

High-throughput single-molecule fluorescence spectroscopy using parallel detection

X. Michalet^{*a}, R. A. Colyer^a, G. Scalia^a, T. Kim^b, Moran Levi^c, Daniel Aharoni^c, Adrian Cheng^c, F. Guerrieri^d, Katsushi Arisaka^c, Jacques Millaud^a, I. Rech^d, D. Resnati^d, S. Marangoni^d, A. Gulinatti^d, M. Ghioni^d, S. Tisa^d, F. Zappa^d, S. Cova^d, S. Weiss^a

^aDept of Chemistry & Biochemistry, Los Angeles, CA, USA 90095; ^bNesher Technologies, Los Angeles, CA; ^cDept of Physics & Astronomy, UCLA, Los Angeles, CA 90095; ^dDipartimento di Elettronica ed Informazione, Politecnico di Milano, Milano, Italy

ABSTRACT

Solution-based single-molecule fluorescence spectroscopy is a powerful new experimental approach with applications in all fields of natural sciences. The basic concept of this technique is to excite and collect light from a very small volume (typically femtoliter) and work in a concentration regime resulting in rare burst-like events corresponding to the transit of a single-molecule. Those events are accumulated over time to achieve proper statistical accuracy. Therefore the advantage of extreme sensitivity is somewhat counterbalanced by a very long acquisition time. One way to speed up data acquisition is parallelization. Here we will discuss a general approach to address this issue, using a multispot excitation and detection geometry that can accommodate different types of novel highly-parallel detector arrays. We will illustrate the potential of this approach with fluorescence correlation spectroscopy (FCS) and single-molecule fluorescence measurements obtained with different novel multipixel single-photon counting detectors.

Keywords: single-molecule, fluorescence, spectroscopy, photon-counting, FRET, FCS, SPAD, array

1. INTRODUCTION

Single-molecule spectroscopy has become a mature and vibrant field with applications in many scientific disciplines and a large number of reviews are available to the interested reader (see for instance ^{1, 2}). This paper is mostly concerned with single-molecule fluorescence spectroscopy of solutions, but many of the arguments discussed here would apply in many different situations (such as immobilized molecules or single-molecule imaging). In this section, we briefly discuss the technique and data analysis used in standard solution-based single-molecule spectroscopy.

1.1 Solution-based single-molecule fluorescence spectroscopy

Solution-based single-molecule fluorescence spectroscopy is particularly useful to study such questions as the conformation of biomolecules, the dynamics of enzymatic function and more generally, address outstanding questions of biology, biochemistry, biophysics, and chemistry at the molecular level (in contrast to the ensemble-averaged picture provided by standard bulk measurements). Single-molecule techniques entail a few basic setup requirements: (i) optimization of the excitation and collection of the fluorescence signal, (ii) rejection of background signal and (iii) optimization of the temporal resolution of the acquisition. Obviously, the sample itself needs to meet a number of criteria (photophysical stability, brightness, etc), which will not be discussed here. In practice, there are many ways to meet these setup requirements. We will focus here on the most common technique using confocal microscopy, although, as mentioned before, most of the arguments developed in the following apply to other optical arrangements.

Confocal microscopy is achieved by focusing a collimated laser light in the sample using a high-numerical aperture objective lens ($NA = n \cdot \sin \alpha$, where n is the refraction index of the medium between the lens and the glass coverslip holding the sample, and α is the half angular aperture of the lens). Due to diffraction, the typical intensity distribution at the focal point has a finite extension, the point-spread-function (PSF), which can in general be well approximated by a 3-dimensional (3D) Gaussian³. Its standard deviation perpendicularly to (σ_{xy}) and along (σ_z) the optical axis, are given by:

* michalet@chem.ucla.edu, Phone: 1 310 794-6693, Fax: 1 310 267-4672

$$\sigma_{xy} \sim 0.21 \frac{\lambda}{NA}, \quad \sigma_z \sim 0.53 \frac{\lambda}{NA^2} \quad (1)$$

where λ is the wavelength of the laser excitation. In our experiments, $\lambda = 532$ nm and $NA = 1.2$, hence the typical excitation volume $V_x = 4/3\pi (2\sigma_{xy})^2(2\sigma_z) \sim 0.06$ fl (1 fl = 10^{-15} liter). By construction, the confocal detection system collects lights emitted from approximately the same microscopic volume. In other words, this configuration only excites molecules present in a very small volume.

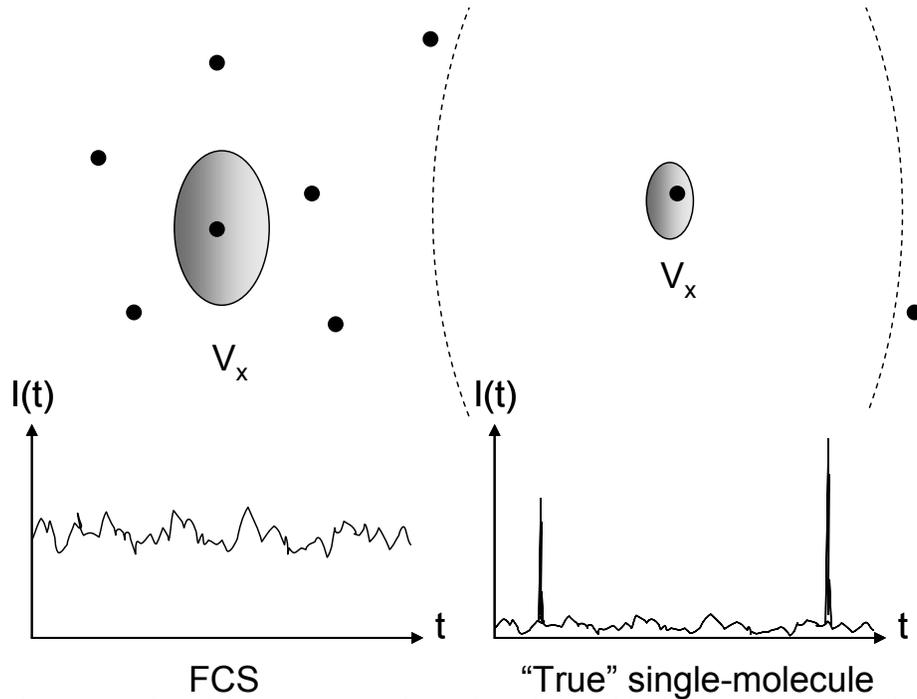


Fig. 1: Schematic representation of the two concentration regimes discussed in the text. V_x represents the effective excitation volume, represented as an ellipsoid. Black dots represent fluorescent molecules. In FCS, the average number of molecules is 1 per V_x or less. For single-molecule burst detections, molecules need to be further apart. The dashed lines represent the contour of an ellipsoid with a volume $10^3 V_x$. Intensity time traces $I(t)$ recorded in these two different situations are represented below.

The definition of single-molecule regime is unfortunately a bit ambiguous. A natural definition would be to set it as the regime where the above excitation volume contains less than one molecule on average, yielding a maximum concentration $C \sim 20$ nM (see Fig. 1). This is actually not far off from typical concentrations used in fluorescence correlation spectroscopy (FCS) discussed later, which typically use concentrations of 10 nM or less. In this loose sense, FCS is a single-molecule technique. However, having one molecule per excitation volume on average is a sure way to have more than one molecule in it most of the time. In other words, this regime does not allow the separate detection of photons emitted from individual molecules. The only way to achieve such a goal is to impose that the time separating the exit of one molecule from the entry of the closest one into the excitation volume is significantly larger than the duration of a single-molecule transit through the excitation volume. Using the diffusion constant of R6G, a fluorescent dye used in this study: $D = 280 \mu\text{m}^2/\text{s}$, we obtain a typical transit time τ_D of a few 100 μs . To guarantee a safe separation between molecules, the separation between individual molecule transits needs to be at least two orders of magnitude larger than τ_D , which translates into a concentration at least three orders of magnitude lower than our previous estimate, i.e. $C < 80$ pM. This estimate is right in the ballpark of values used in single-molecule experiments interested in detecting single-molecule bursts.

1.2 Fluorescence Correlation Spectroscopy

Fluorescence correlation spectroscopy (FCS) is a generic designation for a variety of different techniques having in common the analysis of the fluctuations in fluorescence intensity recorded from a sample (reviewed in ⁴). For typical concentrations used in FCS measurements, the signal is roughly constant but noisy, with occasional spikes when the concentration is very low (Fig. 1). Part of this "noise" is actually due to photoemission and photodetection noise and is

therefore not very interesting. In favorable conditions, a significant contribution to this “noise” comes from chemical, photophysical or mechanical processes. For instance, binding or unbinding of a molecule to a ligand may quench or enhance its fluorescence quantum yield, resulting in fluctuations of the signal. Most importantly, diffusion throughout the excitation volume results in fluctuations characterized by a typical time scale equal to the diffusion time τ_D :

$$\tau_D = \frac{\sigma_{XY}^2}{D} \quad (2)$$

Knowledge of σ_{XY} (for instance using calibration experiments with a fluorescence molecule of known diffusion constant) allows extracting D from the measurement of τ_D . Since D is related to the dimension of the diffusing molecule via the Stokes-Einstein equation, FCS measurements give access to the size of molecules in the sample, allowing for instance the detection of binding or dimerization events. In the simplest case of a single diffusing species, the normalized autocorrelation function (ACF) of the intensity signal $I(t)$ takes the form:

$$G(\tau) = \frac{\langle I(t)I(t+\tau) \rangle}{\langle I \rangle^2} = \frac{1}{CV} \left(1 + \frac{\tau}{\tau_D} \right)^{-1} \left(1 + \frac{\tau}{\omega^2 \tau_D} \right)^{-\frac{1}{2}} \quad (3)$$

where $\omega = \sigma_z/\sigma_{xy}$ and $V = 3\pi^{3/2}/4 \cdot V_x$. Note that in the case of elongated excitation PSFs ($\omega \gg 1$), the last term can be dropped. This modified (2-dimensional) expression is a good approximation for many actual experiments. In other words, the amplitude $1/CV=1/N$ of the normalized ACF gives access to the concentration of the sample. Additional processing of the intensity time trace acquired with different time resolution can give access to molecular brightness information, which helps further characterize the sample. Note that the normalized autocorrelation function has apparently no dependence on the signal intensity itself. In fact, signal intensity has a direct influence on the signal-to-noise ratio of the measurement, as will be discussed shortly, and is therefore an important experimental factor. The above expression is modified in the presence of a Poisson-distributed background signal B superimposed to the mean signal $\langle I \rangle$ as⁵:

$$G_B(\tau) = \left(1 + \frac{B}{\langle I \rangle} \right)^{-2} (G(\tau) - 1) + 1 \quad (4)$$

Therefore, the amplitude of the ACF is reduced in the presence of a significant background, but its temporal behavior is unaffected.

ACF analysis comprises two distinct steps: calculation of the ACF and fit to a model. In standard approaches, the intensity time trace is recorded from a photon counting device and binned with a finite temporal resolution (or dwell time) which is much smaller than the shortest timescale of interest in the experiment (e.g. 1 μ s for studies of single-molecule diffusion in solution, for which $\tau_D \sim 100 \mu$ s). The resulting bins are correlated by a hardware or software correlator using a multitau scheme reducing the number of data points to be computed, while giving access to the evolution of $G(\tau)$ over several decades⁶. Alternatively, the arrival time of each photon can be recorded and the ACF obtained from a dedicated algorithm, the idea remaining to provide access to several decades of temporal evolution, while keeping the actual number of calculated data points reasonably low ($\sim 1,000$)⁷.

Once the ACF has been calculated, the next step is trying to fit an appropriate model to the measured curve. This is a far from trivial task for several reasons. First, real life situations are rarely as simple as the one modeled by Eq. 3 (or its 2D version). In addition to not dealing with a perfectly Gaussian excitation volume, the signal can be contaminated by photophysical fluctuations of the fluorescent dye, background noise or noise coming from the detector (such as dark counts – which is uncorrelated, or afterpulsing – which is correlated). Finally, the sample may contain different molecular species with unknown characteristics. Whereas it is in principle possible to account for all these departures from Eq. 3 by adding more parameters to the fit function, the fitted parameter values become less constrained to the point of being quite meaningless. Although this point is obviously critical, there is surprisingly no theoretical treatment of the resulting fit uncertainties as a function of control parameters such as signal-to-noise ration (SNR) or signal-to-background ratio (SBR). What exists, however, is a theoretical estimation of the variance of the ACF in the ideal case modeled by Eq. 3^{8,9}. To keep this introduction short, we simply summarize one result of this analysis in Fig. 2, namely

the SNR of $G(0)$ as a function of the mean number of particle per sampling volume N , the mean count rate per molecule q , the time resolution θ and the total acquisition time T in different regimes⁸.

Longer acquisition time T is always advantageous, but only increases the SNR as $T^{1/2}$. The brightness per molecule, $q\theta$, has no noticeable effect above 1 but affects the SNR linearly at low count rates, both in the “true” single-molecule regime ($N \ll 1$) and the non single-molecule regime ($N \gg 1$).

We can use these results to qualitatively estimate the uncertainty on fitted parameters, since the uncertainty will certainly increase with decreasing SNR. In all cases, it will be advantageous to increase the duration of the recording. To obtain more reliable parameters from a FCS measurement at low concentration ($N \ll 1$), one may want to increase the

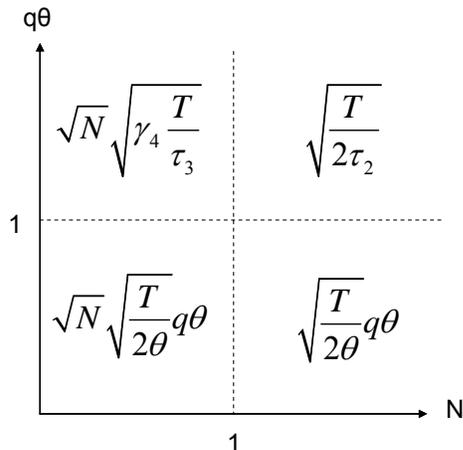


Fig. 2: Signal-to-noise ratio of $G(0)$ in different regimes of number of molecules per sampling volume (N) and signal per molecule per dwell time ($q\theta$). γ_4 , τ_2 and τ_3 are constants depending on the geometry and the functional form of the ACF.

brightness per particle if it is very low ($q\theta \ll 1$), or increase the concentration (in essence, increase the burst frequency). At high concentrations, increasing the brightness per particle will improve the SNR up to a certain point, but will have no effect past $q\theta > 1$. The brightness per particle can be increased in several ways: (i) by increasing the excitation power, (ii) increasing the detection efficiency, or (iii) increasing the dwell time. The first approach is limited by dye photophysics (saturation, blinking, bleaching, etc) and in some cases by the available laser power (especially for multispot excitation as described later). The second criterion is an obvious requirement for any detector. Any decrease in quantum efficiency (QE) needs to be compensated by an increase in acquisition time by the same factor squared. The last option is constrained by the minimal time lag required for the ACF fit, which is typically around 10 μ s.

1.3 Single-molecule burst detection

Although FCS can provide quantitative information on a sample, its sensitivity to a number of hard-to-measure experimental parameters makes it a poor choice in a number of situations. After the introduction of confocal geometry as a way to increase the sensitivity of FCS, it was only a matter of time until researchers started to analyze the information contained in single-molecule bursts in a more direct way than through the intensity ACF. Techniques such as FIDA¹⁰, FIMDA¹¹ or PAID¹² are extensions of FCS, which specifically use the burst size distribution and burst duration distribution (or their joint distribution) to extract further information on the sample. In particular, these techniques can achieve a robust identification of different species within a sample, a feat which is much more difficult to attain using simple ACF analysis. Due to their complexity, we will not discuss these techniques in this brief section. Instead, we will focus on the simpler problem of burst identification, on which all these techniques (and many others) depend.

At low concentration, the recorded intensity time trace is comprised of long stretches of shot noise-limited background signal interspersed with short bursts corresponding to the transit of a single-molecule (or particle) through the excitation volume (see right-hand side of Fig. 1). Most analyses are only concerned with the exact number of photons in each burst and in some case their precise arrival time and duration.

For instance, in the spectroscopic approach known as FRET (Fluorescence Resonant Energy Transfer)², a biomolecule labeled with two organic dyes of different color is probed with a laser excitation tuned for absorption by the first dye (called the donor) while the emission of both dyes is monitored in two separate channels. When the organic dyes are far

apart (typically 10 nm or more), the donor will be the only dye to emit photons. However, when the two dyes are close to one another, the donor can resonantly transfer its excitation energy to the other dye (called the acceptor), provided the donor emission spectrum overlaps the acceptor emission spectrum. The efficiency of this transfer depends very sensitively on the distance R between the two dyes and can be calculated as:

$$E(R) = \frac{I_A}{I_A + \gamma I_D} \quad (5)$$

where I_A is the acceptor signal, I_D the donor signal and γ is the ratio between the product of detection efficiency and quantum yield calculated for the acceptor and the donor. In other words, counting the number of donor and acceptor photons in a burst (using two spectrally separated channels) gives access to information on the biomolecule's conformation during its transit through the excitation spot. Accumulation of information from many such burst allows probing the distribution of conformations in a sample or study the evolution of their conformation over time. Although the technique can use more than two organic dyes, distinguish between polarizations or involve measuring the emission delay after pulsed laser excitation, we will limit our discussion to the simple case of two channel recording FRET measurements.

As for FCS, the transit time (and hence the burst duration) is set by the excitation spot dimension and the diffusion coefficient of the molecule. In most applications, this results in burst duration of the order of 1 ms. The burst size (number of photons in all channels) depends on the transit time and many optical parameters such as excitation intensity, dye quantum yield, detection efficiency, etc, but is of the order of 100 photons or less. At this signal level, shot noise is dominating and any experimental factor reducing the signal is detrimental¹³. Background influences the data at two different levels: burst detection and signal to noise ratio. Burst detection is affected by background in two different ways: large signal fluctuations due to background shot noise can be falsely counted as signal and small bursts can be buried in background fluctuations and remain undetected. Once the bursts are detected, it is easy to subtract the background component knowing the burst durations. However, the shot noise contribution of the background component remains in the signal, which results in a broadening of the FRET histogram¹³. If one does not measure FRET efficiency but instead simply looks for coincidence between two dye emissions, this effect of the background is of course not an issue.

As in FCS, longer acquisition times are beneficial, as single-burst detection and analysis (for instance using Eq. 5) only makes sense if many such bursts information are pooled together and analyzed statistically. Typically, thousands of bursts are required to obtain statistically significant information on the sample (be it FRET histogram or coincidence detection measurements). In practice, this implies that only steady state equilibrium or very slow dynamic processes can be studied with this approach, unless ways of speeding up the acquisition of a similar number of bursts are found.

2. HIGH-THROUGHPUT SINGLE-MOLECULE SPECTROSCOPY

2.1 High-Throughput Single-Molecule Spectroscopy using Parallelization

The previous section has made it clear that longer acquisition time is beneficial to the statistical accuracy of information extracted from single-molecule experiments (for either FCS or burst detection). Unfortunately, although long acquisition time (up to hours) may be acceptable in basic research, this requirement poses significant throughput constraints for clinical or biopharmaceutical applications or any fundamental study requiring access to shorter time scales than the minute time scale accessible by standard approaches. One obvious way to improve on this limitation is to acquire data from multiple identical spots. The only reasons why this approach has been only moderately successful up to now^{14, 15} is due to the lack of appropriate technology to: (i) create high-quality multispot excitation patterns; (ii) detect the corresponding signals in parallel; (iii) acquire this high-throughput data stream; and (iv) process the data efficiently. In this section, we will review these four different facets of the problem in more detail.

2.2 Parallel Excitation

Generating multiple excitation spots from a single laser beam can be achieved in many ways. For instance, microlens arrays, multiple beamsplitters, micromirror arrays, liquid crystal on silicon (LCOS) spatial light modulators (SLM) or diffractive optics elements (DOE) have been used in the past to achieve such a result. The technological choice depends on (i) the spectral and power characteristics of the laser source, (ii) the required flexibility in terms of adjusting the generated pattern spatially and temporally and (iii) cost. In addition to ease of use and flexibility of the generated pattern, several criteria need to be considered: (i) what is the required laser beam expansion to cover the device; (ii) what is the

necessary pattern periodicity to avoid optical crosstalk between spots; (iii) what are the transmission/reflection characteristics of the device; (iv) what is the incident power damage threshold; (iv) thermal and mechanical stability of the device; (vi) size and cost of the device. In the following, we will limit ourselves to presenting data obtained with a microlens array for illustration purposes. The optical crosstalk consideration stems from the fact that we are interested in generating independent spots from which we will be able to gather data that is uncorrelated with data acquired by other spots. A simple criterion to ensure this, is to impose that the average time scale for a molecule to move from one spot to the next is much larger than the largest time scale of interest in the experiment. Since the largest time scale is usually the diffusion time across an excitation spot, we need only to have spots separated by a distance d much larger than the size of an individual spot (Eq. 1). In practice, a separation of a few microns is in general sufficient for this criterion to be verified. An additional consideration is the out-of-focus signal generated from neighboring spots in a square array. Even when the spots are at a “safe” distance from one another, diffuse intensity contributed from first, second, etc, neighbors may add up (or interfere constructively) to generate a significant excitation intensity in regions where the isolated PSF would normally not excite any fluorescence. This may generate a large additional background signal and significantly affect the shape of the ACF or the burst size distribution, making data analysis more problematic. The ratio between spot separation and spot size will play an important role in the image plane (detection), thus it is convenient to have some flexibility for the choice of this ratio in the excitation path.

2.3 Parallel Detection

Independent and simultaneous recording of photons emitted from multiple spots requires either a larger area detector with a corresponding number of (optically and electronically isolated) individual integrating pixels or an array of individual point detectors. An integrating pixel is a sensitive area with an associated electronics counting the number of impinging photons on this area, which does not provide time-tags for individual photons. A point detector on the contrary, provides a pulse for each detected photons, which can be time-tagged by an associated electronics with a temporal resolution limited by the detector response function jitter and walk. Examples of such point detectors are single-photon counting avalanche photodiodes (SPAD), single-photon counting photomultiplier tubes (PMTs) or hybrid photodetectors (HPDs)¹⁶. Pixels count the number of photons detected during a fixed period of time, which is thus by definition the shortest accessible time scale. Silicon photomultipliers (SiPMs) arrays and electron multiplying CCDs (EMCCDs) are example of pixilated detectors (note that SiPM pulses corresponding to one or many photons can in principle also be time-tagged).

Larger area detectors such as EMCCD cameras have been used for single diffusing molecule spectroscopy¹⁵, but their temporal resolution is limited by the maximum data throughput and readout noise (discussed in the next section). Time-tagging detectors such as SPADs have thus an advantage over integrating camera in terms of temporal resolution (they provide ~ 100 ps resolution time tags, allowing fluorescence lifetime decay to be studied when pulsed laser excitation is used). However, individual SPAD units (or PMTs, or HPDs) are expensive and too large to be used as detectors for multispot detection.

An alternative having emerged in the past few years is integrated arrays of single-photon counting point detectors based on SPAD or HPD technologies. Depending on the technology, some arrays require to trade-off some of the performance of individual point detectors. For instance, the larger the number of elements in an array, the more likely it is that individual pixel performance (quantum efficiency, dark count, etc) will be different from one another. Without getting in the details of each technology, we will simply list here important criteria that detector arrays need to obey to be used for single-molecule spectroscopy (SMS).

The geometrical arrangement of the pixels is not a critical parameter, as long as it can be matched by a corresponding excitation spot geometry (up to a magnification factor). In general, high numerical objective lenses used for SMS are combined with a long focal length tube lens, resulting in a magnification factor $M \sim 40$ to 100. Additional lenses can be used to adjust this magnification either way. An obvious constraint is that the pitch of the excitation pattern, d , times the magnification, M , needs to match that of the array, D (Fig. 3). The excitation pattern generator and optics (or laser power) may impose some constraints as far as minimum or maximum spot separation (or pattern extent) are concerned. But most importantly, single spots in the excitation pattern have ideally a diffraction limited size (Eq. 1). The image of this spot must fit quite exactly within the sensitive area of a single pixel to preserve confocality and efficiently collect the emitted photons. Too small an image and the single pixel will collect out of focus light, whereas too large an image and the signal will be clipped and become undetectable. Therefore, the only significant constraint is to match the size of the excitation spot, $2r \sim 4\sigma_{xy}$ and the size of a single pixel of the array, $2R$ (Fig. 3):

$$M \sim \frac{R}{2\sigma_{xy}} \sim 2.4NA \frac{R}{\lambda} \quad (6)$$

This defines the optical magnification, which in turns sets the spot separation in the excitation path:

$$d = \frac{D}{M} \gg 5\sigma_{xy} \sim \frac{\lambda}{NA} \quad (7)$$

The last inequality simply states that spots need to be sufficiently separated to avoid cross-talk (section 2.2). This constraint is reflected on the detection side as a constraint on the individual pixel fill factor, f :

$$f = \frac{\pi R^2}{D^2} \ll 0.55 \quad (8)$$

Therefore, due to the confocality requirement, a small fill factor is in fact an advantage. This is good news for detector manufacturers, as this can be achieved by building smaller area detectors, which is advantageous in terms of reduced dark counting rate or number of defects. However, too small a pixel will require difficult optical quality and alignment constraints, setting a realistic minimum size of $\sim 10 \mu\text{m}$ diameter for the sensitive area. Note that if the detector has too large a fill factor, a pinhole array can in principle be put in front of the detector array, reducing the effective sensitive area dimension R_{eff} so that Eq. 8 is verified.

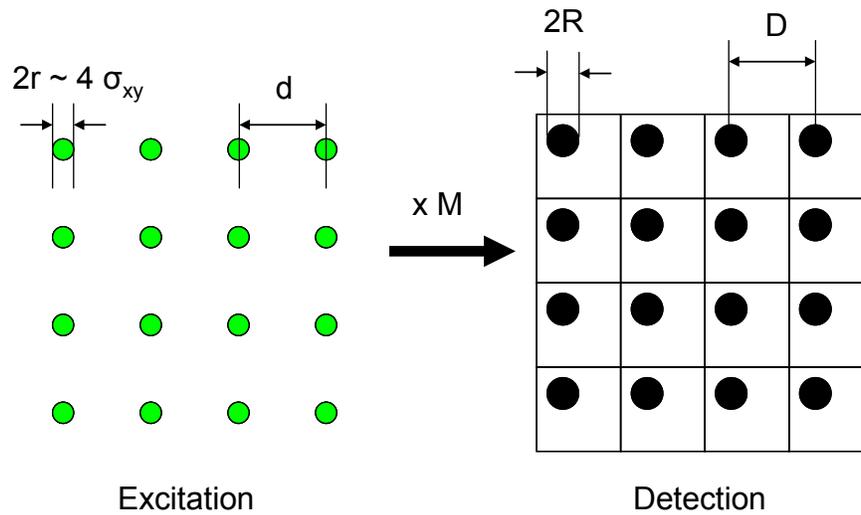


Fig. 3: Relation between geometric parameters in the excitation path and the emission path. The first constraint (on the magnification M) is set by matching the size of the excitation PSF ($2r$) with the pixel sensitive area diameter ($2R$). The spot separation (d) is then set by the pixel pitch (D).

2.4 Parallel Data Acquisition

Cameras are acquiring data in parallel by design, as all pixels accumulate photon counts simultaneously during the integration time. Data readout and transfer breaks down this parallelism to some extent. However, they have very inefficient use of data bandwidth in the case of single-molecule bursts. To understand this, consider a hypothetical 32×32 pixel camera recording the number of photons per pixel per $10 \mu\text{s}$ bin as an 8 bit (1 byte) number¹⁷. The experimental single-molecule sample (randomly) generates 100 photon bursts per pixel with an approximate burst duration of $100 \mu\text{s}$, separated by 100ms . For each set of 10 bytes of useful information, this system will collect 10,000 bytes (10 KB) of background information (with values of mostly 0 and much more rarely 1) per pixel. The data rate will therefore be approximately 100 KB/s. In comparison, time-tagging devices will generate 4 bytes of data per photon (which will represent the value of a 32 bit counter incremented by a information fast clock), i.e. ~ 400 bytes per burst, plus time tags of background photons. Since typical background levels are of the order of a few kHz, the data rate would be $\sim 2,000 \times 4$ bytes < 10 KB/s per channel. This simple comparison shows that down to a temporal resolution of $100 \mu\text{s}$, the camera

efficiently competes with time-tagging detectors (assuming identical detection efficiency). However, if one wants to achieve higher temporal resolution, time-tagging will be advantageous (the amount of photon counts does not depend on the integration time). Note that by using fewer bits to encode the photon count per time bin, cameras can in principle perform increase their performance (4 times better using 2 bits rather than 8). On the other hand, handling a large number of single-photon counting pulse trains (channels) can be challenging, requiring parallel readout connections and separate counting units.

A final consideration is data transfer to the computer memory (and eventually, hard drive). Relatively high throughput is required as the number of pixels or individual photon-counting detectors are used. For instance, a 64 channels time-tagging device would output $\sim 64 \times 10$ KB/s, which is well within the capabilities of PCI or USB 2.0 buses (and even more so PCI-Express or USB 3.0). 10 times larger count rates could easily be supported, but at higher count rates, integrating devices such as cameras or photon binning or burst detection data reduction algorithms running in the counting unit hardware (for instance in FPGA) would be necessary. Even though parallelized data acquisition helps reducing the duration of data acquisition, the amount of data remains identical generating large data. For instance, a 32 x 32 pixel camera encoding each 10 μ s bin of data on 2 bits will generate ~ 24 MB/s of data, i.e 1.5 GB/min.

2.5 Parallel Data Processing

ACF calculation is CPU and memory intensive. Due to memory constraints and the large number of decades spanned by experimental ACF, the standard approach using Fast Fourier Transform (FFT) is not applicable. Software implementations of the multitau algorithm used in hardware correlators achieve relatively fast performance but require binned data (provided natively by camera detectors) and are therefore inefficient at handling time-tag streams generated by arrays of photon-counting detectors¹⁸. Efficient algorithms have been developed to implement a similar approach to the multitau algorithm⁷, but without the need of binning the data. In both cases, the amount of memory and CPU scales linearly with the number of pixels and can therefore become rapidly paralyzing, even on high end PCs. A possible alternative to resorting to PC clusters is using multi GPU boards.

Burst detection on the other hand is a relatively simple task to perform and could in principle be done online (for instance on FPGA), allowing a significant reduction of the amount of transferred data. Indeed, a burst is fully defined by its starting time-tag (or time elapsed since the previous burst), duration and height, information that could easily be stored in less than 10 bytes. With a burst frequency of 10-100 Hz, this would result in extremely low data bandwidth and be amenable to online histogramming and processing (e.g. using Eq. 5).

3. METHODS

Optical Setup: Multispot excitation patterns were first generated using a microlens array (MLA150-7AR, Thorlabs, NJ) with a 150 μ m pitch and a 6.7 mm focal length. The expanded collimated light of the 488 nm line of an Argon ion laser (ILT 5490A, Midwest Laser Products, IL) or of a 532 nm emitting ps pulsed laser (IC-532-1000 ps, High Q Laser, MA) was directed toward the microlens array, forming an intermediate array of illumination spots, which were then re-imaged with a high NA objective lens into the sample, generating quasi-diffraction limited spots. To illuminate only a given number of spots, a custom-made slit was placed in the intermediate image plane to act as a field stop. The excitation profile was measured using stage-scanning confocal imaging of a sub-diffraction sized fluorescent bead with a single-pixel HPD as described in ref. ¹⁶. We also used a LCOS SLM to generate a similar pattern with much more flexibility but essentially the same results (data not shown).

Detectors: Experiments were performed using different detector prototypes including a 8x1 time-tagging custom CMOS SPAD array¹⁹, a 32x32 integrating standard CMOS SPAD array¹⁷ and a 8x8 HPD array from Hamamatsu Photonics. We will only discuss results obtained with the first detector, which provides 100 ns wide TTL pulses for each detected photon on a separate cable for each SPAD.

Data Acquisition: 8 BNC 50 Ω -terminated cables were used to connect the individual SPAD pulse streams (TTL) to a breakout box (CA-1000, National Instruments, TX) used as an interface to a reconfigurable counting board (PXI-7813R, NI). The on-board FPGA was configured using LabVIEW to time-tag each pulse using a 80 MHz clock. For each detected photon, a 32 bits counter value and a channel identification number (32 bits) were sent asynchronously to the PC via a Direct Memory Access (DMA) First-in First-out (FIFO) buffer. A LabVIEW software allowed online visualization of the 8 intensity time traces and saving the data on disk.

Data Analysis: Post acquisition, the same software allowed loading saved files, representing time traces with arbitrary binning, ACF calculation and burst size and duration analysis was performed in LabVIEW using C DLL calls for ACF

calculation using published algorithms^{7, 18}. All curves were exported as ASCII files for representation in Origin 8 (OriginLab, MA) or fitting in custom-designed LabVIEW or C software.

Sample preparation: Fluorescent bead samples of 100 nm diameter used for FCS measurements were purchased from Invitrogen (Carlsbad, CA). Samples were diluted in appropriate buffers, sonicated for 5' and ultracentrifugated to eliminate aggregates. Cy3B-labeled DNA samples used for single-molecule burst analysis were prepared as described in ref.²⁰.

4. RESULTS

4.1 Optical setup characterization

The excitation scheme described on Fig. 4A resulted in a row of quasi-diffraction limited excitation spots ($FWHM_{XY} \sim 500$ nm, $FWHM_Z \sim 2.5$ μ m) separated by ~ 3 μ m as shown by 3-dimensional (3D) confocal imaging of a 100 nm

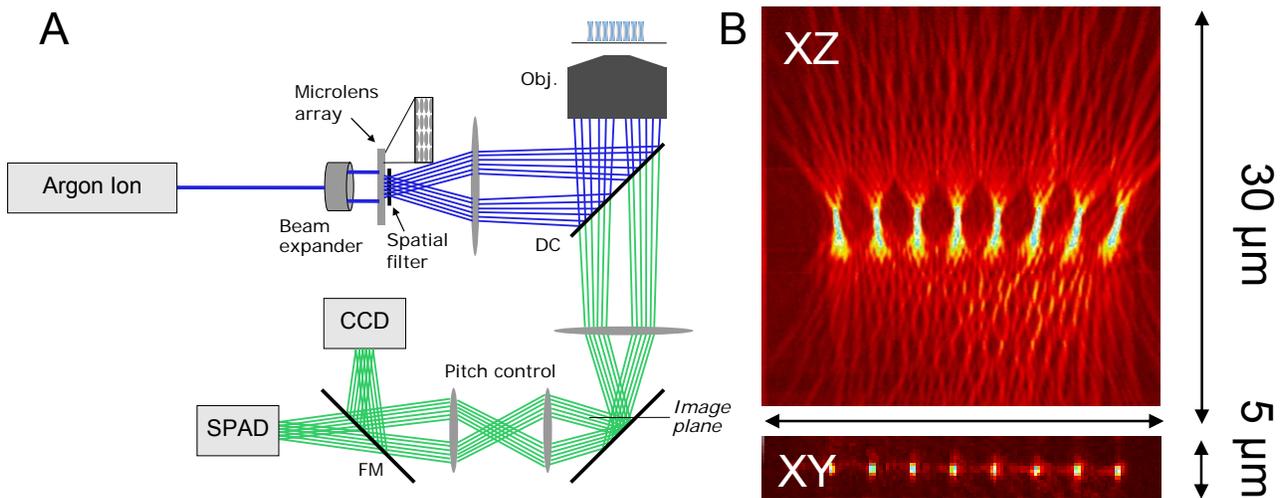


Fig. 4: A) Setup schematic using a microlens array to generate a multispot excitation pattern. Blue segments indicate the path of excitation light. Green lines indicate the path of emitted photons towards the detector (CCD for pattern pitch and size characterization, SPAD array for data acquisition. DC: dichroic mirror, Obj: objective lens, FM: flippable mirror. B) Longitudinal (XZ) and transverse intensity profile of the 8 x 1 excitation spot pattern recorded by stage-scanning confocal microscopy of an isolated 100 nm diameter bead. Gradient color palette: dark red indicates low count rate, while yellow, green, blue and then white indicate increasing brightness. The distance between spots is 3 μ m.

diameter fluorescent bead obtained by raster-scanning the bead through the excitation spots (Fig. 4B). Since the detector used for this image acquisition had a large area (3 mm diameter), no clipping of the emission pattern occurred due to the detector. Therefore, the resulting intensity distribution is directly proportional to the excitation intensity profile. As can be seen in Fig. 4B, a slight tilt of the excitation PSF away from the optical axis occurs away from the center spot due to spherical aberrations created by some of the optical elements. This effect is amplified when the number of spots is increased (e.g. for a 32x32 pattern). In later experiments, we were able to reduce these aberrations using a larger recollimating lens in the excitation path (data not shown). The effect of a tilted PSF is mostly noticeable in FCS experiments as will become obvious in the next section, due to the departure of the excitation PSF from the ideal 3D Gaussian model assumed in the fits. It is less of an issue for single-molecule burst detection, where one merely counts the number of detected photons. However, in this case, there is still an adverse effect of an aberrated PSF, as it reduces the collection efficiency and thus the detected signal.

4.2 FCS of bead samples

Fig. 5A shows an example of raw ACFs obtained with the 1x8 SPAD array developed by the Cova group. The most striking feature is the discrepancy in the ACF amplitudes, which according to Eq. 3 should be inversely proportional to the concentration times the sampling volume. Since the sample concentration is common to all detectors, this discrepancy indicates either a difference in the excitation volumes (i.e. intensity profiles) defined by each excitation spot

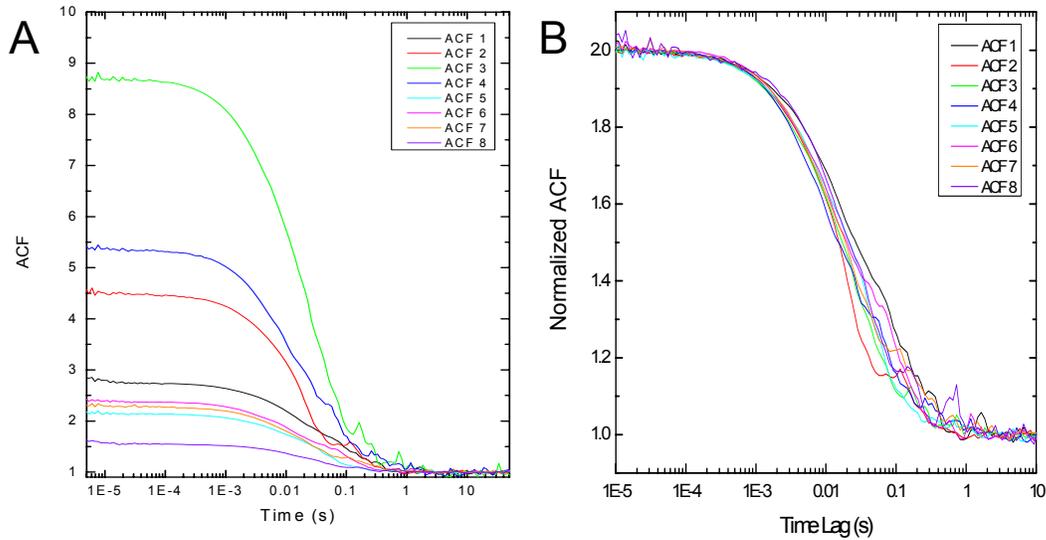


Fig. 5: Autocorrelation functions (ACF) of a 100 nm bead sample obtained with a 1x8 SPAD array. A: raw ACF, B: Normalized ACF after fitting to the 2-dimensional version of Eq. 3.

or a difference in the collection efficiency of each SPAD. As shown in Fig. 4B, excitation profiles are indeed different from spot to spot, which could account for part of this discrepancy. The collection efficiency can also decrease if the individual SPADs are not perfectly aligned with the image of the excitation volume. For instance, this can be the case if the pitch of the spots image on the detector plane does not match the detector's pitch, or if there is a slight tilt of the image with respect to the detector array. With the microlens array arrangement used in this experiment, a perfect alignment of the detector and the excitation pattern turned out to be difficult to achieve because of the many degrees of freedom involved. In particular, the fact that the center pixels 4-5 have smaller amplitude than the more peripheral pixel #3 points strongly toward this explanation. Those are the spots with the least aberration, as seen in Fig. 4B, and thus should correspond to the smallest excitation volume, hence the largest amplitude. Note that the influence of different dark count levels for individual SPADs of the array on the ACF amplitude (Eq. 4) could also contribute in part in the difference in amplitudes, although the dark count levels in this detector were insignificant (see Fig. 7). Whatever the exact origin of this amplitude difference, it does not affect the other fitting parameter (diffusion constant) as significantly, as can be seen in Fig. 5B. Further experiments performed with a LCOS SLM as a pattern generator instead of the microlens array allowed a much better matching between the excitation pattern and the detector geometry and eliminated most of these discrepancies.

4.3 Single-molecule burst detection

Single-molecule burst detection experiments use the same configuration as FCS experiments, but have usually to deal

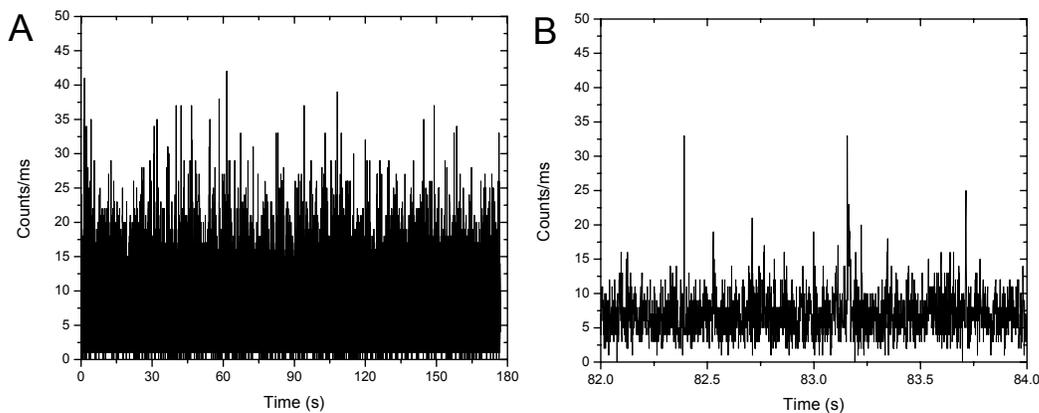


Fig. 6: Cy3B-DNA time trace recorded by a single pixel of the 1x8 SPAD array (Bin: 1 ms). Hundreds of brief bursts of a few 10 to 100 photons are detected above the background. A: full trace, B: 2 s detail.

with much smaller signals. Therefore perfect alignment is critical, as even a slight detector offset can result in reduction of the signal down to the level of background. Fig. 6 shows an example time trace recorded by SPAD 3 in the 1x8 array (the one with the largest amplitude in the FCS experiments of Fig. 5), where a background of ~ 7 kHz (dark counts plus sample background fluorescence and scattering) does not prevent the detection of hundreds of bursts during the 3 min acquisition. Experimentally, we observed that the maximum burst size detected by each channel varied similarly to the ACF amplitude measured in the FCS measurement reported previously, confirming that most of the difference between SPAD signals was due to unsolved alignment issues. Further experiments performed with the LCOS SLM pattern generator eliminated most of these pixel-to-pixel variations by simplifying the alignment procedure (data not shown).

5. DISCUSSION

The preliminary experiments reported here performed with a 1x8 SPAD array as well as others performed with a 32x32 SPAD array detector or on-going ones performed with an 8x8 HPD detector in our lab emphasize the need for further improvements in at least three major areas to fully achieve the goal of high-throughput single-molecule spectroscopy measurements.

5.1 Pattern generation and detector alignment

Our use of an off-the-shelf microlens array to generate a pattern of multiple excitation spots was motivated by cost and simplicity. Although the resulting pattern had reasonably good quality, one of the main disadvantages of this approach turned out to be the demanding requirements in terms of detector alignment. Indeed, the pattern's image needed not only to be properly scaled and aligned but also properly oriented. This can be performed partly by orienting the microlens array itself (in the excitation path), but any (mechanical) modification in orientation needs to be followed by a detector readjustment, ending up in a very tedious alignment process, prone to thermal and mechanical drift. For these reasons, we turned to a more expensive and sophisticated approach based on a LCOS SLM pattern generator, which will be described in a future publication. This technical solution allowed complete flexibility in terms of pattern geometry control and in principle allows individual spot control and modification independently from the other spots. However, any solution will eventually be limited by two major constraints: pattern homogeneity and available power. As was illustrated in Fig. 4B, spherical aberrations are most pronounced for spots further apart from the optical axis. Since spots need to be separated by a minimum amount in the object plane to avoid optical crosstalk, this problem is exacerbated as the size of the array is increased. In our experience, non-uniform pattern or presence of significant out-of-focus excitation are the two main sources of imperfection affecting the final experimental results. Additionally, expanding the laser beam to cover the pattern-generation device results in a reduced excitation power per spot. Too low an excitation eventually prevents single-molecule detection due to insufficient burst size. For instance, a 32x32 detector array will typically require 1,000 times more excitation power (and in general more due to diverse sources of loss) than a typical single spot experiment. Other, more efficient excitation geometries, such as zero-mode waveguides may be preferable in this situation²¹.

5.2 Detector performance

The individual SPADs of the 1x8 array used in these experiments had similar performance to commercial single SPADs. In particular, they have limited sensitivity in the red part of the spectrum. It might be possible for some detector processes to increase their sensitivity at the expense of larger noise (dark count levels and afterpulsing percentage). This could be tolerable to a certain extent depending on the application, as sample background is in general larger than the noise levels of current detectors (and not an obstacle to SMS).

5.3 Data processing

Data sets acquired during these experiments turned out to be quite large and required considerable processing time. To achieve high-throughput SMS, significant progresses will need to be accomplished in online data reduction (e.g. burst detection) or compression as well as parallelized computation. We expect that online FPGA processing and off-line GPU processing will be able to provide the necessary performance to make this goal achievable in the near future.

6. ACKNOWLEDGMENTS

This work was supported in part by NIH grants R01 GM084327 & R01 EB006353 and NSF grant DBI-0552099. The work at Politecnico di Milano was supported in part by EC agreement 232359 (PARAFLUO) FP7-SME-2008-1. XM

thanks Drs. Ted Laurence and Michael Culbertson for providing C codes to implement multitau algorithms for time stamped⁷ and binned data¹⁸, respectively.

REFERENCES

- [1] S. Weiss, "Fluorescence spectroscopy of single biomolecules", *Science* 283, 1676-1683 (1999).
- [2] S. Weiss, "Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy", *Nature Structural Biology* 7, 724-729 (2000).
- [3] M. Born and E. Wolf, [Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light], Cambridge University Press: Cambridge (1999)
- [4] O. Krichevsky and G. Bonnet, "Fluorescence correlation spectroscopy: the technique and its applications", *Reports on Progress in Physics* 65, 251-297 (2002).
- [5] R. Verberk and M. Orrit, "Photon statistics in the fluorescence of single molecules and nanocrystals: Correlation functions versus distributions of on- and off-times", *Journal of Chemical Physics* 119, 2214-2222 (2003).
- [6] K. Schätzel, M. Drewel and S. Stimac, "Photon correlation measurements at large lag times: improving statistical accuracy", *Journal of Modern Optics* 35, 711-718 (1988).
- [7] T. A. Laurence, S. Fore and T. Huser, "Fast, flexible algorithm for calculating photon correlations", *Optics Letters* 31, 829-831 (2006).
- [8] P. Kask, R. Günther and P. Axhausen, "Statistical accuracy in fluorescence fluctuation experiments", *European Biophysical Journal* 25, 163-169 (1997).
- [9] S. Saffarian and E. L. Elson, "Statistical Analysis of Fluorescence Correlation Spectroscopy: The Standard Deviation and Bias", *Biophysical Journal* 84, 2030-2042 (2003).
- [10] P. Kask, K. Palo, D. Ullmann and K. Gall, "Fluorescence-intensity distribution analysis and its application in biomolecular detection technology", *Proceedings of the National Academy of Sciences of the United States of America* 96, 13756-13761 (1999).
- [11] K. Palo, U. Metz, S. Jager, P. Kask and K. Gall, "Fluorescence intensity multiple distributions analysis: Concurrent determination of diffusion times and molecular brightness", *Biophysical Journal* 79, 2858-2866 (2000).
- [12] T. A. Laurence, A. N. Kapanidis, X. X. Kong, D. S. Chemla and S. Weiss, "Photon arrival-time interval distribution (PAID): A novel tool for analyzing molecular interactions", *Journal of Physical Chemistry B* 108, 3051-3067 (2004).
- [13] E. Nir, X. Michalet, K. M. Hamadani, T. A. Laurence, D. Neuhauser, Y. Kovchegov and S. Weiss, "Shot-noise limited single-molecule FRET histograms: Comparison between theory and experiments", *Journal of Physical Chemistry B* 110, 22103-22124 (2006).
- [14] M. Gösch, A. Serov, T. Anhut, T. Lasser, A. Rochas, P.-A. Besse, R. S. Popovic, H. Blom and R. Rigler, "Parallel single molecule detection with a fully integrated single-photon 2x2 CMOS detector array", *J. Biomed. Opt.* 9, 913-921 (2004).
- [15] M. Burkhardt and P. Schwille, "Electron multiplying CCD based detection for spatially resolved fluorescence correlation spectroscopy", *Optics Express* 14, 5013-5020 (2006).
- [16] X. Michalet, A. Cheng, J. Antelman, M. Suyama, K. Arisaka and S. Weiss, "Hybrid photodetector for single-molecule spectroscopy and microscopy", *Proceedings of SPIE* 6862, 68620F (2008).
- [17] F. Guerrieri, S. Tisa and F. Zappa, "Fast Single-Photon Imager acquires 1024 pixels at 100 kframe/s", *Proceedings of SPIE* 7249, 72490U (2009).
- [18] M. J. Culbertson and D. L. Burden, "A distributed algorithm for multi-tau autocorrelation", *Review of Scientific Instruments* 78, (2007).
- [19] I. Rech, D. Resnati, S. Marangoni, M. Ghioni and S. Cova, "Compact-eight channel photon counting module with monolithic array detector", *Proceedings of SPIE* 6771, 677113 (2007).
- [20] N. K. Lee, A. N. Kapanidis, Y. Wang, X. Michalet, J. Mukhopadhyay, R. H. Ebright and S. Weiss, "Accurate FRET Measurements within Single Diffusing Biomolecules Using Alternating-Laser Excitation", *Biophysical Journal* 88, 2939-2953 (2005).
- [21] M. J. Levene, J. Korlach, S. W. Turner, M. Foquet, H. G. Craighead and W. W. Webb, "Zero-mode waveguides for single-molecule analysis at high concentrations", *Science* 299, 682-686 (2003).