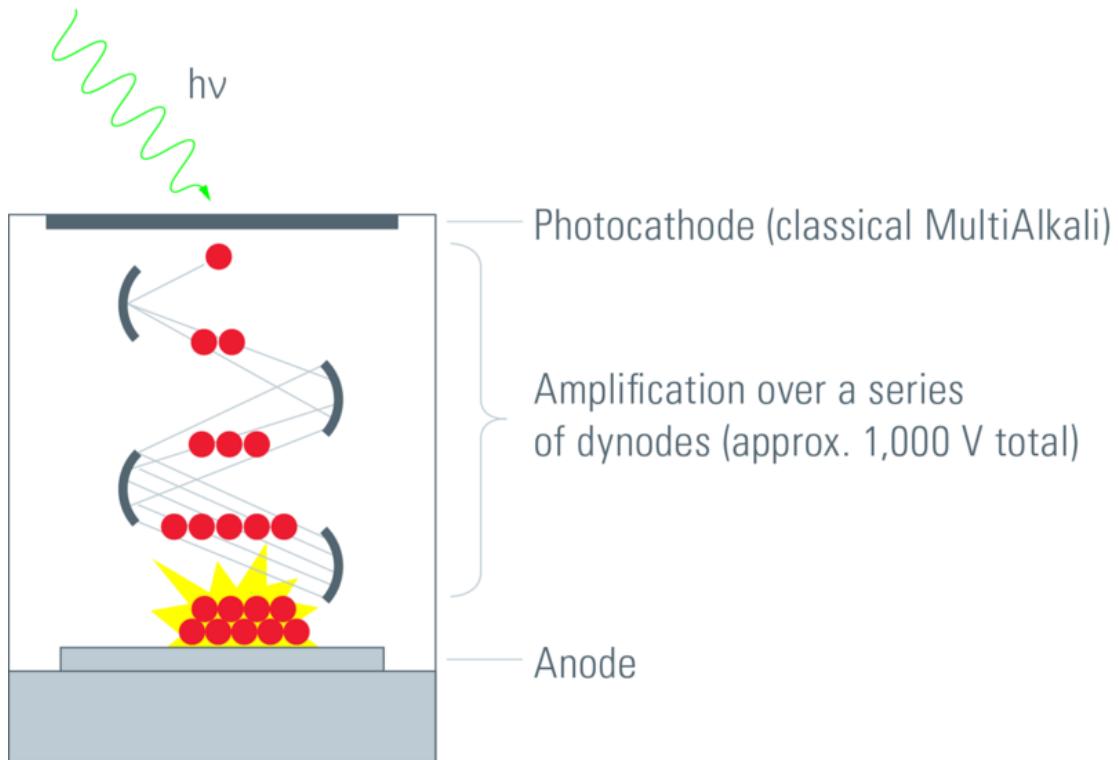


Figure 1.3: PMT detector setup. Electrons are dissociated by photons at the photocathode, accelerated towards various other cathodes where more and more are freed and then finally they arrive at the anode where the current is measured (Leica, 2012).

noise (whereby dissociated electrons are created due to heat energy).

1.2.1.2 Avalanche photo-diodes

Definition 1.7. *Avalanche multiplication* in semiconductors occurs when free electrons which are moving in an electric field across the semiconductor collide with bound electrons, freeing them to move and indeed to free more bound electrons. This compounding increase in free electrons in the electric field leads to an increase in the electric current.

Avalanche photo-diodes (APDs) are semiconductors that exploit the photoelectric effect together with avalanche multiplication (the photoelectric effect starts the avalanche) to convert light into measurable electric current. See figure 1.4. They are somewhat analogous to PMTs, using the avalanche effect in place of the series of cathodes.

The QE of APDs can be as high as 45%, however their dynamic range is low and they can only function with low-intensity light. APDs are less prone to thermal noise than PMTs.

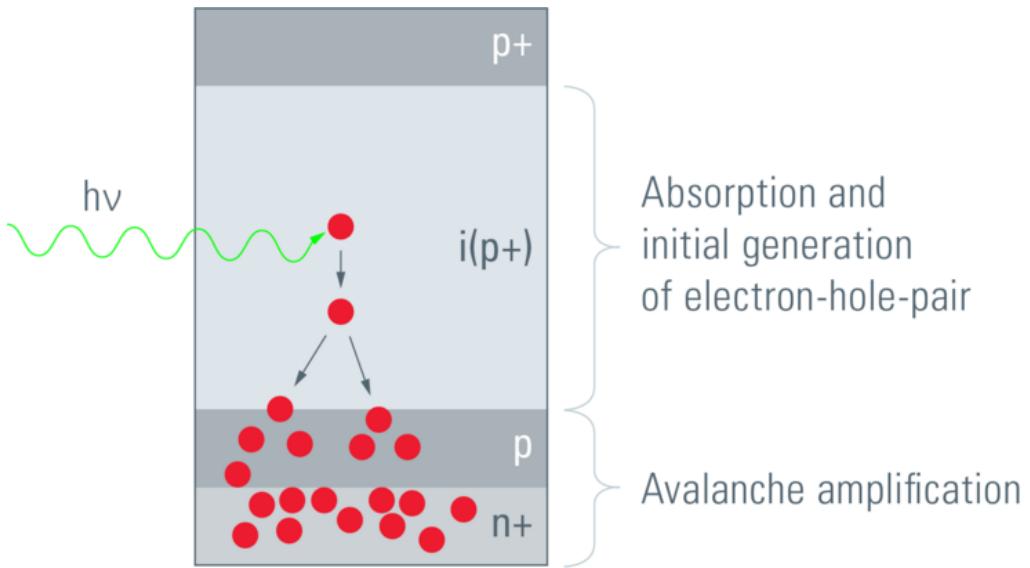


Figure 1.4: APD detector setup. The incident photon dissociates an electron. Given the electric field across the semiconductor, this electron is accelerated, initiating an avalanche (Leica, 2012).

1.2.1.3 Hybrid detectors

APDs have better sensitivity and lower noise, however PMTs have a larger dynamic range. In hybrid detectors (HyDs), PMT and APD technology is combined to get the best of both worlds. With HyDs, photons are converted to electrons at the photocathode, then accelerated in vacuum towards a semiconductor where they initiate an electron avalanche. See figure 1.5. The difference between HyDs and APDs is this acceleration step.

HyDs with GaAsP (gallium arsenide phosphide) photocathodes achieve a quantum efficiency of up to 50% (Leica, 2012). They also suffer least from detector

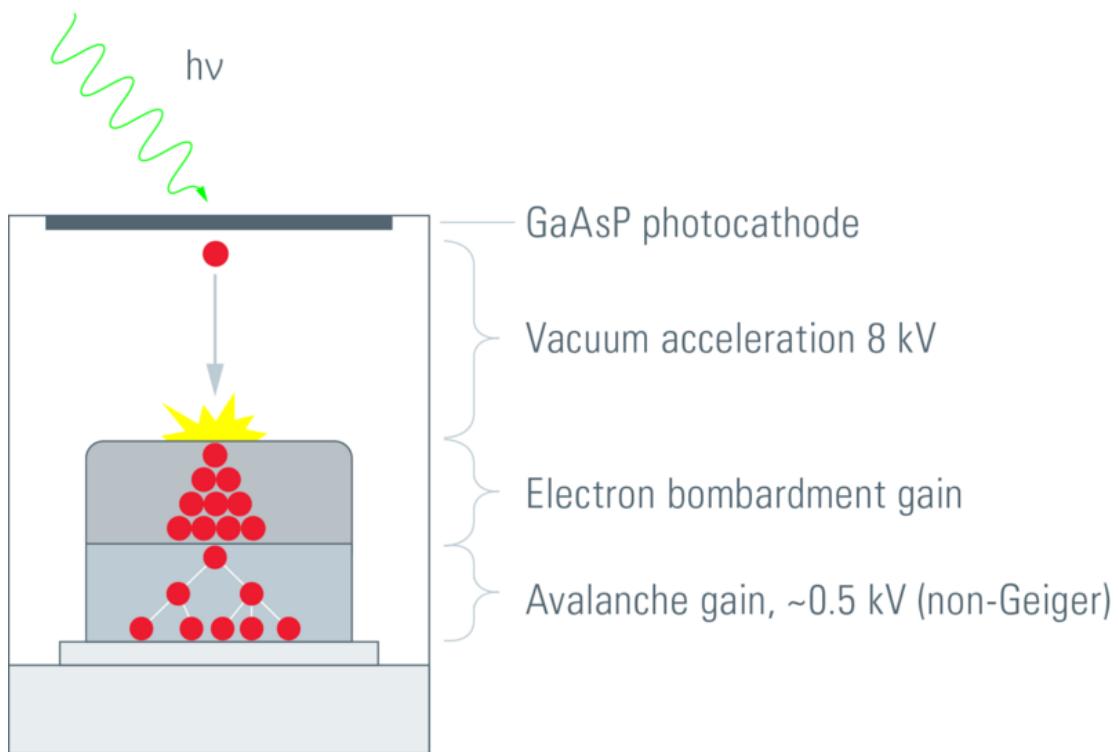


Figure 1.5: HyD detector setup. A photon dissociates an electron from the photocathode, this electron is then accelerated PMT-style towards an APD-style semiconductor setup, triggering an avalanche (Leica, 2012).

afterpulsing, a phenomenon which can cause real signal pulses to be followed by a feedback pulse at a later time (Zhao et al., 2003).

With their superior sensitivity and dynamic range and low noise, HyDs can be calibrated to enable *photon-counting* mode whereby the readout is in units of photons (not electron current). This is a much more physically relevant readout — given that we are interested in measuring fluorescence — which is very convenient for many biological applications, not least fluorescence fluctuation spectroscopy (FFS, discussed in section 1.6).

Remark. Henceforth, intensity counts will be assumed to be in units of photons unless otherwise stated.

1.3 Widefield light microscopy

With widefield microscopy, the entire sample is illuminated simultaneously and the image is acquired by a camera (see figure 1.6). One advantage of this is the general simplicity of the setup: the optics are less complex and there is no need to scan across the sample to acquire each pixel (which requires precise robotics). Another example is that images can be acquired faster because all pixels in a frame are acquired simultaneously. The disadvantage of this is that the absence of confocal optics means that out of focus light can make it to the detection device

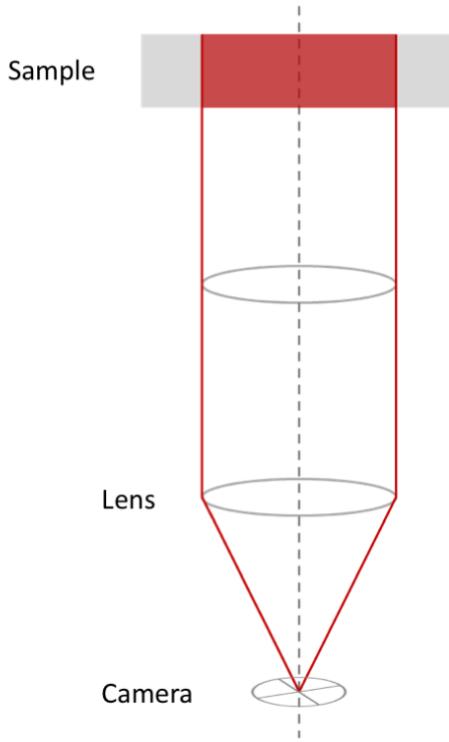


Figure 1.6: Light travels from the uniformly illuminated sample through a lens (or a few lenses) to the camera. Note the simplicity of this setup relative to that of a confocal microscope.

(camera).

1.3.1 Total internal reflection microscopy

Total internal reflection microscopy (TIRF) uses the fact that when total internal reflection (TIR) is occurring, an evanescent wave is produced at the media boundary, but this evanescent wave is only present very close to the boundary.

See figure 1.7 ([Mattheyses et al., 2010](#)). The huge advantage of TIRF is that only

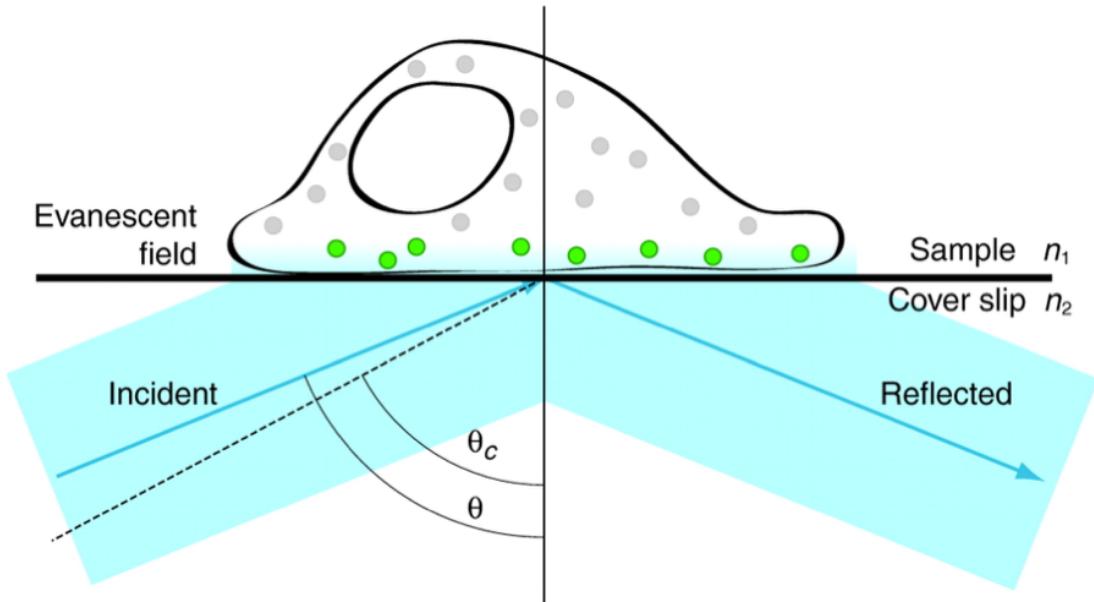


Figure 1.7: The thin evanescent wave close to the boundary (between the cover slip and sample) where TIR is occurring. Only the part of the sample within this thin evanescent wave is illuminated. n_1 is the refractive index of the sample and n_2 is the refractive index of the cover slip; $n_1 < n_2$ is assumed. θ_c is the critical angle and θ is the incident angle of the laser light; $\theta < \theta_c$ is assumed.

a single plane is illuminated, so when focusing on that plane, only in-focus light is collected (because only the in-focus plane is illuminated). The main disadvantage is that it is not applicable when one cannot have the object of interest right at the cover slip. Looking at figure 1.7, it is evident that TIRF would be ideal for imaging the membrane but useless for imaging the nucleus.

1.3.2 Cameras

There are many types of camera available for widefield microscopes, the most common being EMCCD, CMOS, and sCMOS cameras.

1.3.2.1 EMCCD cameras

A charge-coupled device (CCD) is a device for the movement of electric charge. An EMCCD camera is an array of CCDs (one per pixel). When photons strike a CCD in the camera, electrons are dissociated by the photoelectric effect. These free charges are then moved across the CCD and this current can be measured and used as a measure of light intensity (in arbitrary units) ([Andor, 2018a](#)). The EM stands for *electron multiplied* which tells us that EMCCD cameras use an amplification technique, just like PMTs and APDs (section [1.2.1](#)).

1.3.2.2 CMOS cameras

Complementary metal–oxide semiconductor cameras are arrays of semiconductors (one per pixel), again utilizing the photoelectric effect and the electric field across a semiconductor. CMOS cameras have amplification for each pixel, whereas EM-CCDs only have a few amplifiers per camera ([Baker, 2005](#)). This makes CMOS

cameras more sensitive.

1.3.2.3 sCMOS cameras

Scientific CMOS (sCMOS) cameras consist of a CMOS circuit bonded to a CCD imaging substrate. These relatively new cameras offer 95% quantum efficiency ([Photometrics, 2018](#)), very low noise, rapid frame rates, large dynamic range, high resolution and wide field of view, all cheaper than a regular CCD ([Andor, 2018b](#)).

1.3.2.4 Being quantitative with cameras

For electronic reasons, the noise from certain pixels on all of the cameras above may be correlated. This hinders some techniques in the fluorescence correlation spectroscopy (section [1.7](#)). However, with careful characterization of the camera and correct mathematical treatment, quantitative studies can be carried out with cameras. Indeed, our group quantified the oligomeric state of the dynamin-2 protein at the HIV-1 fusion pore using data from a TIRF microscope with an EMCCD camera ([Jones et al., 2017](#)).

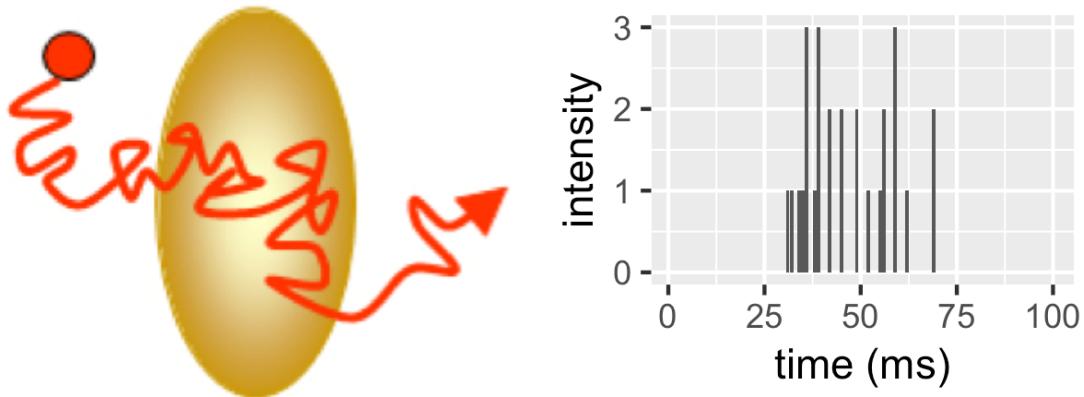


Figure 1.8: Left: a fluorophore diffusing through the confocal volume (Padilla-Parra, 2009). Right: the intensity trace due to this fluorophore.

1.4 Intensity traces

Consider figure 1.8. A fluorophore enters the confocal volume, is excited there and emits photons which are collected by the detector. When the fluorophore is not in the confocal volume (start and end), no photons are detected. When the particle is in the confocal volume, photons it emits are collected at the detector. Different numbers of photons are collected per unit time (per ms here).

Definition 1.8. The time-series of intensity counts in a given pixel or image is referred to as its *intensity trace*.

1.5 Diffusion

Definition 1.9. *Diffusion* is the net movement of molecules or atoms from a region of high concentration (or high chemical potential) to a region of low concentration (or low chemical potential) as a result of random motion of the molecules or atoms.

Fick's first law of diffusion relates the concentration to the diffusive flux, assuming a steady state. It describes how the flux goes from high to low concentration areas, with magnitude proportional to the concentration gradient. The law is:

$$J = -D \frac{d\varphi}{dx} \quad (1.1)$$

where

- J is the diffusive flux
- D is the diffusion constant
- x is the position in space
- φ is the concentration at position x .

Definition 1.10. *Brownian motion* is the random motion of particles suspended in a fluid resulting from their collision with the fast-moving molecules in the fluid

(Einstein, 1905b).

Definition 1.11. In an ensemble of N particles, each of which occupies position $x_i(t)$ at time t , the *mean squared displacement* of the particles in the ensemble is

$$\text{msd}(t) = \frac{1}{N} \sum_{i=1}^N [x_i(t) - x_i(0)]^2 \quad (1.2)$$

For free Brownian motion in 1 dimension, the expected value of the MSD is $\text{msd}(t) = 2Dt$, where D is the diffusion constant. For free Brownian motion in d dimensions, the expected value of the MSD is $\text{msd}(t) = 2dDt$, reflecting the fact that free Brownian motion in d dimensions is one-dimensional free Brownian motion happening simultaneously and independently in each individual dimension.

These definitions are the formally correct ones. In biophysics, however, both of these processes (diffusion and Brownian motion) are most often referred to as *diffusion*. Usually, in biophysics one is interested in proteins which move around in a medium like the cytoplasm or the cell membrane. When this is referred to as *diffusion*, it is what is defined as Brownian motion above. Henceforth when I use the term *diffusion*, this Brownian motion type of movement can be assumed.

1.5.1 Anomalous diffusion

Anomalous diffusion is a diffusion process whereby equation (1.2) no longer holds and the relationship between MSD and time becomes non-linear. Knowing whether diffusion is anomalous or free can offer biological insight. For example, regions of free and anomalous diffusion were measured with FCS (section 1.7) to investigate the spatiotemporal heterogeneity of lipid interaction in the plasma membrane of living cells (Honigmann et al., 2014). The pair correlation function approach (section 1.7.3.1) can be used to image barriers to diffusion (Digman and Gratton, 2009); diffusion barriers necessarily lead to anomalous diffusion.

1.6 Fluorescence fluctuation spectroscopy

Broadly, fluorescence fluctuation spectroscopy (FFS) is the analysis of the intensity fluctuation of a fluorescence signal (Chen et al., 1999). This very often takes the form of moment analysis (Qian and Elson, 1990). Briefly, moment analysis is an attempt to extract data from a distribution of values using its *moments*. The first moment of a distribution is its mean value, the second moment is its variance. The n^{th} moment of a random X with expected value $E[X] = \mu$ for $n > 1$ is $E[(X - \mu)^n]$.

Intensity traces can be viewed as distributions with moments. For example, the intensity trace in figure 1.8 has mean 0.27 and variance 0.48.

1.6.1 Number and brightness

Number and brightness (N&B, [Digman et al. \(2008\)](#)) is an FFS technique for quantifying the oligomeric states of fluorescently labelled proteins. What follows is a mathematical description of the technique.

Definition 1.12. An *entity* is a set of molecules which are chemically bound.

Definition 1.13. The brightness ϵ of an entity is the mean number of photon detector counts it gives per unit time when in the illumination (confocal) volume.

For an image series where the i^{th} slice in the stack is the image acquired at time $t = i$, for a given pixel position (x, y) , we define $\langle I \rangle$ as the mean intensity of that pixel over the image series and σ^2 as the variance in that intensity. Define n as the mean number of entities in the illumination volume corresponding to that pixel. Assuming that all entities are mobile, we have

$$N = \frac{\langle I \rangle^2}{\sigma^2} = \frac{\epsilon n}{1 + \epsilon} \quad (1.3)$$

$$B = \frac{\sigma^2}{\langle I \rangle} = 1 + \epsilon \quad (1.4)$$

where N and B are referred to as the *apparent number* and *apparent brightness* respectively. This gives

$$\epsilon = \frac{\sigma^2}{\langle I \rangle} - 1 \quad (1.5)$$

$$n = \frac{\langle I \rangle^2}{\sigma^2 - \langle I \rangle} \quad (1.6)$$

— Nolan et al. (2017b)³

The quantity ϵ is a relative measure of oligomeric state. That is, ϵ will be twice as big for dimers as it is for monomers, three times as big for trimers as it is for monomers and so on.

The way that N&B experiments to determine unknown oligomeric states are generally done is as follows:

³This is a formulation of N&B that I wrote in a 2017 review of the technique.

1. For a given laser power and fluorophore with a system where all entities are known to be monomeric, measure the brightness ϵ . Call this $\epsilon_{\text{monomer}}$.
2. With the same laser power and fluorophore but now with a system where the oligomeric state is unknown, measure the brightness ϵ again. Call this $\epsilon_{\text{unknown}}$.
3. The unknown oligomeric state is equal to $\epsilon_{\text{unknown}}/\epsilon_{\text{monomer}}$.

Number and brightness has specific pixel dwell-time and frame rate requirements.

These were first articulated in my 2017 review of N&B ([Nolan et al., 2017b](#)).

Definition 1.14. The *pixel dwell time* t_{dwell} of a scanning confocal microscope is the amount of time it spends collecting photons at each given pixel before moving on to the next pixel.

It is essential that the pixel dwell time is the same for each pixel, particularly for FFS. This is something that must be carefully verified on each instrument. This can be done with a chroma-slide. The photon counts measured from a chroma-slide should follow a Poisson distribution. If instead the distribution is super-Poissonian (has a greater variance than its mean), this is indicative of a non-constant dwell-time.

Definition 1.15. The *frame time* of a scanning confocal microscope is the amount of time it takes to acquire the data for all of the pixels in a whole frame. That is the length of time from the start of detectionion of photons from the first pixel

to the end of detection of photons from the last pixel.

Definition 1.16. The mean *residence time* τ_D of a particle in the confocal volume is the average length of time that a particle that enters the confocal volume resides in that volume before exiting.

The requirement is $t_{\text{dwell}} \ll \tau_D \ll t_{\text{frame}}$. This ensures that:

1. When acquiring photons at a given pixel, the underlying configuration of entities is constant (there's not enough time for the entities to move and change their configuration).
2. When the scanner returns to a given pixel (one frame time later), the underlying configuration has changed totally since the last time the scanner was at this pixel (because so much time has passed, all of the diffusing entities have moved a lot in the meantime).

Both of these points are implicitly assumed in the derivation of the N&B equations so it is essential to get these acquisition parameters right. This is discussed at length in [Nolan et al. \(2017b\)](#).

An important property of N&B is that if all fluorescent particles are immobile, then $B = 1$. This is because photon emission from stationary sources happens according to a Poisson distribution. Poisson distributions have variance equal to mean, this implies $\sigma^2 = \langle I \rangle$ which gives $B = \frac{\sigma^2}{\langle I \rangle} = 1$.

The N&B technique is fraught with technical difficulties. Principal of these is the problem of photobleaching. Most of my PhD focused on corrections for photobleaching. This is discussed in chapter 3.

1.7 Fluorescence correlation spectroscopy

Fluorescence correlation spectroscopy (FCS) is the correlation analysis of fluorescence intensity fluctuations. For this reason, FCS can be described as a subfield of FFS (Jameson et al., 2009). In practice, FFS is mostly used to refer to the non-FCS parts of the whole FFS field. I will follow that convention.

First, let us introduce some concepts from statistics.

Definition 1.17 (statistics). The *correlation* between two random variables X and Y with expected values μ_X and μ_Y and standard deviations σ_X and σ_Y is

$$\text{corr}(X, Y) = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (1.7)$$

where E is the expectation operator.

Definition 1.18 (statistics). *Autocorrelation* $G(X; \tau)$, is the correlation of a signal X with a delayed copy of itself as a function of the delay τ .

$$G(X; \tau) = \text{corr}(X_t, X_{t+\tau}) = \frac{E[(X_t - \mu_X)(X_{t+\tau} - \mu_X)]}{\sigma_X^2} \quad (1.8)$$

Definition 1.19 (statistics). The *cross-correlation* of two series is the correlation of one with a delayed copy of the other as a function of the delay.

$$\text{crosscorr}(X, Y; \tau) = \text{corr}(X_t, Y_{t+\tau}) = \frac{E[(X_t - \mu_X)(Y_{t+\tau} - \mu_Y)]}{\sigma_X \sigma_Y} \quad (1.9)$$

For the purposes of FCS, these quantities were redefined as follows.

Definition 1.20 (FCS). The *correlation* between two random variables X and Y with expected values μ_X and μ_Y and standard deviations σ_X and σ_Y is

$$\text{corr}(X, Y) = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\mu_X \mu_Y} \quad (1.10)$$

where E is the expectation operator.

Definition 1.21 (FCS). *Autocorrelation* $G(X; \tau)$, is the correlation of a signal X with a delayed copy of itself as a function of the delay τ .

$$G(X; \tau) = \text{corr}(X_t, X_{t+\tau}) = \frac{E[(X_t - \mu_X)(X_{t+\tau} - \mu_X)]}{\mu_X^2} \quad (1.11)$$

Definition 1.22 (FCS). The *cross-correlation* of two series is the correlation of one with a delayed copy of the other as a function of the delay.

$$\text{crosscorr}(X, Y; \tau) = \text{corr}(X_t, Y_{t+\tau}) = \frac{E[(X_t - \mu_X)(Y_{t+\tau} - \mu_Y)]}{\mu_X \mu_Y} \quad (1.12)$$

The reason for these redefinitions (which just involve replacing standard deviations with means in the denominators of each expression) is that with the FCS definition, the autocorrelation has the nice property that for normal diffusion

$$G(X; 0) = \frac{1}{n} \quad (1.13)$$

where n is the mean number of fluorescent particles in the focal volume. The convenience of the statistics definitions is that there, correlations are guaranteed to be in $[-1, 1]$, with 0 representing no correlation, 1 perfect positive correlation and -1 perfect negative correlation; this is lost with the FCS definitions: 0 still represents no correlation, but correlation values are no longer bounded, so the ideas of perfect correlation are lost.

I felt it necessary to provide these definitions for two reasons:

1. It is important for people from the fields of FCS and pure mathemat-

ics/statistics to know that they have different definitions for the same thing.

2. In FCS, it's very common for people to mistake correlation for cross-correlation. This is unfortunate, but knowing about this common mistake is essential for navigating the field in a sensible manner. It seems that when people in the FCS field correlate the signals from two separate channels, they use the term *cross-correlation*, even though they're only using correlation. I think the idea of working *across* two or more channels (and ideas such as *cross-talk*) leads to this confusion.

Remark. Henceforth, the FCS defintions of these quantities will be assumed.

1.7.1 Correlation

Suppose that two proteins of interest A and B are labelled with red and green fluorophores respectively (and there is no bleed-through between these red and green channels).

If these proteins are interacting, then

- interaction implies that A and B *co-diffuse*
- this implies that A and B co-diffuse
- this implies that for a given volume in the sample
- more of A implies more of B

- less of A implies less of B
- more of B implies more of A
- less of B implies less of A

Since the number of photons emitted is proportional to the amount of fluorophores present, it follows that

- more red photons implies more green photons
- less red photons implies less green photons
- more green photons implies more red photons
- less green photons implies less red photons

Altogether, this implies that the intensity traces from the red and green channels will be correlated. If these proteins are not interacting, then the intensity traces from the red and green channels will not be correlated. Thus, interaction of red A and green B necessarily leads to correlation in fluorescent signals from the red and green channels.⁴

⁴This correlation may not always be detectable due to weak interaction or weak signal.

1.7.2 Autocorrelation

As mentioned already, the autocorrelation function (ACF) can be used to count the number of particles in the confocal volume. It can also be used to measure diffusion coefficients for various types of diffusion (normal, anomalous, polydisperse, etc.). The ACF is not used in my PhD.

1.7.3 Cross-correlation

The cross-correlation of intensity traces from nearby pixels can be used to measure the velocity of the movement of the labelled particles between these two pixels (Hebert et al., 2005).

1.7.3.1 Pair correlation function

The phrase *pair correlation function*⁵ was coined for this idea of cross-correlating intensity traces from nearby pixels. The PCF was used to image barriers to diffusion (Digman and Gratton, 2009) using the idea that the spatiotemporal correlation caused by particles moving from one place to another will not be present for positions p_1 and p_2 if they are on opposite sides of a barrier, because

⁵This was needless, it could just be called *cross-correlation function* but for the confusion about the term *cross-correlation* in the FCS field.

the barrier prevents travel from p_1 to p_2 . It has also been used to create *diffusion tensors*: maps of diffusion velocities at every pixel (Rienzo et al., 2016); this approach is similar to that of Hebert et al. (2005).

1.7.4 Cross-correlated brightness

Cross-correlated brightness (Digman et al., 2009) molds the correlation idea (section 1.7.1) into the framework of N&B. Suppose there are two channels with intensities I_1 and I_2 .

Definition 1.23 (cross-variance).

$$\sigma_{cc}^2 = E[(I_2 - \langle I_1 \rangle)(I_2 - \langle I_2 \rangle)] \quad (1.14)$$

Definition 1.24 (cross-correlated brightness).

$$B_{cc} = \frac{\sigma_{cc}^2}{\sqrt{\langle I_1 \rangle \langle I_2 \rangle}} \quad (1.15)$$

B_{cc} is related to $\text{corr}(I_1, I_2)$ by

$$B_{cc} = \sqrt{\langle I_1 \rangle \langle I_2 \rangle} \times \text{corr}(I_1, I_2) \quad (1.16)$$

This means that B_{cc} is just a scaled version of correlation. The need for this redefinition is unclear (but it is no harm). It does make the formula look like the brightness formula (1.4), but no such oligomeric state information can be gleaned from B_{cc} . It is merely useful as a relative measure of interaction: higher B_{cc} means more interaction, lower B_{cc} means less interaction. It is commonly used to identify interactions. Then, conventional N&B performed on each of the channels (1 and 2) can be used to measure the stoichiometry of the interaction.

Remark. Since cross-correlated brightness uses correlation but not cross-correlation, it is a prime example of the confusing naming that pervades FCS. It should be called *correlated brightness*. Rather than rename it, I will continue to refer to it as cross-correlated brightness.

1.8 Applications of FCS and FFS

FCS and FFS have been used in thousands of research projects. Here I number but a few for the sake of interest and to give biological context to these techniques.

Number and brightness (the prominent imaging FFS technique) has been used to:

- characterize the state of DNA aggregation in live cells ([Mieruszynski et al., 2015](#))

- measure the stoichiometry of scaffold complexes in live neurons (Moutin et al., 2014)
- quantify interactions in gene expression networks (Declerck and Royer, 2013)
- measure the oligomeric state of the dynamin-2 protein at the HIV-1 fusion pore (Jones et al., 2017)
- measure the stoichiometry of the interaction of HIV-1 with its receptor and co-receptor over time in the pre-fusion process (Iliopoulou et al., 2018)

FCS has been used to:

- reveal structural and functional properties of promyelocytic leukemia nuclear bodies (Hoischen et al., 2018)
- demonstrate that HIV-1 evades antibody-dependent phagocytosis (Gach et al., 2017)
- determine the size of nanodomains (Fenz et al., 2017)
- perform chromatographic measurements (Kisley and Landes, 2014)
- quantify interactions of membrane proteins (Ly et al., 2014)

Chapter 2

Instrumentation and Software

2.1 Instrumentation

All of the images used in this thesis were acquired on a Leica SP8 confocal microscope¹ equipped with hybrid detectors (section 1.2.1.3). Importantly, the detectors that we have on this microscope are capable of photon counting.

¹<https://www.Leica-microsystems.com/products/confocal-microscopes/details/product/Leica-tcs-sp8>

2.2 Software programs, languages and tools

2.2.1 C++

C++ (Stroustrup, 2013) is a general-purpose programming language optimized for performance (speed), efficiency (with use of computer resources) and flexibility.²

I used it for its speed, since many of the algorithms that I developed were quite computationally intensive and hence speed was an important consideration.

2.2.2 R

R (R Core Team, 2016) is a programming language and free software environment for statistical computing and graphics.³ I use R primarily as a wrapper for my C++ code to make my algorithms more user-friendly. R is best used with the *RStudio* integrated development environment.⁴

²<https://en.wikipedia.org/wiki/C%2B%2B>

³[https://en.wikipedia.org/wiki/R_\(programming_language\)](https://en.wikipedia.org/wiki/R_(programming_language))

⁴<https://www.rstudio.com>

2.2.3 ImageJ

ImageJ [ImageJ](#) is an open source image processing program designed for scientific multidimensional images.⁵ It is the preferred image viewing and analysis software in the community. I have written my software in C++ and R because they are easier for developers, but I still intend to translate my image-related algorithms ImageJ plugins. ImageJ is best used via the FIJI ([Schindelin et al., 2012](#)) distribution.⁶

2.2.4 Git and GitHub

Git is a free and open source distributed version control system designed to handle everything from small to very large projects with speed and efficiency.⁷ GitHub is a web-based hosting service for version control using Git.⁸ All of the computer code used during my thesis can be found on my GitHub at <https://github.com/rorynolan>. The vast majority of my time during my thesis was spent writing code so this GitHub account is the best record of the work that I have done.

⁵<https://imagej.net>

⁶<https://imagej.net/Fiji>

⁷<https://git-scm.com/>

⁸<https://en.wikipedia.org/wiki/GitHub>

2.3 My software packages

2.3.1 `ijtiff`

An R package for general purpose TIFF file I/O. This is currently the only such package with read and write support for TIFF files with floating point (real-numbered) pixels, and the only package that can correctly import TIFF files that were saved from *ImageJ* (Rueden et al., 2017). R has millions of users worldwide so this TIFF I/O capability is a basic need for masses of people. `ijtiff` gets 335 downloads per month which amounts to 2,563 since it was first released.

This package is part of the *rOpenSci* project. rOpenSci is a non-profit initiative founded to make scientific data retrieval reproducible.⁹ This package was peer reviewed and published (Nolan and Padilla-Parra, 2018).

2.3.2 `autothresholdr`

`autothresholdr` provides the *ImageJ* (Rueden et al., 2017) *Auto Threshold* plugin (Landini et al., 2016) functionality to R users. It gets 630 downloads per month which amounts to 10,995 since it was first released.

⁹<https://ropensci.org/about>

2.3.3 `detrendr`

`detrendr` is an R package for detrending images (correcting for photobleaching).

It provides all detrending algorithms mentioned in section 3. The detrending is done in C++ in the background for speed but it is wrapped in an R package for ease of use. It gets 360 downloads per month which amounts to 3,421 since it was first released.

2.3.4 `filestrings`

`filestrings` is an R package providing convenient functions for moving files, deleting directories, and a variety of string operations that facilitate manipulating file names and extracting information from strings. The motivation for making this package was to facilitate the use of file names for metadata. This is very common in microscopy, e.g. a file name like `well1_cell1_before_drug_addition.tif` is often seen. Using file names for metadata like this is a good idea, however if the naming or the extraction of data from these names is inconsistent, analysis becomes a nightmare and less reproducible. This package was peer reviewed and published ([Nolan and Padilla-Parra, 2017b](#)). It gets 802 downloads per month which amounts to 14,017 since it was first released.

2.3.5 `exaplestr`

Definition 2.1. *Unit testing* is a software testing method by which individual units of source code, sets of one or more computer program modules together with associated control data, usage procedures, and operating procedures, are tested to determine whether they are fit for use.¹⁰

Unit testing is a tool to verify that software is performing as intended. It is a great way to discover bugs in software. `exaplestr` is an R package which makes it easier for R package developers to write unit tests for their packages. It helped me to eradicate many bugs in all of my packages. Interestingly, `exaplestr` was used to unit test and eradicate bugs in itself! This package was peer reviewed and published in 2017.([Nolan and Padilla-Parra, 2017a](#)) It gets 509 downloads per month which amounts to 8,509 since it was first released.

2.3.6 `brownnded`

`brownnded` is an R software package (<https://github.com/rorynolan/brownnded>) for simulating bounded Brownian motion in any number of dimensions, where *bounded* Brownian motion is Brownian motion in an d -dimensional box where the particles collide elastically (without loss of energy) with the boundaries of the

¹⁰https://en.wikipedia.org/wiki/Unit_testing

box. **brownded** allows specification of the number of dimensions, the number of particles, the size of the box and the diffusion coefficient of the particles.

brownded also facilitates the simulation of images created from fluorescent particles undergoing bounded Brownian motion. It allows specification of the time at which each image should be taken, the pixel size and the brightness of the particles. Each fluorescent particle contributes photon counts to its pixel of residence at that time according to a Poisson process.

Finally, **brownded** facilitates the synthetic bleaching of fluorescent particles so bleaching can be investigated with images produced with **brownded**.

Chapter 3

Photobleaching Correction

3.1 Introduction to photobleaching

In the ideal case, an *incident* photon of appropriate wavelength is absorbed by a fluorophore, promoting the fluorophore to an excited state; subsequently, the fluorophore relaxes down to its ground state by emitting a photon. In reality, it is possible that the incident photon can *break* the fluorophore with the result that it will no longer emit light. This breaking is referred to as *photobleaching* (or *bleaching* for short). Bleaching causes a diminution in the number of effective fluorophores which is a direct cause of a loss of fluorescent signal.

Many quantitative methods in fields such as fluorescence fluctuation spectroscopy (FFS) and fluorescence correlation spectroscopy (FCS) implicitly assume that there is no bleaching in the data. Hence, data (image series) with significant levels of photobleaching must be corrected prior to the application of equations and algorithms in these fields. A main focus of this thesis is on how to correct fluorescent image series for the effects of bleaching, given that bleaching *does* occur. There is no attempt to understand *why* and/or *how* photobleaching occurs.

Remark. All of the current literature mentions bleaching correction as being purely for correcting the problem of non-stationary mean (NSM), neglecting the problem of non-stationary variance (NSV). Figure 3.3 shows that NSM and NSV go hand-in-hand. Correction for NSM is referred to as trend removal or *detrending*. Hence, the terms *detrending* and *bleaching correction* have come to be used interchangeably. I will follow this convention and use the term *detrending* from now on to mean correction for NSM and possibly also NSV. Starting at section 3.3, the focus is on correcting for NSM. Discussion of correction for NSV starts in section 3.7.

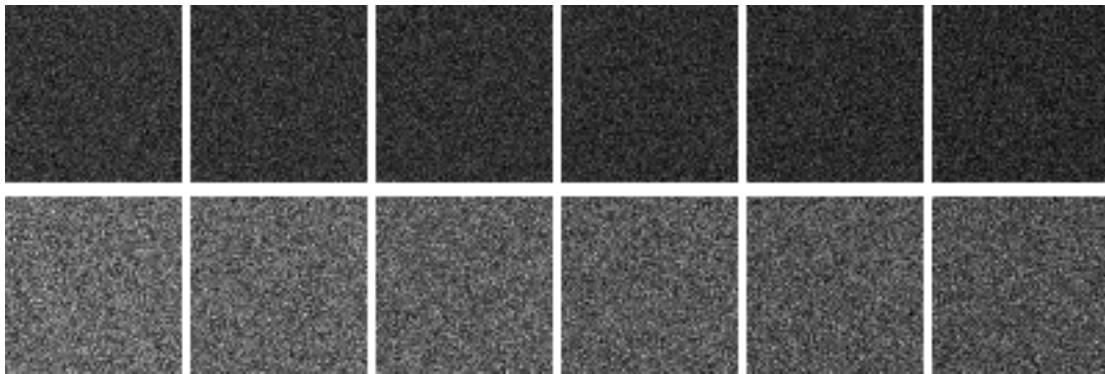


Figure 3.1: Frames 1, 100, 200, 300, 400 and 500 from image series of 100,000 diffusing emitters of brightness $\epsilon = 4$ (top) and $\epsilon = 7$ (bottom) with 15% (top) and 20% (bottom) bleaching.

3.2 The effects of bleaching in FCS and FFS

We simulate two image series, each with 100,000 diffusing fluorescent particles. In the first image series (`img1`), these have brightness $\epsilon = 4$ and in the second (`img2`) they have brightness $\epsilon = 7$. See figure 3.1. We bleach these by 15% and 20% to create `img1_bleached` and `img2_bleached` respectively.

It may not be obvious that these image series are subject to bleaching from figure 3.1, but we can see it more clearly in figure 3.2.

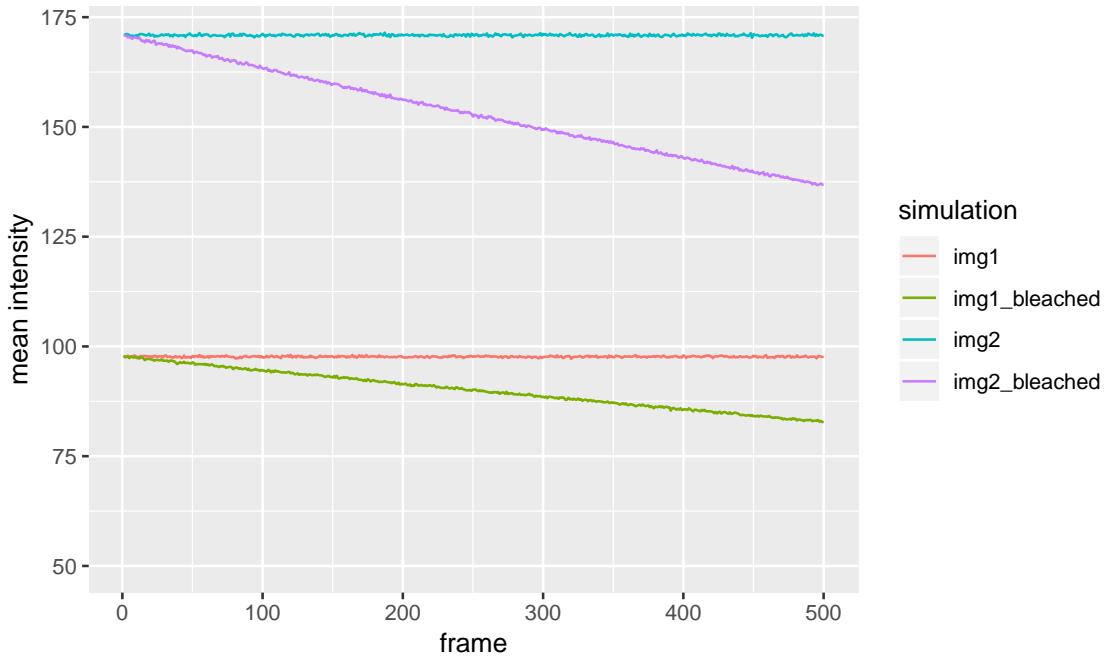


Figure 3.2: Mean intensity profiles of the simulated image series with and without bleaching. Simulation 1 was bleached by 15% and simulation 2 by 20%.

3.2.1 FCS

The unrelated images `img1` and `img2` have a tiny median cross-correlated brightness of $B_{cc} = 0.0036$, signifying no significant correlation, as one would expect. However, the bleached images `img1_bleached` and `img2_bleached` have a significant $B_{cc} = 0.3686$. This shows that bleaching is introducing correlation between otherwise unrelated images. Since correlation is used as a proxy for hetero-interaction, bleaching can make it appear as though there is interaction when in fact there is not.

3.2.2 FFS

Definition 3.1 (median mean pixel intensity). The *median mean pixel intensity* of an image series is found by taking the mean intensity of each pixel in the image series and then taking the median of those means. It can be thought of as a summary statistic for the pixel intensity of the image series.

Definition 3.2 (median pixel intensity variance). The *median pixel intensity variance* of an image series is found by taking the variance in the intensity of each pixel in the image series and then taking the median of those variances. It can be thought of as a summary statistic for the variance in the pixel intensity of the image series.

In FFS, one is always interested in the mean and variance of pixel values. `img1` has a median mean pixel intensity of 98 and a median pixel intensity variance of 487. The mean brightness is $\epsilon = 3.9959$ (very close to 4, as expected since the image series was simulated with brightness $\epsilon = 4$). For `img1_bleached`, we find a median mean pixel intensity of 90 and a median pixel intensity variance of 468. The mean brightness is $\epsilon = 4.2026$. Hence, bleaching has altered both the means and variances of the pixels, resulting in a change in calculated brightness. The non-stationary mean frame intensity introduced by bleaching decreases the mean but increases the variance. The loss in signal has the effect of slightly decreasing the variance: with Poisson statistics (such as photon-emission), a loss of signal (photons) leads directly to a loss in variance. This is a subtle point not discussed anywhere in the literature; it is shown in figure 3.3.

3.3 Exponential fitting detrending

Naively, one could assume that bleaching of fluorophores takes place at some constant rate. This would mean that the intensity of the image would die-off according to an exponential decay. In figure 3.4, we fit an exponential decay to such ideal data.

Having fit the data, one may record the deviations from the fitted line as the

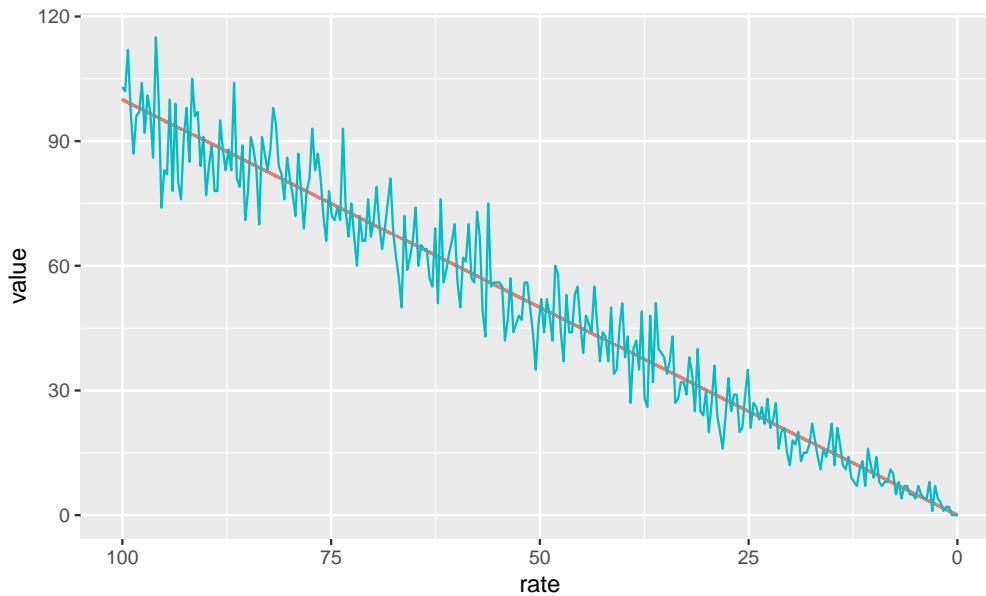


Figure 3.3: A decrease in the Poisson rate (e.g. for emission of photons) leads to a decrease in the mean (blue line) but also a decrease in fluctuations around the mean. Notice that towards the right where rate is low, fluctuations around the mean are at their smallest.

fluctuations and replace these fluctuations about a straight line, which is placed at the mean of the original series; for figure 3.4 above, this mean is 90. Figure 3.5 shows the corrected series.

We can see that here, in the ideal case where the naive assumptions of the exponential decay fitting approach hold, this approach works quite well. Let us now examine the case where these assumptions don't hold because there are other long-term fluctuations e.g. due to cell movement. We add these other fluctuations as a gentle sinusoid. See figure 3.6.

One can see by eye that this is not a good fit for the data. This has disastrous

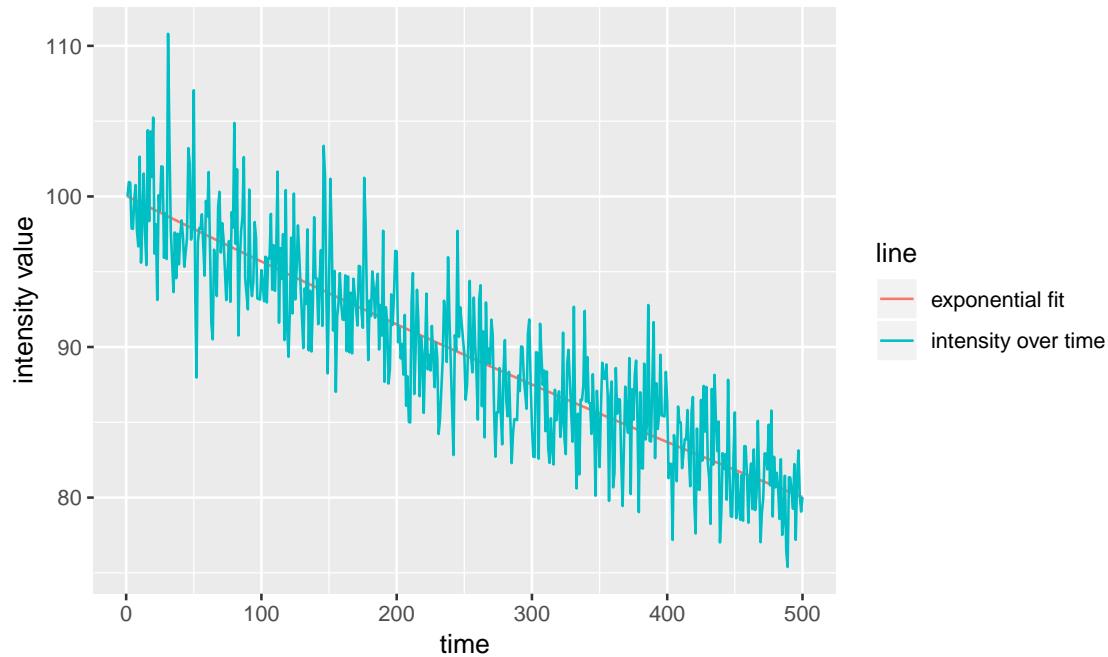


Figure 3.4: Exponential fit to intensity trace of image which is subject to bleaching at a constant rate.

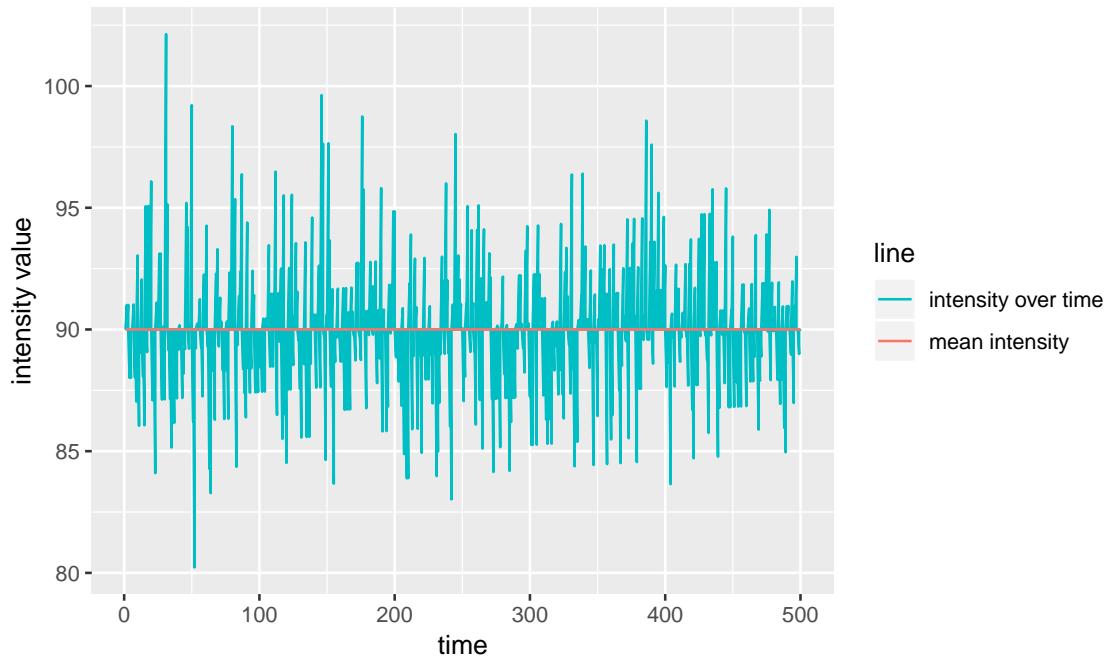


Figure 3.5: The blue line from figure 3.4 has changed to a straight horizontal line cutting the y axis at the mean intensity of the original intensity trace. The fluctuations about the blue line that existed in figure 3.4 are preserved here. As an example, see the large downward fluctuation at $t \approx 50$ seconds in both figures.

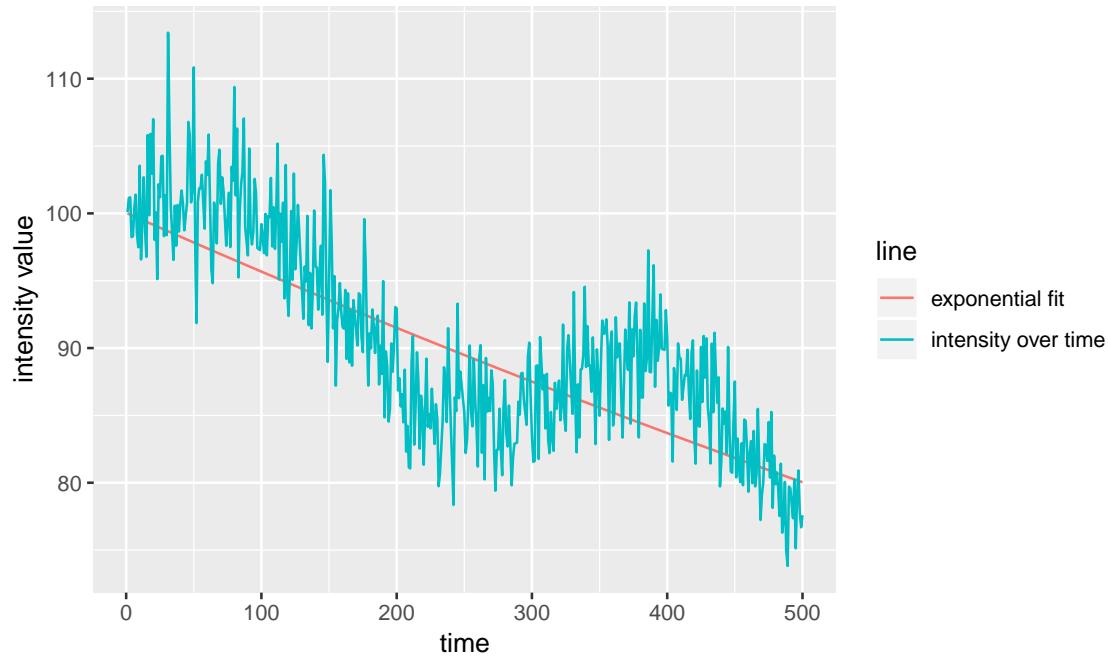


Figure 3.6: An exponential decay with added sinusoidal variance, fit with a simple exponential decay.

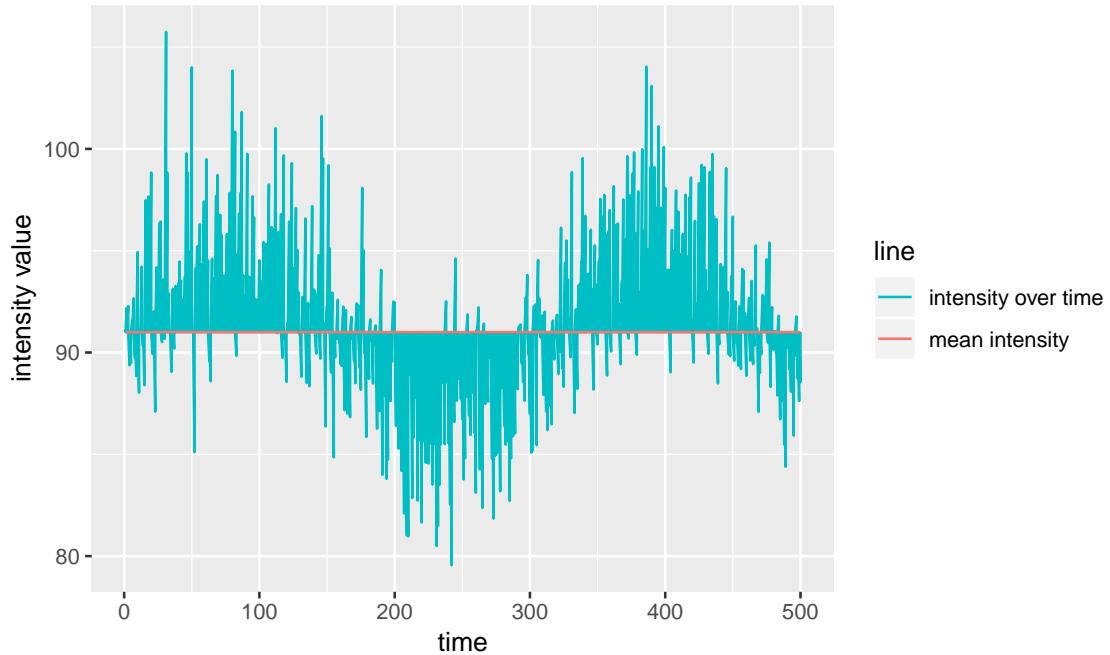


Figure 3.7: Result of exponential fitting detrending applied to a decay with a long-term sinusoidal trend component.

consequences for the detrended series, shown in figure 3.7.

One can see in figure 3.7 that the exponential fit detrend failed to remove the sinusoidal trend in the data (even though it *did* remove the exponential decay component). We have now seen that exponential fitting detrending is appropriate when the decay has a particular form, but is otherwise not fit for use. This is a problem common to all fitting approaches to detrending, even the more flexible types like polynomial detrending (Chan et al., 1977). For this reason alone, for the purpose of detrending, fitting approaches should be avoided.

3.4 Boxcar smoothing detrending

A common approach to obtaining the line from which to measure deviations/fluctuations (as in the red line in figure 3.4) is to *smooth* the time series, i.e. construct the line by taking a *local average* at each point. This is often referred to as *boxcar smoothing* because it can be visualized as drawing a box around a neighborhood of points, taking their average as the smoothed value at that point and then moving the box onto the next series of points and repeating the procedure. See figure 3.8.

The parameter l of a boxcar is such that the length of the boxcar is equal to $2l + 1$. This ensures that the length of the boxcar is always odd) which means it can always be centered upon a point). Hence the allowable lengths of a boxcar are 3, 5, 7, 9, etc.

The boxcar parameter l has a large effect on the type of smoothing achieved. This can be seen in figure 3.9 where boxcar smoothing is applied to the trace in figure 3.4. The traces for $l = 1$ and $l = 3$ are far too wiggly (not smooth enough); the trace for $l = 75$ is better but perhaps still slightly wiggly; finally, the traces for $l = 300$ and $l = 1000$ are too close to straight horizontal lines (too smooth).

This begs the question: what is the correct smoothing parameter l ?

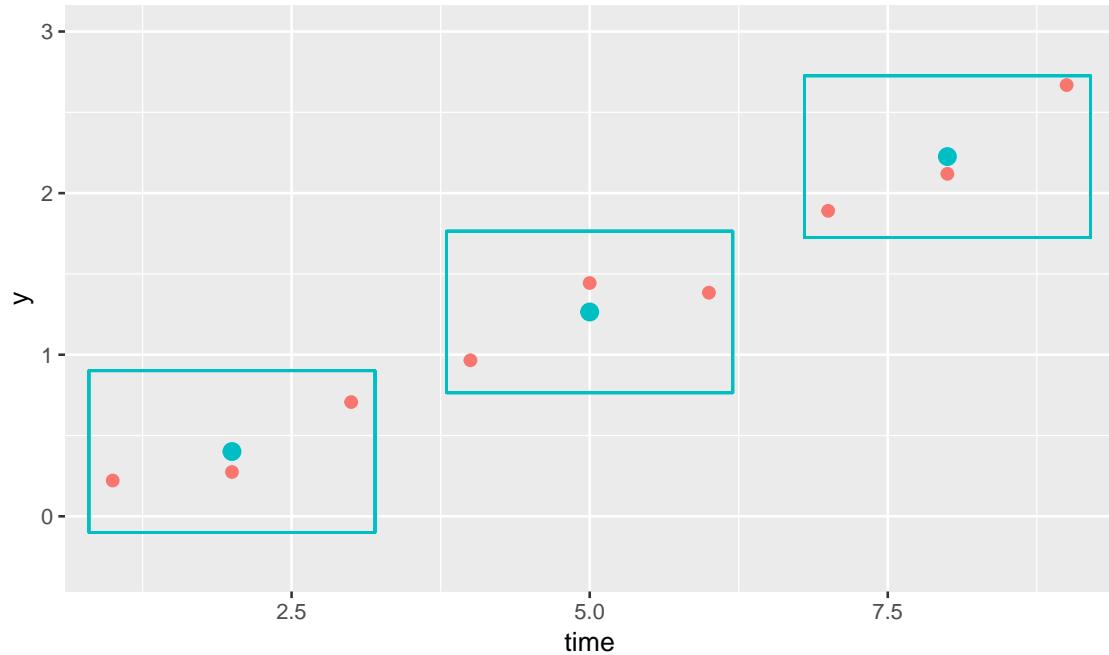


Figure 3.8: The original time series is depicted by the red dots. The blue rectangles represent the *boxcar*. This boxcar is said to be of length 3 because it is wide enough to encompass 3 points at a time. The boxcar is centered on a point and then the smoothed value at that point (blue dot) is calculated as the mean value of all points within the boxcar. In reality, every point gets a smoothed value which means that the boxcar *overlaps* but in this figure, for the sake of clarity, they are not overlapped.

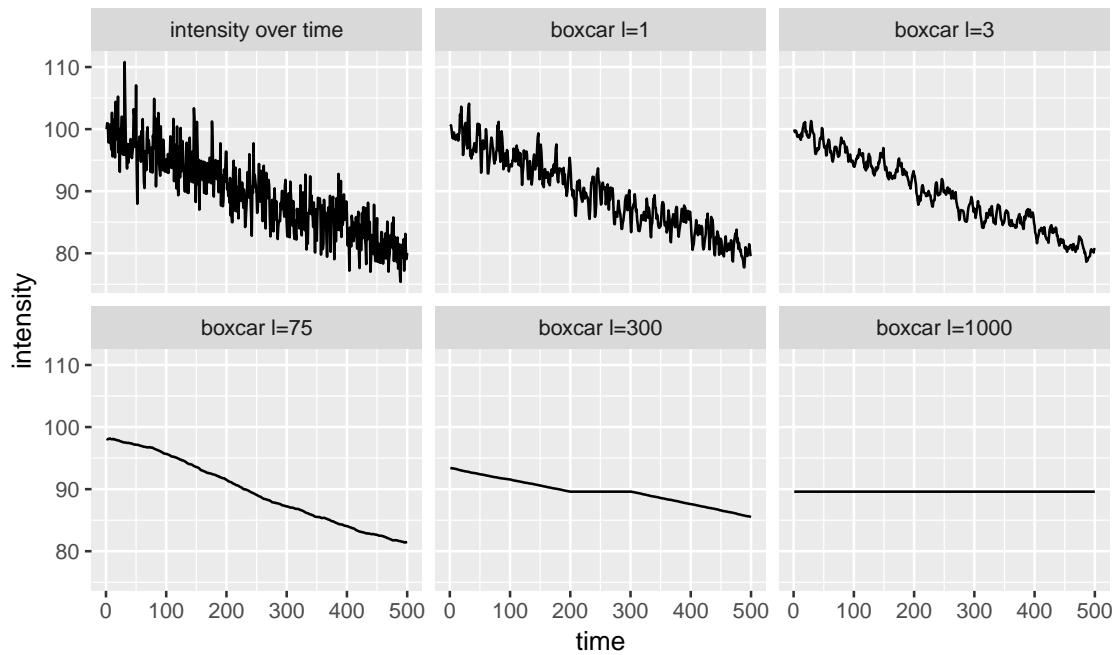


Figure 3.9: The original intensity trace is shown in the top-left. The other panels show the result of boxcar smoothing for $l = 1, 3, 75, 300$ and 1000 .

3.5 Choosing the correct smoothing parameter for detrending

Figure 3.9 shows that the choice of boxcar size is crucial because different sizes lead to very different “smoothed” lines. The most common choice in the community is to choose $l = 10$ (Laboratory for Fluorescence Dynamics, 2018). There is no justification for this choice.

In section 1.6.1, we learned that for immobile particles, the expected brightness is $B = 1$. This fact can be used to solve for the appropriate choice of l to use for detrending a specific image series.

Definition 3.3. The *mean intensity profile* of one channel of an image series is obtained by calculating the mean intensity of each frame in that image series.

The mean intensity profile can be used to visualize the bleaching of an image series. If the fluorophores are bleaching, the mean intensity should be decreasing over time. To proceed with solving for the appropriate l , we need to make one assumption; this is that any two image series with the same mean intensity profile are appropriately detrended with the same detrending parameter l . This assumption seems reasonable to me, however there is no need to debate its validity because later, detrending with the solved-for parameter l will be evaluated with

simulated data and compared to the standard $l = 10$. If this assumption is bad, then the performance of the detrending that relies on it should also be bad. With this assumption in hand, solving for l proceeds as follows:

1. Simulate an image series with immobile particles only which has the same mean intensity profile as the acquired real data.
2. Given that the simulated series is of immobile particles only, once properly detrended, it should have $B = 1$.
3. The l for which the detrended series has mean brightness closes to 1 is the most appropriate for the simulated data.
4. By the assumption above, this l is the most appropriate for the real data.

Mathematically, this can be expressed as

$$l = \operatorname{argmin}_{\tilde{l}} (1 - |\text{mean brightness of simulated series detrended with parameter } \tilde{l}|) \quad (3.1)$$

In fact, what I have done here is to give a general method for solving for any detrending parameter α :

$$\alpha = \operatorname{argmin}_{\tilde{\alpha}} (1 - |\text{mean brightness of simulated series detrended with parameter } \tilde{\alpha}|)$$
(3.2)

This will be useful later when other detrending regimes with their own parameters are introduced.

3.6 Exponential smoothing detrending

Exponential smoothing is a slight alteration to boxcar smoothing. The idea is that when computing a local average, points nearer to the point of interest should have greater weights.¹ The weights fall off with distance $|t|$ from the point of interest according to $\exp(-\frac{|t|}{\tau})$ where the parameter τ is a positive real number. This function is visualized in figure 3.10. For small values of τ , only values very close to the point of interest have importance when calculating the local average. For larger values of τ , further values also have importance (but closer values always have higher weights). In this sense, increasing the value of τ has a similar effect to increasing the value of l for the boxcar in that further away points are taken into account.

¹https://en.wikipedia.org/wiki/Weighted_arithmetic_mean

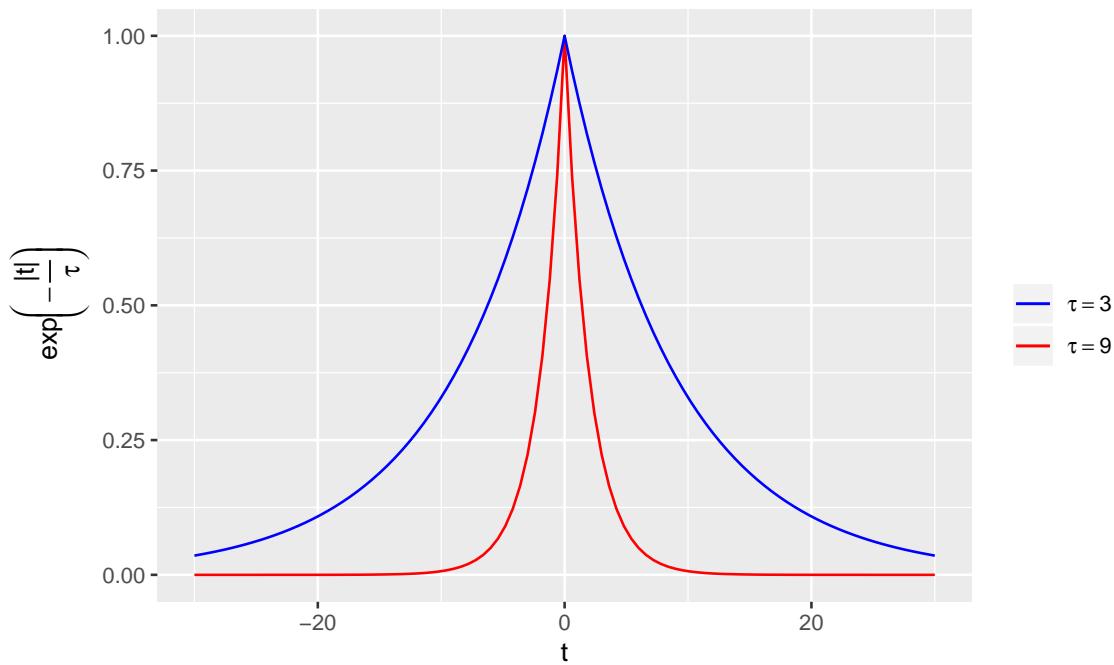


Figure 3.10: The function $\exp(-\frac{|t|}{\tau})$ visualized with $\tau = 3$ and $\tau = 9$. For $\tau = 3$, points at distance $|t| = 10$ have approximately zero weight, whereas for $\tau = 9$, these points have significant weight.

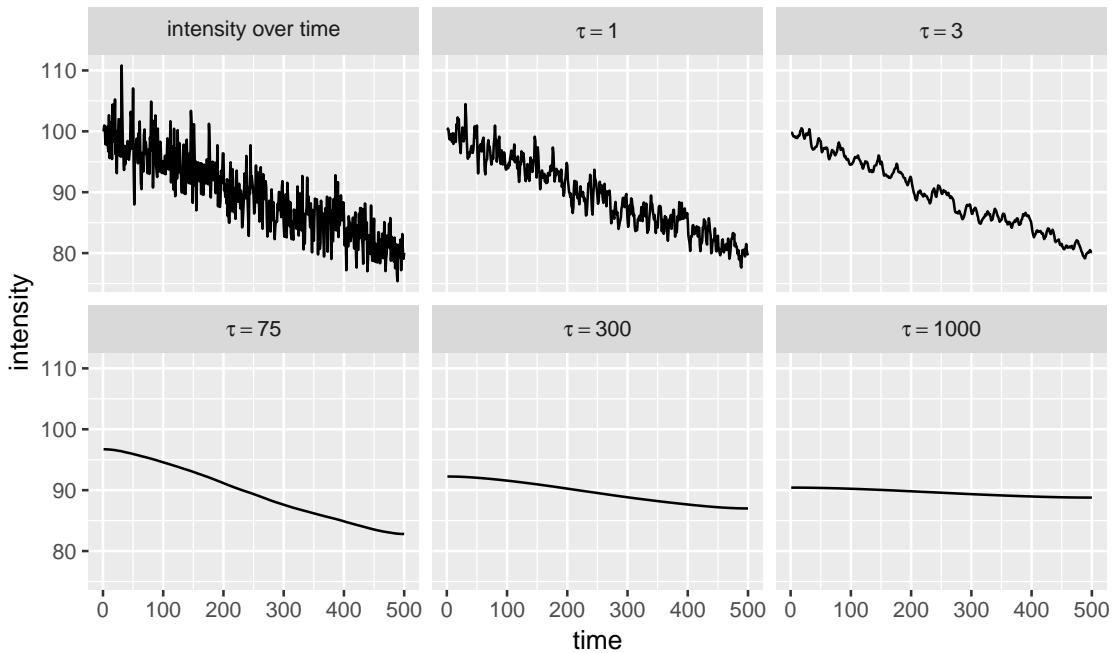


Figure 3.11: The original intensity trace is shown in the top-left. The other panels show the result of exponential smoothing for $\tau = 1, 3, 75, 300$ and 1000 .

In figure 3.11, exponential smoothing with different parameters τ is applied to the trace in figure 3.4. The results are similar to those in figure 3.9.

Heuristically, exponential smoothing detrending seems favorable to boxcar detrending because the idea that points further away from the point of interest are less important (but still somewhat important) when computing the local average is reasonable. Indeed, this was the method proposed in the original number and brightness paper (Digman et al., 2008). For this reason, exponential smoothing was the method of choice for my paper where the method of choosing the correct detrending parameter was published (Nolan et al., 2017a).

3.7 Correcting for non-stationary variance

Definition 3.4. The variance of a random variable X is the expected value of the squared deviation of X from its mean μ :

$$\text{Var}(X) = E[(X - \mu)^2] \quad (3.3)$$

All of this chapter so far has focused on correcting for non-stationary mean. As shown in figure 3.3, as the mean decreases, so too does the variance. For an instance x of the random variable X with expected value $E[X] = \mu$, $x - \mu$ is the *deviation* of x from μ . If we write x as $x = \mu + \tilde{x}$, then we get the deviation $x - \mu = (\mu + \tilde{x}) - \mu = \tilde{x}$, so \tilde{x} is the deviation. For a given point in figure 3.3, its deviation is its distance from the red line. For positive real number k , making the transformation $\tilde{x} \rightarrow \sqrt{k}\tilde{x}$ i.e. $x \rightarrow \mu + \sqrt{k}\tilde{x}$ causes the variance (i.e. the *squared deviation*) to be transformed as $\text{Var}(X) \rightarrow k \times \text{Var}(X)$. Hence, we have a way to modify the variance of a time series as a whole by modifying the deviation of each time point from the mean. For months, I toyed with this idea as a solution of correcting for non-stationary variance. However, in reality the contribution to the variance in intensity at a given pixel is down to both Poisson photon statistics and fluorophore movement. This combination of factors makes it very difficult

to ascertain the amount by which the variance should be altered. I eventually abandoned my efforts to alter the variance like this in favor of the *Robin Hood* detrending algorithm (section 3.9) which includes correction for non-stationary variance as an intrinsic part of its detrending routine.

3.8 Caveats of fitting and smoothing approaches to detrending

Both fitting and smoothing approaches to detrending have serious caveats. Fitting approaches assume that the fluorescence intensity decay has a certain form. Unpredictable issues such as cell movement mean that no particular decay form can be assumed. Smoothing methods do not perform well at the edges of time series that they are applied to. They also require the user to choose a smoothing parameter. The problem of how to best choose this parameter was solved recently (Nolan et al., 2017a), but this method has not been widely adopted. Most importantly, both fitting and smoothing fail when the data cannot be approximated as continuous (fitted and smoothed lines are continuous approximations of data). Fluorescence intensity data at low intensities—where most pixel values are either 0 or 1—are quasi-binary² and hence a continuous approximation does not make

²By *quasi-binary*, I just mean that almost all values are 0 or 1.

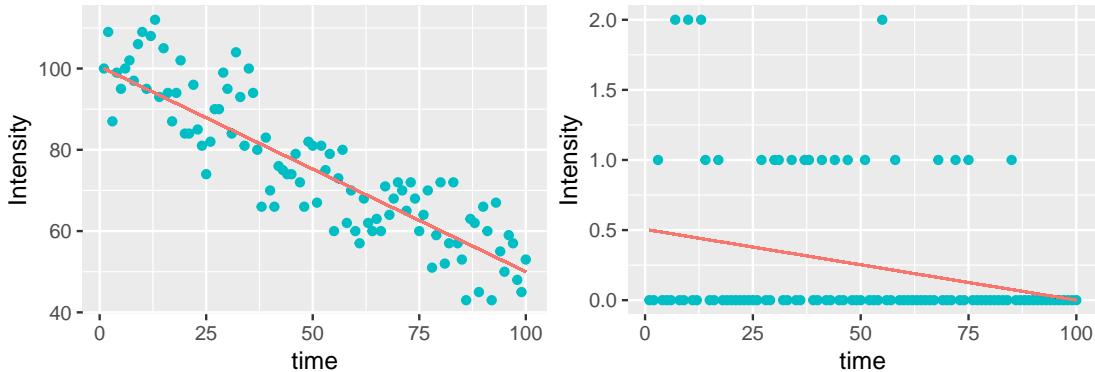


Figure 3.12: Left: for high ($\gg 1$) intensity values, the line is a satisfactory approximation of the data, representing it well. Right: for low (quasi-binary) intensity values, the line is not a good approximation for the data and indeed no line or curve could represent the data well.

sense (see figure 3.12). This means that neither fitting nor smoothing are applicable detrending methods at low intensities. This is the crucial caveat of these methods because, when bleaching is a problem, it is common to reduce laser power to reduce bleaching, which leads directly to lower intensity images. With fitting and smoothing techniques, it may sometimes be advisable to increase the laser power to achieve higher intensities such that the detrending routines will function properly. This means one may need to bleach more in order to be able to correct for bleaching. This farcical situation necessitates a new detrending technique which can function at low intensities.

3.9 Robin Hood detrending

Intensity images in units of photons are count data. This means that the values are all natural numbers, i.e. elements of $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$. Fitting and smoothing give real-numbered values (elements of \mathbb{R}), which must then be transformed back into count data (elements of \mathbb{N}_0), normally by rounding. This means that fitting and smoothing methods of detrending push values through $\mathbb{N}_0 \rightarrow \mathbb{R} \rightarrow \mathbb{N}_0$. When current methods were failing to properly detrend low-intensity images, I began to wonder was it necessary to go through the real numbers \mathbb{R} , given that the start and end points were the natural numbers \mathbb{N}_0 ?

Consider figure 3.13. There is a bleached and unbleached version of an intensity trace. Suppose that our real data is the bleached trace, but we *wish* it looked like the unbleached trace. You may wonder why the unbleached trace is not at the starting intensity of the bleached series. For reasons that will become clear, the Robin Hood algorithm can only place the detrended image at the mean intensity of the original image. This is not a problem because the issue with bleaching in FCS and FFS is mainly that the changing signal leads to incorrect calculations, not that the loss in signal leads to a chronic lack of information (photons). Indeed a feature of the Robin Hood algorithm is that it preserves the mean intensity of the real data on a pixel-by pixel basis.

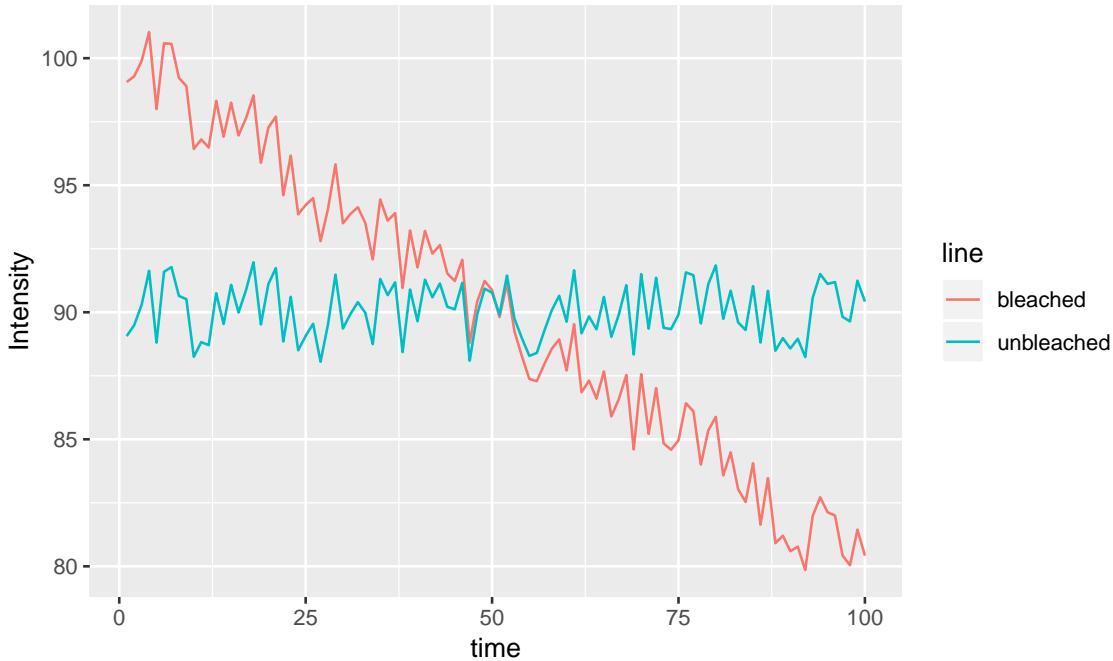


Figure 3.13: Bleached and unbleached intensity traces.

To get to the unbleached intensity trace from the bleached intensity trace, intensity must be subtracted from time-points with too much intensity and added to time points with too little intensity. This can be done by *taking* counts from frames with too much intensity and *giving* them to frames with too little intensity. In this way, no counts are gained or lost, they are just moved around the image series. See figure 3.14. Counts are passed from one frame to another *along* a given pixel, i.e. if a count is taken from pixel at position p in some frame i , it must be given to a pixel at the same position p in some other frame j . It is this condition that ensures that the mean intensity images of the original and detrended image series are the same.

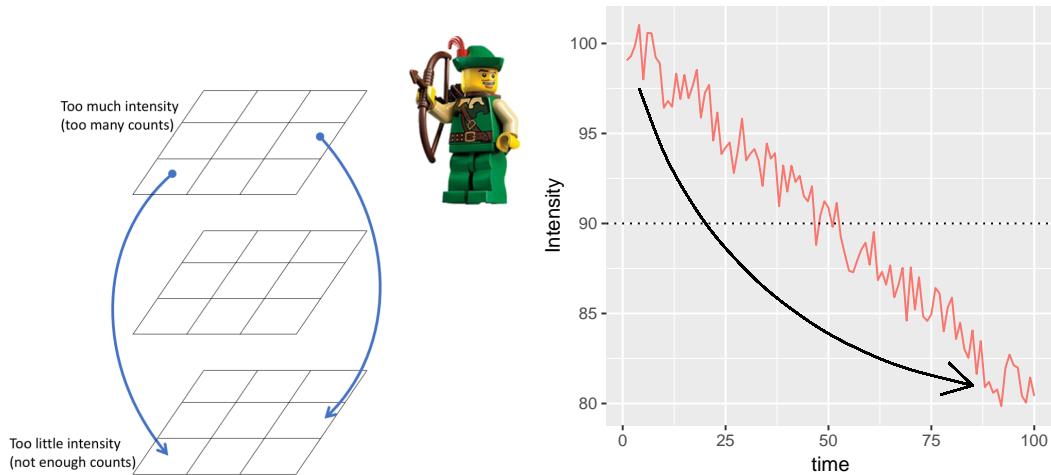


Figure 3.14: Robin Hood: counts are taken from frames of higher intensity (usually closer to the start of the image series) and given to frames of lower intensity (usually closer to the end of the image series).

To determine how many swaps need to be made to detrend a given image series, equation (3.2) can be used, with α being the number of swaps.

The random gifting of counts from higher to lower intensity frames has the effect of redistributing mean intensity but *also* variance in intensity. With photon statistics (which follow a Poisson distribution), random counts provide both mean and variance. This is in contrast to all previous methods which consist of determining local deviation and adding it to a *fixed* global mean: this provides no redistribution of variance.

3.10 A comparison of detrending methods

To compare the various detrending methods, I use the following workflow:

1. Simulate a number $N = 100,000$ of particles diffusing with known diffusion rate. Simulations were done with the `brownded` software package (section [2.3.6](#)).
2. Simulate photon emission from these particles with chosen brightness ϵ and create an image series from this, being careful to (virtually) sample at a rate appropriate for number and brightness analysis.
3. Bleach the simulation with a chosen constant bleaching rate.
4. Simulate photon emission from the bleached simulation (bleached particles don't emit photons) with the same brightness ϵ and create an image series.
5. Detrend the bleached image series.
6. Evaluate the detrending algorithm by measuring how close the brightness of the detrended bleached image series is to the known simulated brightness.

For all combinations of brightnesses of $\epsilon = 0.001, 0.01, 0.1, 1, 10$ and bleaching fractions of 0%, 1%, 5%, 10%, 15%, 20%, 25%, 30%, 20 images of 64x64 pixels and 5,000 frames were simulated using 100,000 fluorescent diffusing particles.³

These were detrended with the following detrending routines:⁴

³The simulation took 3 weeks.

⁴The detrending took 2 weeks.

1. Boxcar with $l = 10$ (`boxcar10`, the most common detrending routine).
2. Exponential smoothing with automatically chosen parameter τ (`autotau`).
3. Robin Hood with automatically chosen swaps (`robinhood`).

The performance was evaluated using the *mean relative error*.

Definition 3.5. For a given brightness and bleaching fraction,

$$\text{mean relative error} = \frac{|(\text{calculated brightness after detrending}) - (\text{true brightness})|}{(\text{true brightness})} \quad (3.4)$$

Figure 3.15 shows the results. Before I discuss them, note that the common brightnesses that we see are in the range $\epsilon = 0.003$ to $\epsilon = 0.1$.

The most striking thing about figure 3.15 is that the best choice in all cases is to not detrend at all! This is an interesting result and seems to render all detrending routines worthless. However, when working with real data, not detrending does not work well at all. This will be shown in chapter 4. This is probably because with real data, bleaching is likely not taking place at a constant, steady rate and other factors such as cell movement are contributing to medium and long term intensity fluctuations and these have a detrimental effect on calculations if not detrended out. It would be possible to study this by mimicking real bleaching

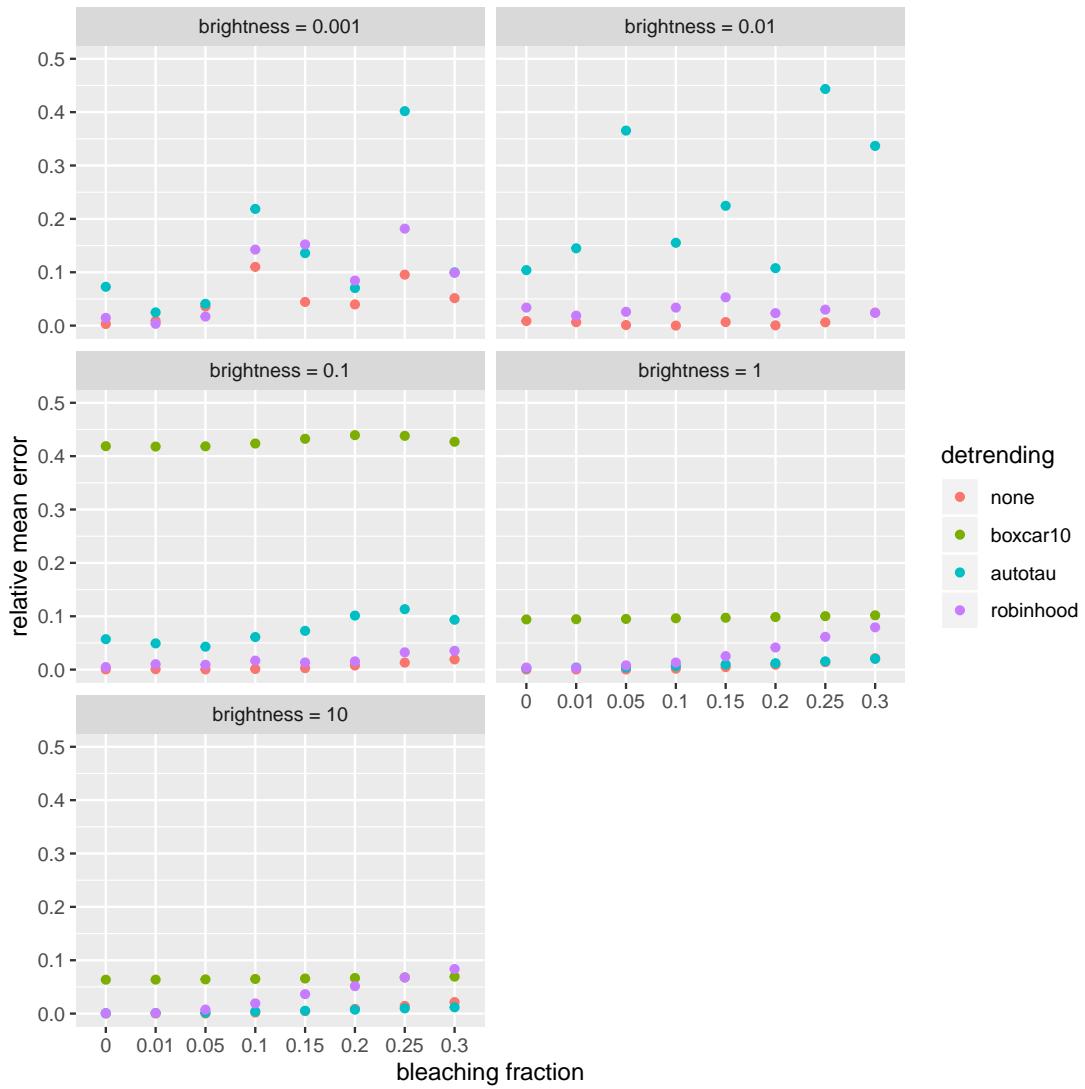


Figure 3.15: A comparison of different detrending methods with various brightnesses and bleaching fractions (steady, constant-rate bleaching), including the results of not detrending at all.

profiles with simulations (see section 6.3).

The worst performer by far is `boxcar10`. For example, at $\epsilon = 0.1$, it makes an error of worse than 40% and for $\epsilon = 0.001, 0.01$, its error is worse than 50%, so it does not even appear on the plot. This is good evidence that arbitrarily choosing the parameter l very bad practice. For realistic brightnesses (≤ 0.1), `robinhood` is the best with errors almost always lower than 5%. `autotau` also performs very well, with errors almost always less than 10%. At the lowest brightness $\epsilon = 0.001$, all methods are somewhat erratic. That is because at this extremely low brightness, there is a critical lack of information (photons) for the algorithms to work with. Finally, at unrealistically high brightnesses of $\epsilon = 1, 10$, `autotau` begins to perform well because at these high photon counts, the caveats of smoothing have totally disappeared. However, I cannot explain the degradation in the performance of `robinhood` in this case. Fortunately, there is no need to dwell on this, as this situation ($\epsilon = 1, 10$) does not arise in practice because available fluorophores are not this bright.

Chapter 4

Applications

4.1 Dimerization of FKBP12

4.1.1 Introduction

Myristoylated FKBP12 is known to dimerize upon addition of the drug AP1510 (Amara et al., 1997). As a test application of exponential smoothing detrending (section 3.6) with automatically chosen parameter τ , we used this system with number and brightness to verify this dimerization. We tested this in 20 Cos7 cells with mClover-labelled FKBP12.

4.1.2 Experimental results and discussion

We found a brightness increase in ϵ of ≈ 1.6 -fold using the *automatic* detrending method. The choice of $\tau = 10$ resulted in a ≈ 0.7 -fold calculated increase, which is indeed a decrease. The ≈ 1.6 -fold increase suggests that dimerization had occurred (see figure 4.1), however, we expected the increase to be ≈ 2 -fold upon dimerization. In that publication (Nolan et al., 2017a), we postulated that the 1.6 figure was due to the fact that not all of the protein had dimerized. Recently, a paper came out (Dunsing et al., 2018) explaining how the assumption that all fluorophores emit signal is invalid and that because of this, oligomeric state changes calculated from brightness must be adjusted by a correction factor specific to the fluorophore. Unfortunately, this study did not characterize mClover, so we do not know its correction factor. I suspect that applying this correction would bring our figure of 1.6 a lot closer to 2. We also tried $\tau = 10$, which gave a decrease in brightness, showing that arbitrary parameter choices lead to unpredictable and unreliable results and should hence be avoided.

The heterogeneity of fold-changes in ϵ measured from single images as shown in figure 4.1 shows that many replicates are needed in number and brightness experiments in order to converge upon the true value of the fold change.

A pair of cells from this study together with their brightness statistics are shown

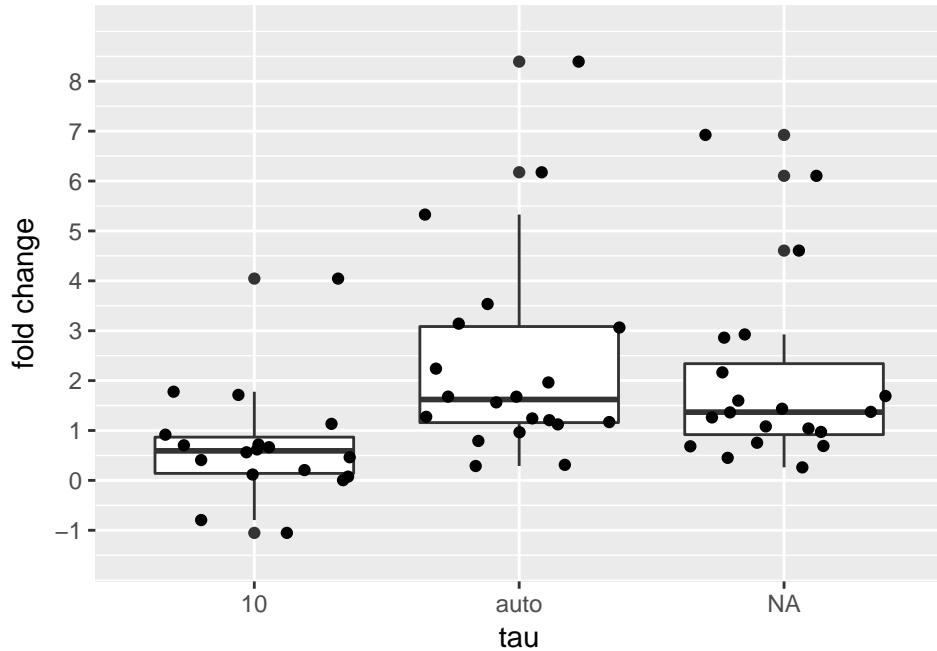


Figure 4.1: The fold changes in brightness ϵ upon addition of AP1510 drug shown for different detrending routines. $\tau = \text{NA}$ is no detrend.

in figure 4.2. Notice how there is no discernible change in intensity before and after addition of the drug, but there is a discernible change in brightness B , best seen using the histogram of pixel brightnesses.

4.1.3 Visualization of bleaching correction on real data

That publication also includes a visualization of bleaching correction on real data. Figure 4.3 shows this, comparing the choice of $\tau = 10$ with the automatic detrending algorithm, which chooses an appropriate τ based on the data. It is evident that both $\tau = 10$ and the automatic choice of τ give the corrected intensity profile

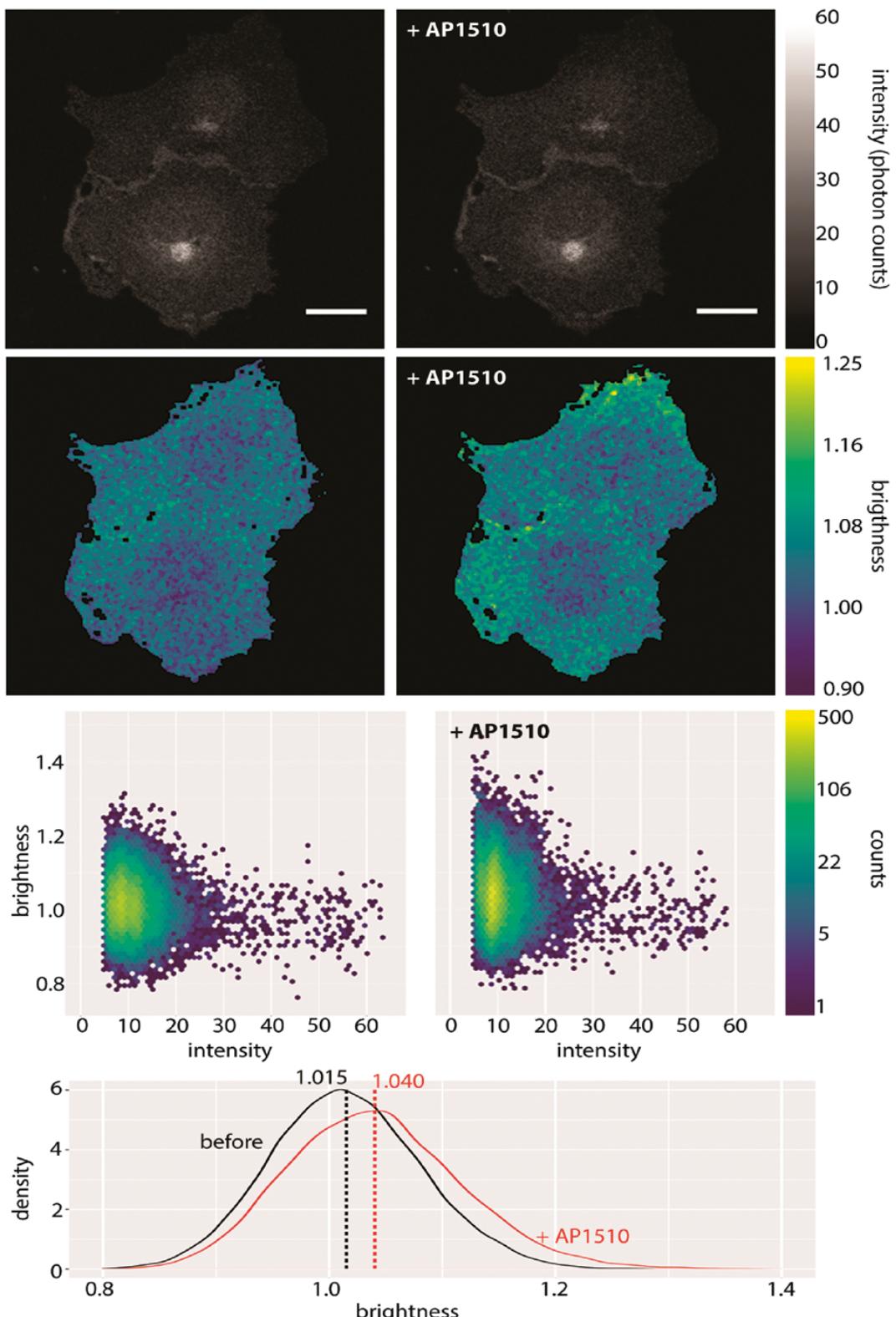


Figure 4.2: mClover-labelled myristoylated FKBP12 before and after application of 50nM AP1510. Shown here are intensity (first row), brightness (second row), a plot of intensity versus brightness (third row) and brightness histograms (fourth row). Notice how the change in brightness upon addition of the drug is seen most clearly by comparing the brightness histograms. The vertical lines in the histogram plot show the means of those histograms. Brightness here refers to B.

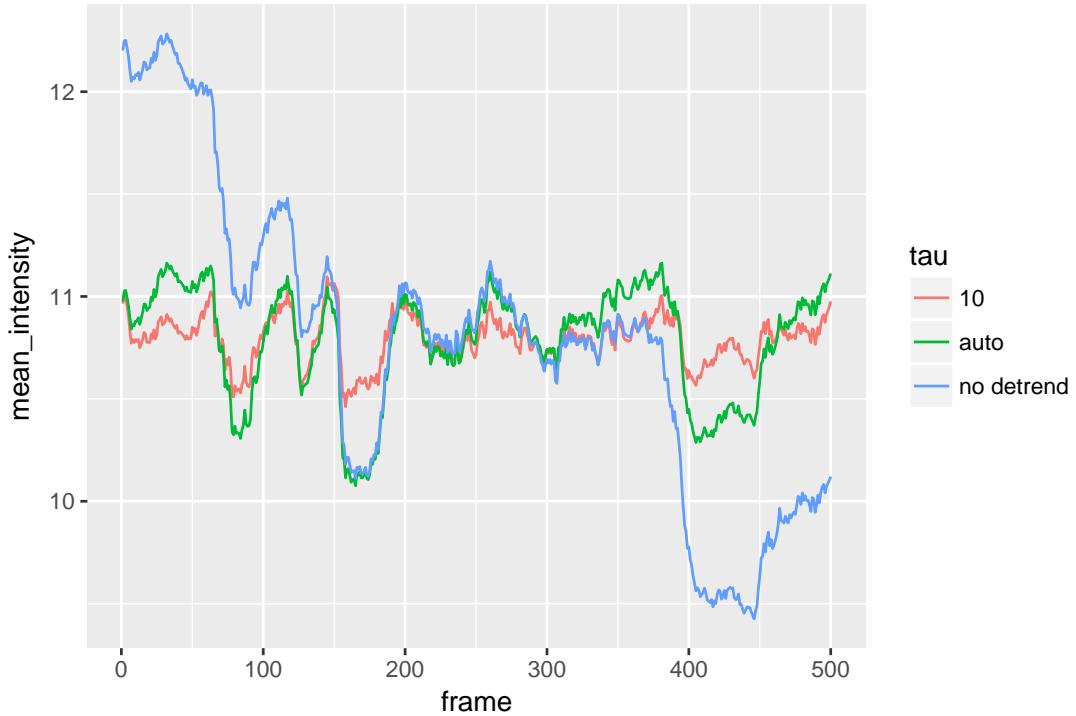


Figure 4.3: Real data image series mean intensity profile with bleaching correction with $\tau = 10$ and auto τ .

a stationary mean (both the red and green series start and end at ≈ 11), however the $\tau = 10$ correction (red line) also has a much decreased variance compared to the auto τ line (green), which is bad; the $\tau = 10$ line is removing local variation, which is exactly what we're trying to avoid.

4.2 *In vitro* number and brightness

4.2.1 Introduction

In our research group, we believe that the most practical quantitative method for measuring homo-dimerization *in vivo* and *in vitro* is N&B (Digman et al., 2008) because it is calibration-free and does not require specialized instrumentation. There are many examples of the application of N&B *in vivo* (the original N&B paper has over 250 citations, most of which are *in vivo* applications) but none *in vitro*. Hence, we published a protocol (Nolan et al., 2018) detailing how N&B can be applied *in vitro*. This time, we used FKBP12F36V which is an FKBP mutant with a new dimerizing drug AP20187 (known colloquially as the *BB dimerizer*); this pair is designed to have better specificity than the original (Clackson et al., 1998).

4.2.2 Experimental results and discussion

In this experiment, the FKBP12F36V was labelled with mVenus. We found that the brightness doubled from $\epsilon = 0.005$ to $\epsilon = 0.010$ upon addition of the drug. See figure 4.4. This analysis was done with exponential smoothing detrending with automatically chosen parameter τ . Without detrending, the pre-BB brightness

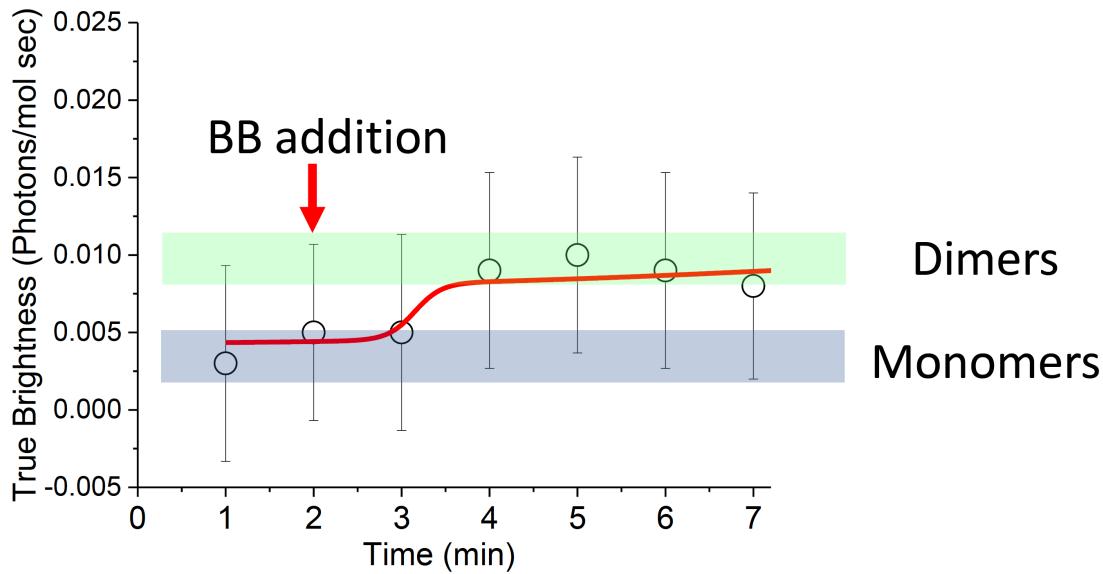


Figure 4.4: Dimerization of FKBP12F36V upon BB addition is seen by a brightness doubling from $\epsilon = 0.005$ to $\epsilon = 0.010$ over a period of minutes.

was calculated as $\epsilon = 0.026$, showing that detrending is absolutely necessary and that neglecting this step can lead to nonsensical results.

Remark. This paper also included an important correction to the equation for brightness ϵ when analog equipment is used. The correct equation is

$$\epsilon = \frac{\sigma^2 - \sigma_0^2}{S(\langle I \rangle - \text{offset})} \quad (4.1)$$

The S in the denominator was omitted in the original paper (Dalal et al., 2008) and this error was reproduced in our N&B review (Nolan et al., 2017b).

4.3 HIV-1 receptor stoichiometry

4.3.1 Introduction

4.3.1.1 HIV-1 cell entry

HIV-1 infects many cell types (e.g. CD4 T cells, macrophages, dendritic cells) and has different modes of entry for each cell type and indeed possibly more than one mode of entry in any given cell type (Jakobsdottir et al., 2017). Endocytosis is thought to be a common entry route (Miyauchi et al., 2009), particularly in macrophages (Marechal et al., 2001) and Dendritic cells (Ménager and Littman, 2016). In CD4 T cells, HIV-1 has been shown to fuse at the plasma membrane without needing endocytosis (Herold et al., 2014).

Entry of HIV into any cell involves the initial binding of the CD4 receptor on that cell by the HIV-1 virus. Subsequently, a co-receptor (often CCR5 or CXCR4) is used in the fusion process (Jakobsdottir et al., 2017). The question of how many receptors and co-receptors are required to facilitate fusion (the *stoichiometry* of the interaction of HIV-1 with its receptor and co-receptor, possibly different for different cell types) had not been answered.

4.3.1.2 The use of number and brightness to study HIV-1 cell entry

Our main motivation for studying N&B in the first place was that we thought it was a valuable method to study the process of HIV-1 fusion in live cells. N&B was first used in our research group to study the oligomeric state of dynamin at the HIV-1 fusion pore in TZM-bl cells (Jones et al., 2017). This study concluded that dynamin-2 stabilizes the HIV-1 fusion pore with a low oligomeric state.

Following on from this, we wanted to study the stoichiometry of the interaction of HIV-1 with its receptor (CD4) and co-receptor (CCR5 or CXCR4) upon the engagement of the virus with the cell and to follow this interaction stoichiometry up to the point of fusion. See figure 4.5.

Entry of HIV-1 into a host cell requires an initial interaction between the viral-envelope glycoprotein spike complex—Env—with cell surface displayed CD4 and co receptors (Jakobsdottir et al., 2017). Although structural studies have revealed the intra-molecular basis for CD4 receptor and CXCR4/CCR5 co-receptor-induced conformational changes to the HIV-1 Env during host cell entry (Ozorowski et al., 2017), little is known about how the inter-molecular dynamics and stoichiometry of this process culminates in fusion with the host cell membrane in live cells (Brandenberg et al., 2015). This is due to the

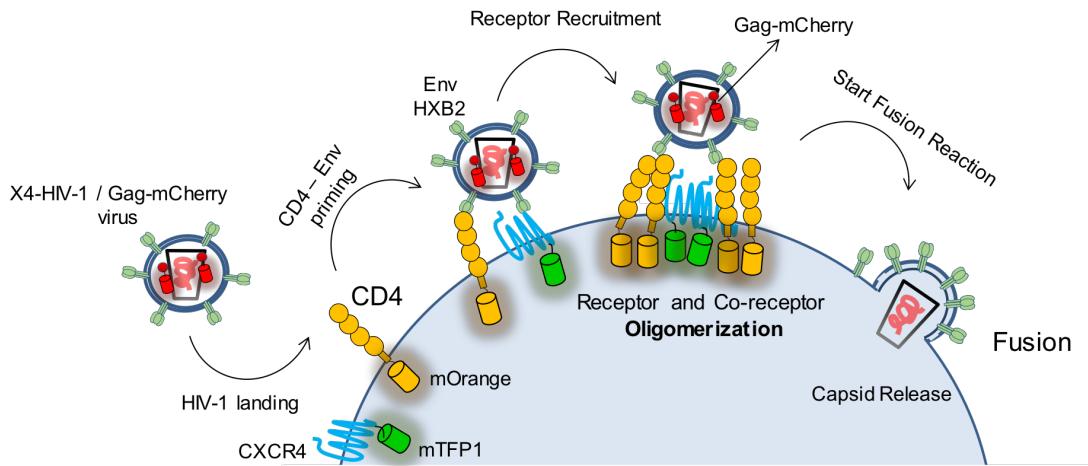


Figure 4.5: The HIV-1 envelope glycoprotein Env must bind the receptor (CD4) and form a complex with the co-receptor (CXCR4 or CCR5, this figure shows an X4-tropic virus and co-receptor) to initiate the fusion process. Labeling the viral Gag protein with mCherry, the receptor with mTFP1 and the co-receptor with mOrange, it is possible to follow these three players in the fusion reaction and to quantify their interaction.

difficulty of working with live cells and the lack of temporal resolution of the techniques commonly employed (i.e. crystallography and cryo-EM).

— Iliopoulou et al. (2018)

We saw N&B as the ideal technique to probe this stoichiometry temporally. With our microscope, we could acquire 100 frames per 1.7 minutes, therefore, using each consecutive sequence of 100 frames to create a brightness image, we could obtain 1 brightness image every 1.7 minutes and use this to calculate this temporal stoichiometry.

4.3.2 Experimental setup

Receptor (CD4) and co-receptor (CXCR4 or CCR5) were labelled in Cos7 cells. Virus was added at time $t = 0$ and imaging proceeded for a number of minutes at 100 frames per 1.7 minutes. Alternating laser excitation (ALEX, Kapanidis et al. (2005)) was used to eliminate the possibility of channel bleed-through. See figure 4.6.

4.3.3 Analysis

The virus channel was used to locate the virus at a given point in time. The receptor and co-receptor were used to calculate brightness and cross-correlated brightness every 100 frames (every 1.7 min). The brightness was used to determine the number of receptor and co-receptor units involved in a complex. The cross-correlated brightness was used to delimit whether or not the receptor and co-receptor units were together in the *same* complex. See figure 4.7, this is the corresponding brightness and cross-correlated brightness image of figure 4.6. Notice that once the virus lands, the oligomeric state of the receptor and co-receptor increases. We also see significant positive cross-correlated brightness in this area, indicating that the virus has triggered a complex of receptor and co-receptor.

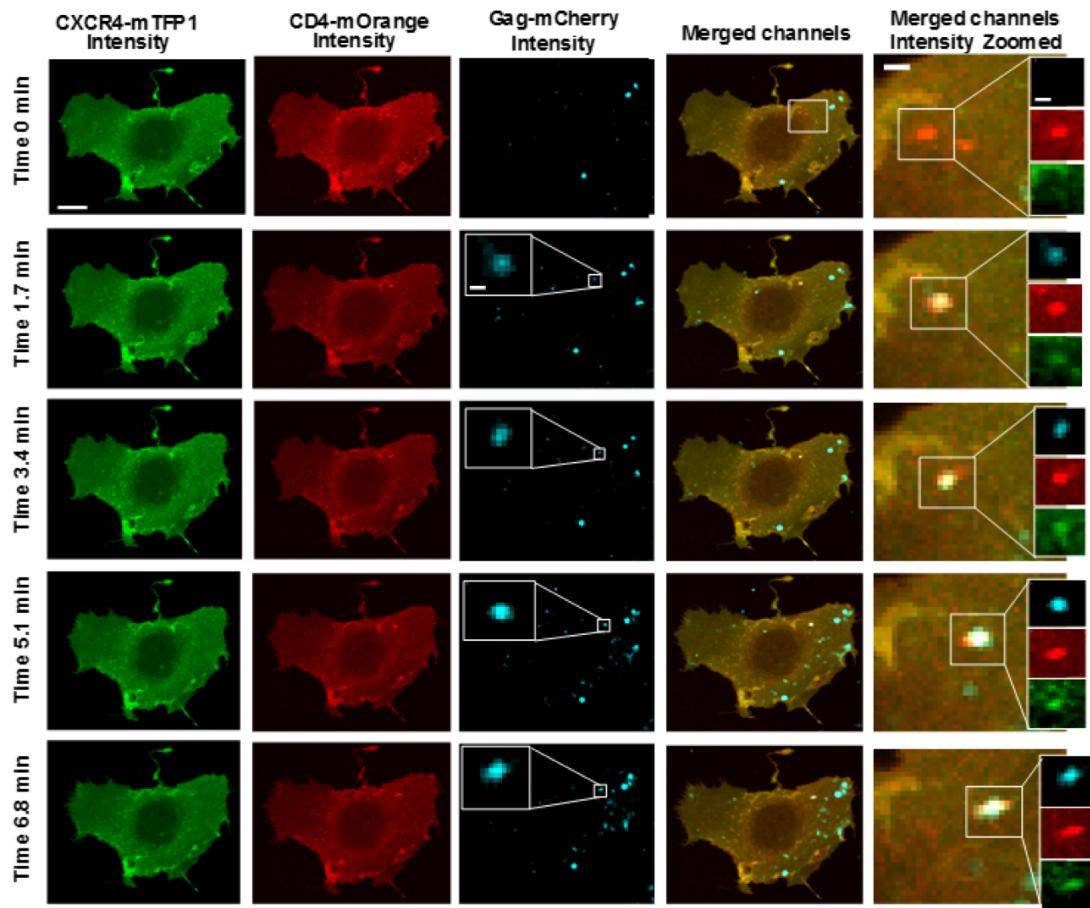


Figure 4.6: Intensity images from the virus, receptor and co-receptor. Every 100th frame is shown. A virus which lands at $t \approx 1.7$ minutes is highlighted.

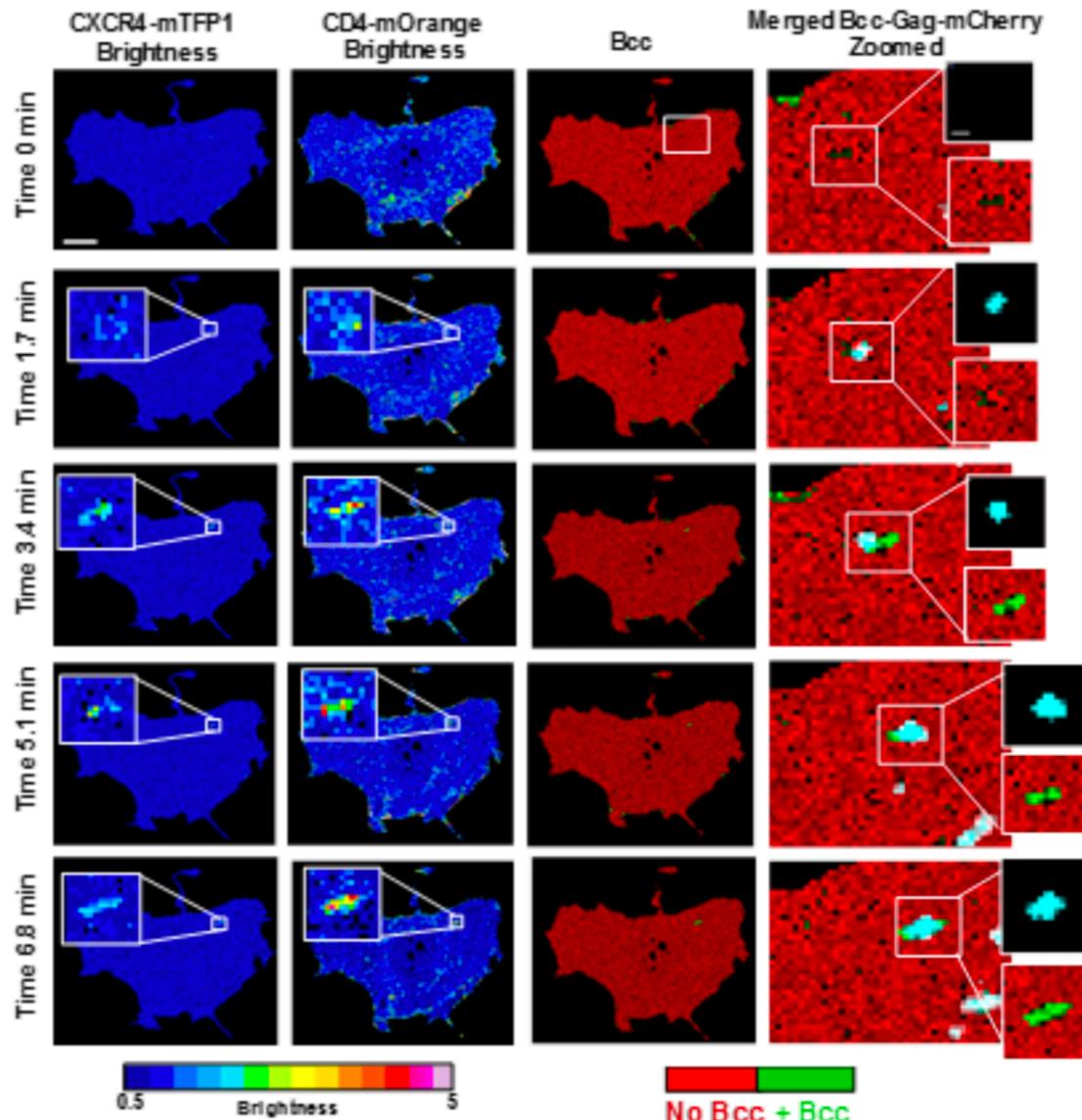


Figure 4.7: Brightness images of receptor and co-receptor and cross-correlated brightness image of the interaction between the two.

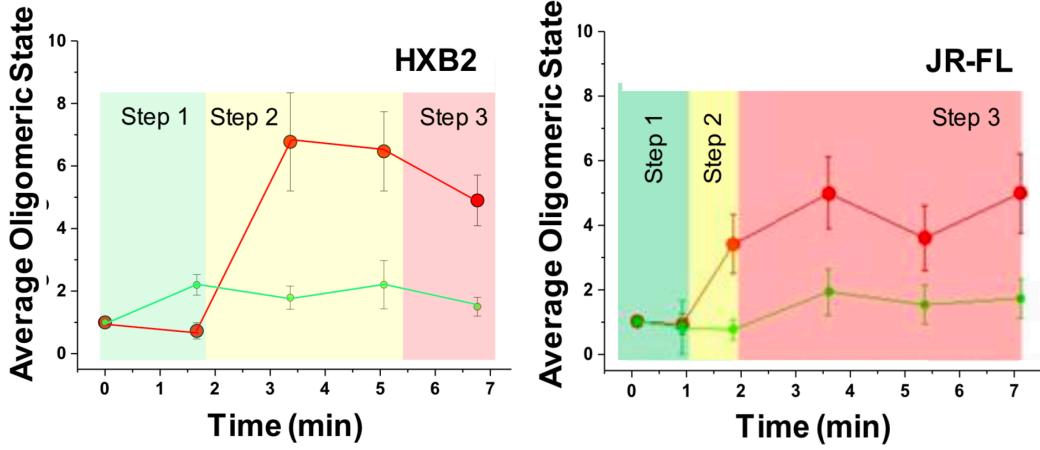


Figure 4.8: Number of receptor and co-receptor units involved in complexes with virus over time, obtained by brightness analysis. Left panel: HIV_{HXB2}. Right panel: HIV_{JR-FL}.

4.3.4 Results

Figure 4.8 shows the results of the analysis detailed in figures 4.6 and 4.7 for $n = 10$ cases where virus triggered receptor and co-receptor complexes in the X4-tropic setting and $n = 12$ in the R5-tropic setting. A three-step pre-fusion process is hypothesized for each.

Our studies support a dynamic three step model for both HIV_{HXB2} and HIV_{JR-FL} (figure 4.9). For X4 tropic virions, Env – CD4 interactions induce CXCR4 dimerization, CD4 then engages with two Env (shown by 3 color TIRF-dSTORM microscopy) to generate a hexamer that might serve as a scaffold to stabilise a final 4 CD4 – 1/2 CXCR4 conformation, with a single Env. We speculate that for HIV_{HXB2}, step 2 is

crucial to culminate the fusion reaction and there could be an anchoring domain and a fusion domain that undergoes gp120 disassembly leading to 6 helix bundle formation. For R5 tropic virions, Env – CD4 interactions form a the previously described asymmetric pre-hairpin intermediate (Munro et al., 2014) (Kwon et al., 2015) (Ma et al., 2018); following binding and oligomerisation of 2 additional CD4 molecules with concomitant CCR5 dimerization. After this, the secondary intermediate leads to the final fusion competent complex with a total of 4 ± 0.3 CD4, 2 ± 0.3 CCR5 and 1 JR-FL Env.

Our data indicate that both HXB2 Env and JR-FL Env start with an asymmetric intermediate bound to a single CD4, as previously suggested²⁷. Our models also support the existence of important differences in the entry mechanisms of X4 and R5 strains. In the X4 strains, CXCR4 dimerization (Tan et al., 2013) (Qin et al., 2015) occurs prior to CD4 hexamer formation and following initial Env – CD4 recognition (Liu et al., 2017). For R5 tropic JR-FL, CCR5 dimerization (Qin et al., 2015) occurs after Env-CD4 complexation and recruitment of two additional CD4 molecules (Wu et al., 1996) around the complex.

For the X4 and R5-tropic case, structural modelling of the hypothesis has been done. See figure 4.9.

4.3.5 Conclusion

Time-resolved N&B enabled us to answer questions about the interaction between the HIV-1 virus and its receptor and co-receptor in live cells which up to now could not be answered. Being able to correctly correct for bleaching is crucial for reliable N&B analysis, so the development of these new algorithms was crucial to this study.

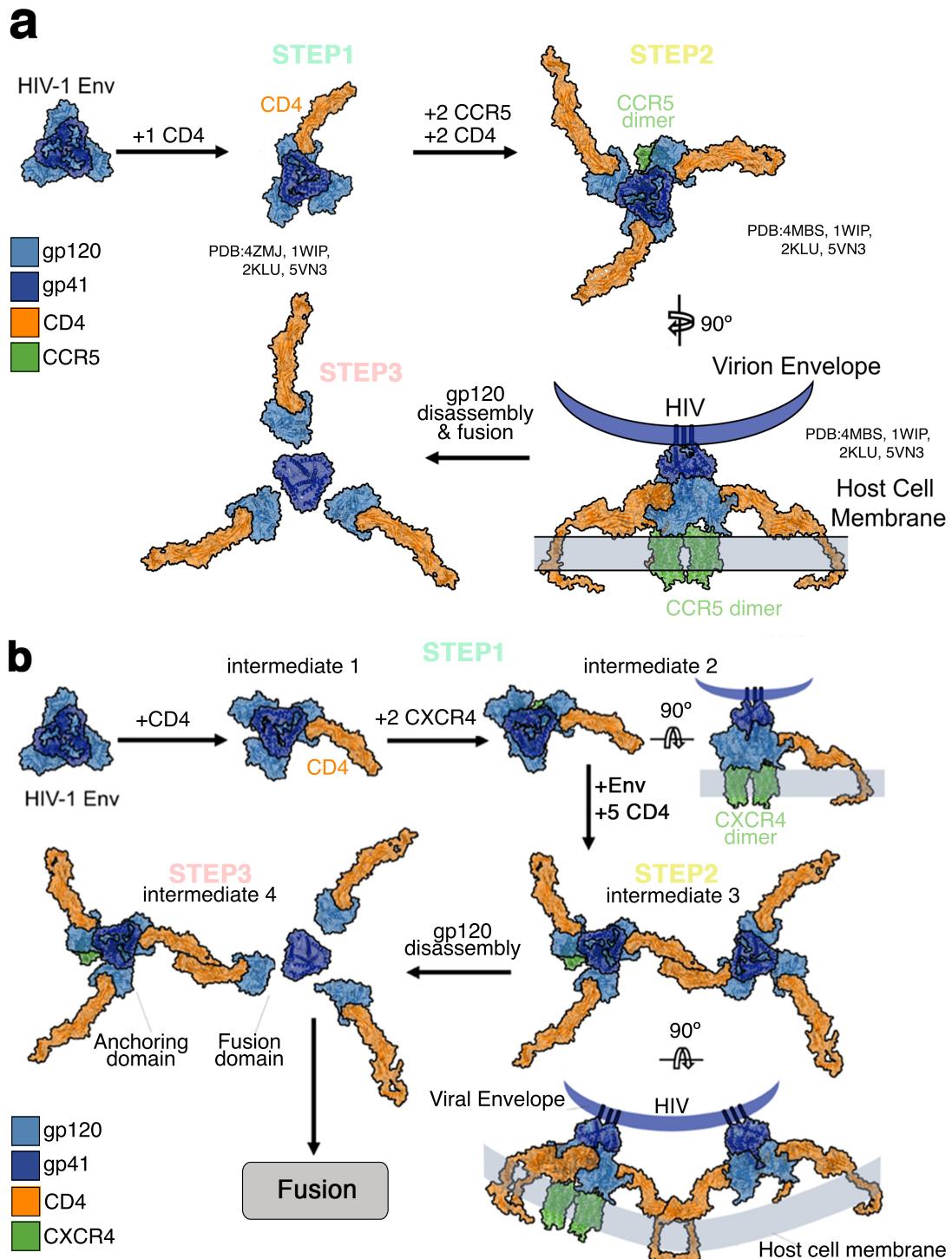


Figure 4.9: The HIV-1 JR-FL Env glycoprotein (blue) (PDB ID: 4ZMJ) is present as a trimer on the mature virion. b, Time-resolved stoichiometry pre-fusion reaction for CXCR4-CD4 (red dots) induced by HIVHXB2-Gag-iCherry virions. (gp120: light blue, gp41: dark blue, CD4: orange, CXCR4: green, b12: yellow)

Chapter 5

Discussion

5.1 Fluorescence fluctuation spectroscopy

Even with photon-counting detectors, FFS is a difficult technique. There are many pitfalls: the acquisition settings (dwell time and frame rate) must be correct and the correct settings for these parameters depend on the residence time τ_D of the protein of interest, which is non-trivial to measure. The acquired data must be checked to ensure there is not an excess of photobleaching and if there is not, there will inevitably still be some bleaching, so this must be corrected for. Until now, it was practically impossible to perform this correction correctly, because all methods required the user to select a vital correction parameter (τ or l) without

providing any instructions as to how the parameter should be chosen. I have now solved this problem such that now image series can be safely detrended by novice users, as this parameter is chosen for them in the background. Now, detrending an image `img.tif` with the *Robin Hood* method is as simple as typing the command `img_detrend_rh("img.tif")` in my software. This should make FFS techniques safer and easier to use, opening FFS techniques up to more users, however the expertise required is still such that FFS may struggle to expand from the domain of microscopy and biophysics into a more commonly used biological technique.

5.2 The evolution of detrending algorithms

Previously, FFS detrending methods were based on smoothing methods taken from the field of time-series analysis. The fact that these smoothing methods required a choice of smoothing parameter was ignored by sticking to the arbitrary choice of $l = 10$ for this parameter.

My work in investigating the significance of this smoothing parameter found that this arbitrary choice was totally incorrect. The use of simulated image series and the fact that immobile particles have brightness $B = 1$ opened up a means of solving for the correct choice of this parameter without the need for user input. This was the first set of *automatic detrending* methods, whereby to detrend, the

users task was as simple as clicking a **detrend** button.

Still, using smoothing approaches to detrend low-intensity data is problematic because it involves approximating very discrete time series with continuous functions; this is unwise and unnecessary. The *Robin Hood* idea of giving photon counts directly from one pixel to another in an image series circumvents the need for smoothing. The detrending process can be simplified from $\mathbb{N}_0 \rightarrow \mathbb{R} \rightarrow \mathbb{N}_0$ to $\mathbb{N}_0 \rightarrow \mathbb{N}_0$. Conveniently, the automatic approach used in the smoothing approaches to detrending can readily be extended to *Robin Hood* detrending.

There are no obvious caveats to the *Robin Hood* bleaching correction method and thus no more work is needed on its theory. However, it is vital to bear in mind that one should always try to avoid the source of error in the first place rather than rely on correction methods.

5.3 Applications of the new detrending techniques

5.3.1 FKBP

The FKBP applications of these detrending techniques in Cos-7 cells (Nolan et al., 2017a) and in-vitro were mainly to demonstrate that N&B used with these detrending techniques is a reliable method to measure oligomerization. The in vitro study (Nolan et al., 2018) was particularly interesting because it was the first in vitro application of number and brightness.

5.3.2 HIV-1 receptor stoichiometry

The study of HIV-1 receptor stoichiometry (Iliopoulou et al., 2018) is a *real-life* application of N&B and ccN&B, made possible by the automatic detrending algorithm. We have shown that this kind of fine-grained information about the process of HIV-1 fusion can be measured on a temporal basis in live cells. Whilst this alone is very exciting, it paves the way for similar studies to be done with HIV-1 and different cell types and indeed for other virus fusion processes to be probed in this way.

5.3.3 Multiplexing with structural biology

Our collaboration with structural biologists (Iliopoulou et al., 2018) is a demonstration of how live cell fluorescence microscopy and structural biology can be complementary. The sub-molecular insight from structural biology is not available from live cell fluorescence microscopy, while the dynamic information from live cell fluorescence microscopy cannot be gotten from structural biology.

5.3.4 Fluorescence fluctuation spectroscopy

FFS has many applications. Indeed the original N&B paper (Digman et al., 2008) alone has over 250 citations. All of these and future FFS studies require correct detrending to be reliable. *Robin Hood* detrending is the answer for this. A major challenge will be making *Robin Hood* visible, available and easy to use for the community. This means that the algorithm must be peer-reviewed, made available in all of the major free imaging softwares (ImageJ, python, R) and very well documented: a good manual is essential with any software package.

Chapter 6

Future plans

6.1 Robin Hood publication

The *Robin Hood* algorithm is already incorporated in the R package `detrendr` but it hasn't been published or peer-reviewed. Getting this done is my top priority.

6.2 Translate software to *ImageJ*

Whilst I really like R, the fluorescence community does not use it, which has been a major barrier to the use of my algorithms. I will code my detrending algorithms

as *ImageJ* plugins and also as python modules so that more of the community have easy access to them.

6.3 Study real data bleaching profiles with simulations

It was mentioned in section 3.10 that it would be possible to study the effect of real bleaching as opposed to simulated ideal bleaching and why it is more necessary to detrend in the real data case by mimicking real bleaching profiles with simulations. This is something I would like to do. It would be difficult because it would require the collection and cataloging of a diverse set of real data bleaching profiles from various biological samples. This would be the bottleneck because I have already written the simulation and analysis pipelines.

6.4 Compare FRET with FCS

I have the idea that when the question “Do these proteins interact?” is answered by Forster resonance energy transfer (FRET, Förster (1948)), it should also be answerable by FCS. I would like to try to reproduce some standard FRET results

with FCS. I am particularly interested to find out if there are instances where one technique succeeds and the other fails and why this might be. For example, it might be possible that for a given interacting pair of proteins, it is impossible to label them such that the FRET couple is close enough for FRET to be detected, but that this interaction is detectable by FCS. These techniques have been compared before ([Sahoo and Schwille, 2011](#)), but not by trying to reproduce previous work. In addition, attempts by new groups to reproduce work that is accepted in the literature are always interesting ([Baker, 2016](#)).

Bibliography

Amara, J. F., Clackson, T., Rivera, V. M., Guo, T., Keenan, T., Natesan, S., Pollock, R., Yang, W., Courage, N. L., Holt, D. A., and Gilman, M. (1997). A versatile synthetic dimerizer for the regulation of protein-protein interactions.

Proceedings of the National Academy of Sciences, 94(20):10618–10623.

Andor (2018a). Emccd cameras.

Andor (2018b). scmos cameras.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604):452–454.

Baker, R. (2005). *CMOS circuit design, layout, and simulation*. IEEE Press, New York.

Brandenberg, O. F., Magnus, C., Regoes, R. R., and Trkola, A. (2015). The HIV-1 entry process: A stoichiometric view. *Trends in Microbiology*, 23(12):763–774.

- Chan, K. H., Hayya, J. C., and Ord, J. K. (1977). A note on trend removal methods: The case of polynomial regression versus variate differencing. *Econometrica*, 45(3):737.
- Chen, Y., Muller, J. D., Berland, K. M., and Gratton, E. (1999). Fluorescence fluctuation spectroscopy. *Methods*, 19(2):234–252.
- Clackson, T., Yang, W., Rozamus, L. W., Hatada, M., Amara, J. F., Rollins, C. T., Stevenson, L. F., Magari, S. R., Wood, S. A., Courage, N. L., Lu, X., Cerasoli, F., Gilman, M., and Holt, D. A. (1998). Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity. *Proc. Natl. Acad. Sci. U.S.A.*, 95(18):10437–10442.
- Dalal, R. B., Digman, M. A., Horwitz, A. F., Vetri, V., and Gratton, E. (2008). Determination of particle number and brightness using a laser scanning confocal microscope operating in the analog mode. *Microsc. Res. Tech.*, 71(1):69–81.
- Declerck, N. and Royer, C. A. (2013). Interactions in gene expression networks studied by two-photon fluorescence fluctuation spectroscopy. In *Methods in Enzymology*, pages 203–230. Elsevier.
- Digman, M. A., Dalal, R., Horwitz, A. F., and Gratton, E. (2008). Mapping the number of molecules and brightness in the laser scanning microscope. *Biophys. J.*, 94(6):2320–2332.

- Digman, M. A. and Gratton, E. (2009). Imaging barriers to diffusion by pair correlation functions. *Biophys. J.*, 97(2):665–673.
- Digman, M. A., Wiseman, P. W., Choi, C., Horwitz, A. R., and Gratton, E. (2009). Stoichiometry of molecular complexes at adhesions in living cells. *Proceedings of the National Academy of Sciences*, 106(7):2170–2175.
- Dunsing, V., Luckner, M., Zühlke, B., Petazzi, R., Herrmann, A., and Chiantia, S. (2018). Quantifying protein oligomerization in living cells: A systematic comparison of fluorescent proteins.
- Einstein, A. (1905a). Concerning an heuristic point of view toward the emission and transformation of light. *Annalen der Physik*, 17:132–148.
- Einstein, A. (1905b). Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen der Physik*, 322(8):549–560.
- FCSXpert (2018). What is the confocal volume?
- Fenz, S. F., Smith, A.-S., and Monzel, C. (2017). Measuring the invisible: Determining the size of growing nanodomains using the “inverse FCS”. *Biophysical Journal*, 112(11):2245–2246.
- Förster, T. (1948). Zwischenmolekulare energiewanderung und fluoreszenz. *Annalen der Physik*, 437(1-2):55–75.

- Gach, J. S., Bouzin, M., Wong, M. P., Chromikova, V., Gorlani, A., Yu, K.-T., Sharma, B., Gratton, E., and Forthal, D. N. (2017). Human immunodeficiency virus type-1 (HIV-1) evades antibody-dependent phagocytosis. *PLOS Pathogens*, 13(12):e1006793.
- Hammamatsu (2007). Photomultiplier tubes: Basics and applications.
- Hebert, B., Costantino, S., and Wiseman, P. W. (2005). Spatiotemporal image correlation spectroscopy (STICS) theory, verification, and application to protein velocity mapping in living CHO cells. *Biophys. J.*, 88(5):3601–3614.
- Herold, N., Anders-Osswein, M., Glass, B., Eckhardt, M., Muller, B., and Krausslach, H.-G. (2014). HIV-1 entry in SupT1-r5, CEM-ss, and primary CD4 t cells occurs at the plasma membrane and does not require endocytosis. *Journal of Virology*, 88(24):13956–13970.
- Hoischen, C., Monajembashi, S., Weisshart, K., and Hemmerich, P. (2018). Multimodal light microscopy approaches to reveal structural and functional properties of promyelocytic leukemia nuclear bodies. *Frontiers in Oncology*, 8.
- Honigmann, A., Mueller, V., Ta, H., Schoenle, A., Sezgin, E., Hell, S. W., and Eggeling, C. (2014). Scanning STED-FCS reveals spatiotemporal heterogeneity of lipid interaction in the plasma membrane of living cells. *Nat Commun*, 5:5412.
- Iliopoulou, M., Nolan, R., Alvarez, L., Watanabe, Y., Coomer, C. A., Jakobsdottir,

- G. M., Bowden, T. A., and Padilla-Parra, S. (2018). A dynamic three step mechanism drives the HIV-1 prefusion reaction. *Nat. Struct. Mol. Biol.*, 0(0).
- Jakobsdottir, G. M., Iliopoulou, M., Nolan, R., Alvarez, L., Compton, A. A., and Padilla-Parra, S. (2017). On the whereabouts of HIV-1 cellular entry and its fusion ports. *Trends in Molecular Medicine*, 23(10):932–944.
- Jameson, D. M., Ross, J. A., and Albanesi, J. P. (2009). Fluorescence fluctuation spectroscopy: ushering in a new age of enlightenment for cellular dynamics. *Biophys Rev*, 1(3):105–118.
- Jones, D. M., Alvarez, L. A., Nolan, R., Ferriz, M., Urruela, R. S., Massana-Muñoz, X., Novak-Kotzer, H., Dustin, M. L., and Padilla-Parra, S. (2017). Dynamin-2 stabilizes the hiv-1 fusion pore with a low oligomeric state. *Cell reports*, 18(2):443–453.
- Kapanidis, A. N., Laurence, T. A., Lee, N. K., Margeat, E., Kong, X., and Weiss, S. (2005). Alternating-laser excitation of single molecules. *Accounts of Chemical Research*, 38(7):523–533.
- Kisley, L. and Landes, C. F. (2014). Molecular approaches to chromatography using single molecule spectroscopy. *Analytical Chemistry*, 87(1):83–98.
- Kwon, Y. D., Pancera, M., Acharya, P., Georgiev, I. S., Crooks, E. T., Gorman, J., Joyce, M. G., Guttman, M., Ma, X., Narpala, S., Soto, C., Terry, D. S., Yang,

- Y., Zhou, T., Ahlsen, G., Bailer, R. T., Chambers, M., Chuang, G.-Y., Doria-Rose, N. A., Druz, A., Hallen, M. A., Harned, A., Kirys, T., Louder, M. K., O'Dell, S., Ofek, G., Osawa, K., Prabhakaran, M., Sastry, M., Stewart-Jones, G. B. E., Stuckey, J., Thomas, P. V., Tittley, T., Williams, C., Zhang, B., Zhao, H., Zhou, Z., Donald, B. R., Lee, L. K., Zolla-Pazner, S., Baxa, U., Schön, A., Freire, E., Shapiro, L., Lee, K. K., Arthos, J., Munro, J. B., Blanchard, S. C., Mothes, W., Binley, J. M., McDermott, A. B., Mascola, J. R., and Kwong, P. D. (2015). Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1 env. *Nature Structural & Molecular Biology*, 22(7):522–531.
- Laboratory for Fluorescence Dynamics (2018). *Globals for Images: SimFCS 4*.
- Landini, G., Randell, D., Fouad, S., and Galton, A. (2016). Automatic thresholding from the gradients of region boundaries. *Journal of Microscopy*, 265(2):185–195.
- Leica (2012). Sensors for true confocal scanning.
- Liu, Q., Acharya, P., Dolan, M. A., Zhang, P., Guzzo, C., Lu, J., Kwon, A., Gururani, D., Miao, H., Bylund, T., Chuang, G.-Y., Druz, A., Zhou, T., Rice, W. J., Wigge, C., Carragher, B., Potter, C. S., Kwong, P. D., and Lusso, P. (2017). Quaternary contact in the initial interaction of CD4 with the HIV-1 envelope trimer. *Nature Structural & Molecular Biology*, 24(4):370–378.

- Ly, S., Bourguet, F., Fischer, N. O., Lau, E. Y., Coleman, M. A., and Laurence, T. A. (2014). Quantifying interactions of a membrane protein embedded in a lipid nanodisc using fluorescence correlation spectroscopy. *Biophysical Journal*, 106(2):L05–L08.
- Ma, X., Lu, M., Gorman, J., Terry, D. S., Hong, X., Zhou, Z., Zhao, H., Altman, R. B., Arthos, J., Blanchard, S. C., Kwong, P. D., Munro, J. B., and Mothes, W. (2018). HIV-1 env trimer opens through an asymmetric intermediate in which individual protomers adopt distinct conformations. *eLife*, 7.
- Marechal, V., Prevost, M.-C., Petit, C., Perret, E., Heard, J.-M., and Schwartz, O. (2001). Human immunodeficiency virus type 1 entry into macrophages mediated by macropinocytosis. *Journal of Virology*, 75(22):11166–11177.
- Mattheyses, A. L., Simon, S. M., and Rappoport, J. Z. (2010). Imaging with total internal reflection fluorescence microscopy for the cell biologist. *Journal of Cell Science*, 123(21):3621–3628.
- Ménager, M. M. and Littman, D. R. (2016). Actin dynamics regulates dendritic cell-mediated transfer of HIV-1 to t cells. *Cell*, 164(4):695–709.
- Mieruszynski, S., Briggs, C., Digman, M. A., Gratton, E., and Jones, M. R. (2015). Live cell characterization of DNA aggregation delivered through lipofection. *Scientific Reports*, 5(1).

- Miyauchi, K., Kim, Y., Latinovic, O., Morozov, V., and Melikyan, G. B. (2009). HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes. *Cell*, 137(3):433–444.
- Moutin, E., Compan, V., Raynaud, F., Clerete, C., Bouquier, N., Labesse, G., Ferguson, M. L., Fagni, L., Royer, C. A., and Perroy, J. (2014). The stoichiometry of scaffold complexes in living neurons - DLC2 functions as a dimerization engine for GKAP. *Journal of Cell Science*, 127(16):3451–3462.
- Munro, J. B., Gorman, J., Ma, X., Zhou, Z., Arthos, J., Burton, D. R., Koff, W. C., Courter, J. R., Smith, A. B., Kwong, P. D., Blanchard, S. C., and Mothes, W. (2014). Conformational dynamics of single HIV-1 envelope trimers on the surface of native virions. *Science*, 346(6210):759–763.
- Nolan, R., Alvarez, L., Elegherert, J., Iliopoulou, M., Jakobsdottir, G. M., Rodriguez-Muñoz, M., Aricescu, A. R., and Padilla Parra, S. (2017a). nandb—number and brightness in r with a novel automatic detrending algorithm. *Bioinformatics*, 33(21):3508–3510.
- Nolan, R., Alvarez, L., Griffiths, S. C., Elegherert, J., Siebold, C., and Padilla Parra, S. (2018). Calibration-free in-vitro quantification of protein homooligomerization using commercial instrumentation and free, open source brightness analysis software. *Journal of Visualized Experiments*, 0(0).

- Nolan, R., Iliopoulou, M., Alvarez, L., and Padilla-Parra, S. (2017b). Detecting protein aggregation and interaction in live cells: A guide to number and brightness. *Methods*.
- Nolan, R. and Padilla-Parra, S. (2017a). examplestr - an easy start to unit testing r packages. *Wellcome Open Research*, 2:31.
- Nolan, R. and Padilla-Parra, S. (2017b). filesstrings: An r package for file and string manipulation. *The Journal of Open Source Software*, 2(14):260.
- Nolan, R. and Padilla-Parra, S. (2018). ijtiff: An r package providing TIFF i/o for ImageJ users. *Journal of Open Source Software*, 3(23):633.
- Ozorowski, G., Pallesen, J., de Val, N., Lyumkis, D., Cottrell, C. A., Torres, J. L., Copps, J., Stanfield, R. L., Cupo, A., Pugach, P., Moore, J. P., Wilson, I. A., and Ward, A. B. (2017). Open and closed structures reveal allostery and pliability in the HIV-1 envelope spike. *Nature*, 547(7663):360–363.
- Padilla-Parra, S. (2009). *Developments METHODOLOGIQUES POUR LA QUANTIFICATION DES INTERACTIONS PROTEIQUES EN CELLULES VIVANTES PAR MICROSCOPIE DE FLUORESCENCE*. PhD thesis, Université Paris Diderot.
- Photometrics (2018). Photometrics launches next generation scientific cmos camera with 95 percent quantum efficiency.

Qian, H. and Elson, E. L. (1990). Distribution of molecular aggregation by analysis of fluctuation moments. *Proc. Natl. Acad. Sci. U.S.A.*, 87(14):5479–5483.

Qin, L., Kufareva, I., Holden, L. G., Wang, C., Zheng, Y., Zhao, C., Fenalti, G., Wu, H., Han, G. W., Cherezov, V., Abagyan, R., Stevens, R. C., and Handel, T. M. (2015). Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science*, 347(6226):1117–1122.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rienzo, C. D., Cardarelli, F., Luca, M. D., Beltram, F., and Gratton, E. (2016). Diffusion tensor analysis by two-dimensional pair correlation of fluorescence fluctuations in cells. *Biophysical Journal*, 111(4):841–851.

Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., and Eliceiri, K. W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, 18(1):529.

Sahoo, H. and Schwille, P. (2011). FRET and FCS-friends or foes? *ChemPhysChem*, 12(3):532–541.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., and Cardona, A. (2012). Fiji:

- an open-source platform for biological-image analysis. *Nat. Methods*, 9(7):676–682.
- Stroustrup, B. (2013). *The C++ Programming Language*. Addison-Wesley Professional, 4th edition.
- Tan, Q., Zhu, Y., Li, J., Chen, Z., Han, G. W., Kufareva, I., Li, T., Ma, L., Fenalti, G., Li, J., Zhang, W., Xie, X., Yang, H., Jiang, H., Cherezov, V., Liu, H., Stevens, R. C., Zhao, Q., and Wu, B. (2013). Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. *Science*, 341(6152):1387–1390.
- Wu, L., Gerard, N. P., Wyatt, R., Choe, H., Parolin, C., Ruffing, N., Borsetti, A., Cardoso, A. A., Desjardin, E., Newman, W., Gerard, C., and Sodroski, J. (1996). CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature*, 384(6605):179–183.
- Zhao, M., Jin, L., Chen, B., Ding, Y., Ma, H., and Chen, D. (2003). Afterpulsing and its correction in fluorescence correlation spectroscopy experiments. *Appl Opt*, 42(19):4031–4036.