Development of new photon-counting detectors for single-molecule fluorescence microscopy

X. Michalet¹, R. A. Colyer¹, G. Scalia¹, A. Ingargiola¹, R. Lin¹, J. E. Millaud¹, S. Weiss¹, Oswald H. W. Siegmund³, Anton S. Tremsin³, John V. Vallerga³, A. Cheng², M. Levi², D. Aharoni², K. Arisaka², F. Villa⁴, F. Guerrieri⁴, F. Panzeri⁴, I. Rech², A. Gulinatti⁴, F. Zappa⁴, M. Ghioni⁴ and S. Cova⁴

¹Department of Chemistry and Biochemistry, and ²Department of Physics and Astronomy, UCLA, Los Angeles, CA 90095-1547, USA
³Space Sciences Laboratory, UCB, Berkeley, CA 94720, USA
⁴Dipartimento di Elettronica ed Informazione, Politecnico di Milano, Milano, Italy

Two optical configurations are commonly used in single-molecule fluorescence microscopy: point-like excitation and detection to study freely diffusing molecules, and wide field illumination and detection to study surface immobilized or slowly diffusing molecules. Both approaches have common features, but also differ in significant aspects. In particular, they use different detectors, which share some requirements but also have major technical differences. Currently, two types of detectors best fulfill the needs of each approach: single-photon-counting avalanche diodes (SPADs) for point-like detection, and electron-multiplying charge-coupled devices (EMCCDs) for wide field detection. However, there is room for improvements in both cases. The first configuration suffers from low throughput owing to the analysis of data from a single location. The second, on the other hand, is limited to relatively low frame rates and loses the benefit of single-photon-counting approaches. During the past few years, new developments in point-like and wide field detectors have started addressing some of these issues. Here, we describe our recent progresses towards increasing the throughput of single-molecule fluorescence spectroscopy in solution using parallel arrays of SPADs. We also discuss our development of large area photon-counting cameras achieving subnanosecond resolution for fluorescence lifetime imaging applications at the single-molecule level.

1. Introduction

Single-molecule-sensitive detection techniques have found increasing domains of applications since their original developments in the early 90s [1–6]. In basic science, observing one molecule at a time allows elucidating the steady-state characteristics of heterogeneous mixtures by enumerating and precisely characterizing each molecule of a sample. This type of ‘static disorder’ study obviously requires repeating single-molecule observations a large number of times in order to have a statistically significant distribution of measurements. The same approach can also, in principle, be used to study ‘dynamic disorder’ within a sample of identical molecules evolving stochastically and independently from one another. Single-molecule analysis is also ideally adapted to detect rare events or rare molecular conformations in a sea of irrelevant measurements, provided, however, that a large enough total number of single molecule can be detected.

Single molecules can be detected using different physical effects [7], but optical means in particular have the advantage of being the least invasive and usable in different sample states (gaseous, liquid, solid) or experimental arrangements (flow, diffusion, surface, etc.) [1–3]. Several types of...
spectroscopic signatures can be used to distinguish different molecular species and most can be used at the single-molecule level: fluorescence emission intensity, spectrum, polarization, lifetime, Raman spectrum, etc. Whichever modality is used, successful single-molecule optical detection requires two criteria to be fulfilled:

(i) the signal-to-noise ratio (SNR) corresponding to a single molecule needs to be large enough;
(ii) signals from different molecules need to be distinguishable.

The first criterion, which ensures that single molecules can be detected, can be fulfilled by optimizing different experimental parameters, although this can, in practice, be challenging because some of these parameters are correlated. For instance, the SNR depends on signal (proportional to the detected count rate $s$) and detector readout noise ($N_b$), but also background (count rate $b$). Signal and background are proportional to the measurement duration $\tau$, whereas readout noise is usually constant or, at worst, increases with readout rate. A detailed discussion of the SNR and signal-to-background ratio (SBR) and of the different areas (sample, optics, detector) to consider for successful single-molecule detection is presented in the electronic supplementary material (text and figure). We limit ourselves in the following to detector considerations.

From a detector point of view, the critical parameter is sensitivity, quantified by the wavelength-dependent quantum efficiency $Q_i$. Quantum efficiency (QE) of detectors discussed in the review is shown in figure 1. Point-detectors (which are photon-counting detectors) have maximum QE ranging from 40 to more than 70 per cent, with peak locations varying between 450 and 650 nm (figure 1a). We will compare the respective merits of these detectors in §2. The picture is even more contrasted for current wide field imagers, where QE can be larger than 90 per cent for charge-coupled device (CCD) or electron-multiplying charge-coupled device (EMCCD) cameras but as low as a few per cent for some of the older generation wide field photon-counting detectors discussed in a later section (figure 1b). A good QE helps increase the signal count rate ($s = Q_i b$, where $i$ is the incident count rate impinging on the detector).

Sensitivity can also depend on other detector design characteristics: fill factor (sensitive fraction of the detector area), which can be in principle increased with microlenses, efficiency of photoelectron detection by the readout electronics, etc. For this reason, it is often preferable to quote the photon-detection efficiency (PDE), which is the product of all these efficiencies, rather than the QE.

It is also important to choose a detector with small read-out noise, as readout noise will eventually dominate the recorded signal at short integration time and decrease the SNR. If present, readout noise should be minimized, or at least compensated by additional signal gain (such as in intensified CCDs or EMCCDs). We will not discuss this latter strategy, as it is irrelevant for the photon-counting detectors discussed here, but it is worth mentioning that it comes at a cost: the gain generates additional signal variance, quantified as an excess noise factor (ENF), which translates into reduced SNR [8,9]. Finally, large dark count rate can affect some detectors, reducing the SBR and making it difficult to detect single molecules.

The second criterion (single-molecule separability) depends a lot on the details of each experiment, but generally, molecules can be distinguished only if they have different optical signatures and/or they are sufficiently separated spatially and/or temporally [10]. Detector characteristics play a role in achieving these goals, as we will discuss after briefly examining typical geometries encountered in single-molecule fluorescence experiments. Section 2 will discuss confocal geometries and detectors adapted to measurements on freely diffusing molecules in solution, whereas §3 will examine wide field detection geometries and their

Figure 1. Quantum efficiencies (QEs) of detectors discussed in the text. Curves were provided by the manufacturers. (a) Point-detectors, (b) wide field detectors. τ-SPAD, SPAD module sold by PicoQuant GmbH, based on a thick reach-through SPAD manufactured by Laser Components. SPCM-AQR, thick reach-through SPAD manufactured by Excelitas Technologies. RE-SPAD, red-enhanced epitaxial SPAD manufactured by Micro Photon Devices (MPD). MPD, former version of the epitaxial thin SPAD manufactured by MPD. HPD, hybrid photo detector prototype manufactured by Hamamatsu Photonics based on a GaAsP photocathode. Commercial devices including this type of modules are now available from different vendors. CMOS: 32 × 32 CMOS SPAD array prototype developed by Politecnico di Milano. EMCCD: Cascade 512B camera (Photometrics). sCMOS, scientific CMOS camera (model: neo) manufactured by Andor Technology. GaAsP, typical QE of a GaAsP photocathode manufactured by Photonis. S20, H33D Gen 1 sensitivity based on a multialkali S20 photocathode manufactured by UCB. RE-SPAD, red-enhanced epitaxial SPAD manufactured by Micro Photon Devices (MPD). MPD, former version of the epitaxial thin SPAD manufactured by MPD. HPC, hybrid photo detector prototype manufactured by Hamamatsu Photonics based on a GaAsP photocathode. Commercial devices including this type of modules are now available from different vendors. CMOS: 32 × 32 CMOS SPAD array prototype developed by Politecnico di Milano. EMCCD: Cascade 512B camera (Photometrics). sCMOS, scientific CMOS camera (model: neo) manufactured by Andor Technology. GaAsP, typical QE of a GaAsP photocathode manufactured by Photonis. S20, H33D Gen 2 sensitivity, based on a SuperGen2 photocathode manufactured by Photonis. S20, H33D Gen 1 sensitivity based on a multialkali S20 photocathode manufactured by UCB.
applications. Section 4 will discuss new detectors for diffusing molecule measurements in multispot geometries. Section 5 will describe recent advances in wide field photon-counting detectors developed for single-molecule imaging and spectroscopy. We conclude this review with a brief overview of future prospects for the field.

2. Single-molecule detection and spectroscopy in point-like geometries

(a) Point-like excitation and detection

A point-like geometry has advantages and drawbacks for single-molecule detection. On the one hand, by minimizing the volume of the sample in which excitation takes place, it reduces background sources and the number of simultaneously excited individual molecules. Ideally, the average number of molecules in the excited volume should be much smaller than 1 in order to facilitate separability of individual molecules (see the electronic supplementary material, figure S1A). This geometry also simplifies the optical set-up as the requirement is to image only a single point in the sample onto a small area detector (single-pixel or point-like detector). Because these two points are located on the optical system axis, this relaxes somewhat aberration correction requirements in the optics.

On the other hand, this arrangement constrains the detected molecule to be located precisely at the observation point. For static molecules, this necessitates a high-precision scanning stage or beam-scanning optics in order to first image the sample and localize single molecules, and then sequentially park the excitation spot on each identified molecule for further study [11,12]. This is useful when high temporal resolution fluorescence intensity time traces are desired, because point-detection allows using single-photon-counting detectors with time-stamping capabilities (discussed later). However, this approach is slow and requires immobilization of the molecules of interest. For this reason, point-excitation is mostly used for fluid samples in which free molecules undergo two-dimensional or three-dimensional diffusion, allowing a fixed excitation volume configuration to be used: single molecules randomly diffuse in and out of the excitation volume, the duration and frequency of single-molecule detection depending on both detection volume and sample concentration.

A standard approach to achieve point-like excitation involves tightly focusing an expanded and collimated laser beam in the sample using a high numerical aperture lens such as a microscope objective lens [11]. The achievable diffraction-limited volume (defined using the full-width at half-maximum (FWHM) of the point-spread function (PSF) of the instrument) is then of the order of [13]:

$$V_X \sim \lambda^4 \text{NA}^{-4},$$

(2.1)

where \(\lambda\) is the excitation wavelength, and NA is the numerical aperture of the lens. Using standard values for these parameters (\(\lambda = 532\) nm, \(\text{NA} = 1.2\) for a water immersion objective lens), one obtains a typical diffraction-limited excitation volume of a fraction of a femtolitre (1 fl = \(10^{-15}\) litre = \(1\) \(\mu\)m\(^3\)). Different techniques have recently been developed to further reduce excitation volume and thus increase the range of accessible concentrations to larger values (e.g. near-field excitation [14], stimulated emission depletion [15], zero-mode wave guides (ZMWGs) [16], etc.). We will limit ourselves in this section and §4 to confocal microscopy, which uses the same objective lens to focus excitation light and collect emitted fluorescence [13]. Its advantage is its simplicity and its ability to detect signals relatively deep into a liquid sample (approx. \(50\) \(\mu\)m) through a standard thickness glass coverslip [17].

Creating a small excitation volume is not the most challenging step in single-molecule experiments. Efficient collection of the weak light emitted by individual molecules is also demanding. Owing to the limited NA of conventional lenses (which translates in a collection solid angle \(\Omega < \pi r^2\)), the collection efficiency \(\eta\) is usually at best a few per cent of the total emitted signal:

$$\eta \leq 1/2 \left(1 - \sqrt{1 - (\text{NA}/n_0)^2}\right),$$

(2.2)

where \(n_0\) is the refraction of the buffer. This fraction is further decreased by losses in additional relay optics and spectral filtering elements.

It is in particular essential to properly match the size of the image of the emission volume (emission PSF) to that of the detector: too small a detector (or equivalently too large a PSF magnification, \(M\)) will clip the image and result in lost photons (see the electronic supplementary material, figure S1B). On the other hand, the detector size cannot be increased arbitrarily, as otherwise background signal from outside the region of interest will be collected (additionally, detector dark count increases with detector area), reducing the SBR. We will get back to this issue of detector versus PSF size when discussing high-throughput single-molecule applications.

When an adequate trade-off between these parameters is achieved, distinct photon bursts can be detected as individual molecules transit across the excitation/detention volume (see the electronic supplementary material, figure S1C). For diffusing molecules, burst duration depends on molecule size, solvent viscosity and excitation volume. Typical values range from a few \(10\) \(\mu\)s to a few milliseconds, with an approximately exponential distribution of burst durations [18].

Burst intensity (or burst size), i.e. the total number of detected photons during the transit of a single molecule through the excitation volume, depends on excitation power, absorption cross section and fluorescence quantum yield of the molecule as well as the total detection efficiency [18]. Burst sizes can reach up to a few hundred photons, but as for burst durations, their typical distribution is quasi-exponential, resulting in a large fraction of bursts having a few dozen photons or less [18]. Because small bursts have small SNR, they increase the variance of quantities computed by averaging data from many single-molecule bursts [19,20] and are generally rejected. A simple way to increase burst size would seem to be increasing the excitation power. However, this is not always possible, and advantageous only up to a certain point: owing to emission saturation and photo-bleaching, background eventually increases faster than signal, reducing SBR and SNR [9].

For obvious reasons, the measurement temporal resolution needs to be better than the shortest burst duration \(\tau\) one wants to detect, as otherwise the detected signal may be dominated by background, or close successive bursts may become indistinguishable. For photon-counting detectors such as those
discussed later, the measurement resolution is user-adjustable, as the detectors and their typical readout electronics are in general capable of providing a time-stamp for each photon. However, some read out electronics output data in a binned format, where the bin duration is either fixed or defined by the user. Because very short bursts are also in general very small and of no interest for further analysis, a few hundred microseconds bin duration is usually sufficient for most experiments.

Picosecond resolution may be needed for time-correlated measurements, in which each photon is timed with respect to the exciting laser pulse or other detected photons [21]. This resolution is achieved by determining single-photon pulses arrival time using dedicated time-correlated single-photon-counting (TCSPC) electronics. Most photon-counting detectors provide a few to a few tens of nanoseconds long electrical signal for each detected photon, which is then detected and time-stamped with nanosecond or better resolution using additional electronics (e.g., constant fraction discriminator for variable amplitude pulses and simple edge detection for low-jitter logic signal pulses). Laser excitation is similarly timed with great accuracy and both are compared using time-to-amplitude converter or time-to-digital converter (TDC) electronics. Because this high-accuracy time interval measurement can be done over limited time windows (10–100 ns), each photon timing information generally comprises two components: (i) a macrotime, which provides the time elapsed since the beginning of the measurement and is generally obtained using a digital clock with a few tens of nanoseconds resolution, and (ii) a nanotime (also called microtime in the literature, although this terminology makes it easily confused with the macrotime component), precisely timing the photon arrival with respect to the previous or next laser pulse.

(b) Single-molecule burst detection and analysis
Detecting and characterizing bursts (by arrival time, duration and intensity) is only a first step, which could in principle be considered part of the data acquisition process. A detailed discussion of the parameters involved in this step is provided in the electronic supplementary material (text and figures).

Single-molecule burst detection can be used to perform more interesting measurements than mere counting when combined with various spectroscopic techniques. As is well known, the conformation of a molecule (protein, nucleic acid, etc.) or respective location of two interacting molecules can be studied by monitoring the distance between two dyes attached to specific sites of the molecule(s) using fluorescence resonant energy transfer (FRET) [22–24]. When the two dyes have overlapping absorption and emission spectra and are in close proximity, intermolecular non-radiative (resonant) energy transfer can occur by dipole–dipole interaction, with an efficiency varying with the sixth power of the distance between the two dyes [25,26]. This phenomenon has been extensively used in bulk, but its use at the single-molecule level (smFRET) has exploded since its first demonstration on surface-immobilized molecules [27] and on freely diffusing molecules in solution [28]. In standard smFRET experiments, doubly labelled molecules or molecular complexes are excited by a single laser exciting the donor dye. The fluorescence signals from the donor and acceptor dyes are collected simultaneously in their respective emission spectral band by two detectors (figure 2a,b) [28]. After identification of individual bursts, the donor and acceptor signals (I_D and I_A) resulting from donor excitation, corrected for background and other contaminating signals, are used to compute the FRET efficiency E:

\[
E = \frac{I_A}{I_A + \gamma I_D},
\]

where \(\gamma\) is a correction factor accounting for the different quantum yield and detection efficiency of both fluorophores [29]. Alternatively, when using pulsed-laser excitation and TCSPC electronics, the donor lifetime \(\tau_D\) gives directly access to the FRET efficiency via:

\[
E = 1 - \frac{\tau_D/A}{\tau_D},
\]

where \(\tau_D\) is the donor lifetime in the absence of any acceptor (figure 2c,d). Note that although the latter approach is generally thought to be less sensitive to background or other signal contamination, this in fact is only partially the case [30]. When using the TCSPC approach, it is just as critical as in the intensity-based method to perfectly understand the contribution of background and non-donor photons to the so-called donor decay histogram. Moreover, the analysis of this decay in terms of a single lifetime is not always possible, which raises additional interpretation issues. Both methods should obviously report the same value, which is related to the distance between donor and acceptor dyes via the well-known Förster equation [25]:

\[
E = \left(1 + \left(\frac{R}{R_0}\right)^6\right)^{-1},
\]

where \(R_0\) is a distance of the order of a few nanometres characterizing the donor/acceptor dye pair and its environment. In the case of intensity-based FRET measurements, E values measured from many multiple single-molecule bursts (usually several hundred or thousand bursts) are then histogrammed to identify populations of molecules characterized by specific E values, as well as their respective fraction. TCSPC-based measurements can be analysed similarly [31,32].

Refinements of this general technique have been developed to carefully analyse the contribution of shot noise to these distributions, or help identify molecules labelled with only a donor or acceptor dye, using alternating laser excitations (ALEX) [33–36], and progresses are continuously made in order to disentangle artefacts affecting the extraction of true distances between sites of interest in a molecule [37–40]. A new natural extension of the two-colour FRET experiment, three or more colour smFRET/ALEX measurements allow studying more complex biochemical molecular assemblies, conformations and interactions [41,42]. Their discussion is however beyond the scope of this review (see [43] for a review).

(c) Fluorescence correlation spectroscopy
A related but distinct experimental regime, illustrated in electronic supplementary material, figure S2A, is encountered when the point-like excitation volume contains one or a few molecules on average at any time. In a confocal microscope, this situation corresponds to a sample concentration of a few nanomolar (nM). In this regime, individual molecule bursts cannot be distinguished anymore and are replaced by a highly fluctuating signal \(l(t)\) centred on an average
Analysis of the autocorrelation function (ACF) of these fluctuations $\langle \delta I(t) \delta I(t') \rangle = \langle I(t) I(t') \rangle$ [44,45] can yield physical as well as photochemical information (molecule size and concentration, blinking or binding/unbinding rates) about the diffusing molecules (see the electronic supplementary material, figure S2C). In the simplest case of pure diffusion through a Gaussian volume of lateral standard deviation $\sigma_{XY}$ and vertical standard deviation $\sigma_Z = \omega \sigma_{XY}$, the theoretical expression of the fluctuations ACF is given by [45]:

$$G(\tau) = \frac{1}{CV} \left( 1 + \frac{\tau}{\tau_D} \right)^{-1} \left( 1 + \frac{\tau}{\omega \tau_D} \right)^{-1/2},$$

(2.6)

where $C$ is the sample concentration, $V$ the excitation/detection volume (depending on the exact detection geometry), $\tau_D = \sigma_{XY}^2/D$ the diffusion time through the excitation volume and $D$ the diffusion coefficient.

This technique gives exploitable information only at very low concentration, the relative amplitude of the ACF being inversely proportional to the average number of molecules in the excitation/detection volume $V$. On the other hand, too low a concentration (e.g. single-molecule concentration) results in most of the signal coming from uncorrelated inter-bursts signal, leading to a noisy ACF. In particular, background has a detrimental effect on the ACF amplitude, as expressed in the following relation [46]:

$$G_B(\tau) = \frac{1}{1 + \frac{SBR}{C_0}} G(\tau),$$

(2.7)

where $G$ is the ACF in the absence of background and $G_B$ the ACF in the presence of background.

In principle, fluorescence correlation spectroscopy (FCS) techniques have the advantage of simplicity owing to the existence of plug-and-play hardware correlators and the existence of well-established data analysis methods. They are also capable of detecting populations of molecules with sufficiently different sizes, the ACF of a mixture being a weighted sum of the ACFs of each population [45]. In practice, however, it can be extremely demanding to perfectly characterize departure of the experimental system from the ideal situations assumed in theoretical models. For instance, a measurement as simple as a diffusion coefficient can be affected by a number of experimental problems such as dye saturation or imperfect illumination and detection geometries that can severely compromise the reliability of the extracted results (see, for instance, discussions in recent studies [47,48]).

As for single-molecule methods, multiple spectroscopic channels can be cross-correlated, including donor and acceptor
emission in the case of a FRET experiment [49]. Many sophisticated analysis schemes going beyond mere auto- or cross-correlation analysis of the signal have been developed, making this technique a very versatile approach to study millisecond to microsecond timescale inter- or intramolecular dynamics [50–52].

Unsurprisingly, considering the similarity to single-molecule methods, comparable requirements of detector size and excitation/detection volume matching [47], SBR, SNR and temporal resolution mentioned earlier apply for this experimental approach, although the theoretical analysis of this problem is rather cumbersome [53–57]. The main result of interest for the following discussion is that the SNR of FCS measurements is always proportional to the square root of the measurement duration.

It is thus time to briefly examine the respective merits of point-detectors currently available to the single-molecule spectroscopist.

(d) Detectors used in single-point geometry single-molecule experiments

Although early single-molecule fluorescence experiments were performed using photomultiplier tubes (PMTs) [58,59], the introduction of single-photon-counting avalanche diode (SPAD) [60] for single-molecule detection [61–63] and FCS [64,65] significantly improved the detection efficiency in the visible range and replaced high-voltage, bulky detectors by more compact and simpler devices, permitting a rapid development of the field. Nowadays, typical detectors used in point-like geometry experiments are: (i) SPADs, (ii) PMTs [66] or (iii) the recently introduced hybrid photodetectors (HPDs) [67]. The latter two are high-voltage and large sensitive area devices with a QE depending on the photocathode material. The best efficiency is obtained with GaAsP and reaches approximately 45 per cent in the visible (green curve in figure 1a). This sensitivity is comparable to that of standard shallow-junction epitaxial (thin) SPAD technology [68] (orange curve), which has the advantage of using low voltage and results in ambient-light-resistant devices. This QE is however lower than that of the thick reach-through SPADs (black and blue curves) [69], or the new generation of red-enhanced thin SPAD devices (red curve) discussed further in the following.

This QE hierarchy is however inadequate to classify these detectors, as other parameters can come into play (e.g. ability to handle large photon flux without irreversible damage or sensitive detection area). In particular, although all these detectors have in general sufficient temporal resolution for most time-resolved single-molecule experiments, more demanding applications such as photon antibunching measurements [66] or very short lifetime measurements as encountered in electron transfer processes [70] may in some cases benefit from detector resolution better than 100 ps. Until recently, the best sensitivity SPAD counting modules (thick reach-through SPAD, model SPCM-AQRH, Excelitas Technologies) suffered from a wavelength-dependent instrument response function (IRF) FWHM of 200–600 ps with count rate-dependent walk [66]. Improved properties seem to characterize similar detectors introduced by Laser Components [71], although the relatively large and thick sensitive area of these detectors seems to limit the achievable IRF FWHM to a few 100 ps.

An advantage of the thin SPAD technology developed by Politecnico di Milano is the much narrower IRF, which can reach a few tens of picoseconds for 50–200 μm diameter detectors [72,73]. As mentioned earlier, however, until recently these detectors suffered from lower QE than the thick reach-through SPADs in the red region of the spectrum, due mainly to their thinner absorption region. New developments have solved this problem [74], resulting in red-enhanced SPAD (RE-SPAD) detectors with good sensitivity for single-molecule detection across the whole visible range and excellent timing response (as illustrated in the comparison of thick reach-through SPADs and the new RE-SPAD for single-molecule FRET experiments shown in figure 2). An additional advantage of this technology is its compatibility with array geometry, as discussed in a later section.

PMTs have temporal resolution comparable to that of thick SPADs [66], have generally lower QE and are high-voltage devices with inherent fragility and are therefore rarely used in single-molecule experiments, unless large detection area is needed (as for instance, for two-photon excitation non-descanned detection [13]). HPDs are similar to PMTs in terms of sensitivity, large detection area and high-voltage characteristics, but have better temporal resolution (approx. 100 ps) and no noticeable afterpulsing down to approximately 100 ns, which makes them attractive detectors for some applications, such as afterpulsing-free single-detector ACF analysis [67,75]. It is worth noting that they occasionally suffer from artefacts owing to the ionization of residual atoms in the vacuum tube by X-ray emission from the silicon avalanche photodiode (APD) [76].

(e) Common limitations of single-point-detection geometries

The single-point-excitation/single-point-detection design used in all the techniques described so far results in highly sensitive detection but requires in general long acquisition times in order for sufficient statistics to be accumulated. This is compatible with studies of equilibrium dynamics such as conformational fluctuations occurring at timescales much shorter or longer than the diffusion time and, with some additional efforts, at intermediate timescales [20,77]. However, fast irreversible reactions cannot be studied by this approach, as they are over before enough single-molecule bursts have been acquired. In other words, only reactions with timescales significantly longer than the minimum duration of a measurement (a few minutes or more) can be conveniently studied in this geometry [78]. Microfluidic devices can to some extent address this problem, as discussed in the electronic supplementary material.

Section 3 will examine the simplest solution consisting in using a wide field detection approach to allow the simultaneous detection of multiple single molecules. Because it is mostly limited to surface-immobilized molecules or molecules confined into a plane, §4 will then describe another parallelization approach based on multispot excitation compatible with the study of freely diffusing molecules.

3. Single-molecule imaging

There are many experimental situations where the kind of point-excitation/detection earlier-mentioned methods
becomes inefficient. For instance, samples in which molecules diffuse slowly (as encountered for instance in live cell membranes) or are immobilized on a surface are not best studied using point-like excitation/detection geometries. Slowly diffusing molecules can of course be detected with fixed point-like excitation geometry. In fact, because they stay longer in the excitation volume than fast diffusing species, the total collected signal will in general be larger than for fast diffusing molecules and thus easier to detect. Additionally, the longer transit duration through the excitation volume can reveal intramolecular fluctuations taking place over that timescale [79]. However, it also takes longer for a new molecule to reach the excitation volume; therefore, accumulating a statistically significant number of individual molecule measurements takes more time, because the concentration still needs to remain very low in order to avoid having several molecules within the excitation volume. In addition, the probability that molecules photobleach during their transit through the excitation volume is increased, potentially reducing the measurement yield. Another possibility is to track individual mobile molecules using some active feedback mechanism, updating either the position of the excitation volume or that of the sample within the excitation volume [80–82].

Immobile molecules or molecules trapped in microscopic domains need to be first imaged (using sample or beam raster-scanning), located, and finally positioned one at a time in the excitation volume (or the excitation volume moved to each different location) sequentially in order for a single-molecule time trace to be collected [11,12]. This is clearly a very inefficient approach, with the additional disadvantages that imaging prior to time trace acquisition may result in premature photobleaching of molecules, and sample or set-up drift may prevent from reliably analysing more than a few molecules, before a new image needs to be acquired to update the location of the remaining molecules.

In summary, in both fixed or mobile molecule situations, an imaging approach using a wide field detector is in general more effective [83]. In particular, because the observation time of each molecule can last as long as the molecule is not photobleached, there is no need for high temporal resolution to simply detect the molecule, and the optimal frame rate is determined by other considerations. Single-molecule experiments then consist simply of recording movies of the sample, from which intensity time traces corresponding to individual molecules can be extracted.

Like for point geometry, signal can be obtained by one or more excitation wavelengths and detected in several spectral bands [84,85]. Similar to point geometry, single-molecule detection using wide field microscopy requires that the two criteria introduced in §1 are fulfilled. First, good SNR and SBR are needed, which in general requires similar optimizations as described for point geometry, but can also influence the minimum recommendable frame duration. Second, molecules need to be sufficiently dilute to be optically distinguishable. We will first discuss how this can be achieved experimentally, before reviewing some typical data acquisition and analysis techniques, in order to better understand detector requirements.

(a) Wide field imaging approaches

The simplest situation occurs when single molecules are confined in a single plane or move vertically over short distances only (equivalent to approx. 1 μm for a typical microscope magnification M = 60) and there is no major background source. In these situations, a simple epifluorescence imaging geometry can be used and excellent SNR and SBR obtained using standard CCD cameras. This is for instance the case when imaging bright single-molecule probes in in vitro assays on bound or diffusing molecules [83,86]. It also applies to fluorophore-labelled molecules imaging in live cells, in the presence of negligible amount of free fluorophores and if buffer/cell autofluorescence spectrum does not overlap the fluorophore-emission spectral range. This is in practice difficult to achieve but with very bright (but large) probes such as quantum dots (QDs) [87–89] or when very low excitation intensity is used as in super-resolution imaging approaches based on single-molecule localization (PALM [90], F-PALM [91], STORM [92], etc.). In these latter cases, however, the only information recorded from each molecule is its instantaneous localization, with some limited extension to tracking [93].

A common, although slightly more constraining, approach consists of using total internal reflection (TIR) excitation, in order to excite fluorescence only in a thin layer (approx. 100–200 nm) adjacent to the coverslip interface [83,94]. This limits its use to surface-bound molecules, or membrane-bound molecules, when the membrane (either artificial or that of an adherent cell) is close to or in contact with the coverslip. Naturally, a combination of both types of illumination scheme can be used [88]. For common fluorophores, TIR microscopy is often the method of choice for long-term observation of single molecules, by virtue of its complete elimination of out-of-focus background.

In order to study specimens that do not reside close to the coverslip, imaging in three-dimensions is needed. In microscopes, the objective lens work in combination with the tube lens so that images of objects located in the objective focal plane are in focus in the tube lens’ image plane [13,95]. In other words, for a given vertical position of the objective lens with respect to the sample, only molecules located close to the objective focal plane are (approximately) in focus in the tube lens’ image plane. The typical vertical distance from the focal plane over which a molecule can be considered approximately in focus (the depth of field) is given by [95,96]:

\[ \Delta z = \frac{n_0 \lambda}{NA^2} \sim \lambda, \quad (3.1) \]

where \( n_0 \) is the index of refraction of the medium between the objective lens and the specimen. Three-dimensional imaging is thus achieved by sampling successive planes of the specimen by incremental displacements approximately \( \Delta z \) of the objective lens. The non-zero depth of field allows imaging of single molecules moving slightly in and out of the focal plane without any need for refocusing, but it also results in background detection from molecules further out-of-focus, which might become a significant problem if the concentration of single molecules is large or if the background signal is high. Vertical optical sectioning (or out-of-focus light rejection) is therefore needed in many cases of wide field three-dimensional imaging.

By analogy to single-spot scanning confocal microscopy, strategies based on multispot [97] or line/slit [98] scanning have been devised to go beyond the sequential acquisition of a confocal image pixel by pixel. However, unless two-photon excitation is used, delicate alignment of a set of
conjunct pinholes/slit is required. This alignment is built-in in spinning disk (Nipkow disk) confocal microscopes, which are capable of forming up to several thousand images per second, and typically use standard high-sensitivity cameras (e.g. EMCCD), readily providing single-molecule sensitivity [99]. In practice, the frame rate is limited by the detector.

More affordable and promising approaches to optical sectioning dispensing with the complex opto-mechanical Nipkow head exist, but are not yet commercially available. For instance, Jovin and co-workers have developed a programmable micromirror array-based microscope (PAM) [100, 101] allowing flexible definition of the amount of light reaching any part of the specimen. Another promising approach is based on light-sheet illumination (also known as selective plane illumination microscopy) [102], whereby a laser beam is focused only in one direction and injected perpendicularly to the imaging optical axis into the sample. Although the limited NA of the illumination lens results in a larger depth of field than in the traditional fluorescence microscopy, the obtained SNR is sufficient for single-molecule imaging in live cells [103].

There are many other approaches to achieving three-dimensional sectioning, such as structured illumination [104–106] or temporal focusing (based on two-photon excitation using ultrafast infrared, pulsed excitation) [107, 108], and undoubtedly many more will be developed. Most, however, will probably have similar detector requirements, discussed in §3f.

(b) Single-molecule detection and analysis in wide field geometries

A wide field geometry has the advantage of permitting the observation of several molecules simultaneously, but also tracking them if they are mobile, allowing extraction of motion parameters such as the diffusion coefficient [109]. Being able to simultaneously observe immobilized molecules for a long time simplifies the study of irreversible dynamics, because the number of independent measurements is equal to the number of molecules in the field of view. This geometry is also efficient at detecting rare binding events, when one of the components can be surface immobilized or even within a population of mobile molecules [110].

Similar to point-like geometries but because of a different reason, there is a maximum concentration above which single-molecule analysis becomes impossible, especially for mobile molecules. This regime is attained when PSFs of nearby molecules overlap. For typical visible wavelengths and numerical aperture, the corresponding density is of the order of one molecule per μm² (or μm³ for three-dimensional imaging). The total number of observable molecules is then set by the detector area, number of pixels and optics magnification. Simple geometrical considerations show that at most N/4 regularly patterned molecules can be unambiguously observed with a detector having N pixels². This number is an upper limit, and requires precise micrometre-scale patterning of the molecules, which limits it to in vitro assays on surface-immobilized molecules or trapped molecules [111]. In practice, random localization is more common, which limits the effective density and the total number of simultaneously fluorescing molecules per field of view to a few hundred. This does not necessarily mean that the total of molecules present in the imaging plane is limited to such a low number, as illustrated by super-resolution imaging approaches using single-molecule localization [90, 91, 112]. In these cases, less than a hundred molecules are turned on at any time, but they are replaced during later acquisition stages by another random set of molecules. The image is then reconstructed using all single-molecule coordinates [113–115].

SNR and SBR considerations in wide field microscopy are slightly different from those discussed in the point-like geometry situation. In particular, the optimal size of the PSF image compared with the detector pixel size will depend on the intended application. To simplify, if the intent is to accurately localize the single molecule, spatial oversampling of the PSF is necessary, whereas the contrary is preferable only if the information of interest is the single-molecule fluorescence intensity. A detailed discussion can be found in the electronic supplementary material (text and figure).

The practical localization precision is bounded by the Cramér-Rao lower bound, which depends in a nonlinear manner on both SNR and SBR [9, 116–119]. A good approximation, provided the PSF image is sufficiently well sampled, is given by the product of the PSF size (σ) and the inverse of the SNR. For most detectors, this translates into nanometre-scale resolutions with as few as 100 detected photons. As shown in the electronic supplementary material, a large SBR is also required, while a photon-counting detector (F = 1) may be preferable to a better QE detector affected by a large ENF.

Like for experiments on single-molecule diffusing in solution, the number and variety of applications have grown tremendously since the original pioneering works, including combining single-molecule imaging and manipulation [120, 121]. We refer the interested reader to recent reviews on the topics mentioned earlier, our purpose being limited to describing the influence of detector choice in these experiments.

(c) Image correlation spectroscopy

As seen previously, the single-molecule regime for diffusing molecules in solution is defined by an effective number of molecules per excitation/detection volume much less than 1, beyond which methods designed to analyse signal from isolated molecules break down and need to be replaced by fluctuation analysis techniques such as FCS. Similarly, single-molecule localization and time trace studies are superseded by image correlation analysis techniques beyond a certain concentration preventing identification and/or tracking.

One major difference with single-point geometry is that image correlation techniques can be of two different types depending on the modality of image acquisition: (i) based on raster-scanned images obtained by confocal laser scanning microscopy [122] or (ii) based on image acquisition using wide field illumination and camera [123]. A recent review nicely summarizes the (expanding) variety of mathematical treatments to which these two different types of datasets can be submitted [124], including alternatives to the super-resolution imaging techniques by single-molecule localization mentioned previously [125, 126]. As with FCS, image correlation spectroscopy techniques can also be used on single-molecule image series in order to get access to fast timescale information [127]. And as in FCS, noise and sampling
(now not just temporal but also spatial) issues are of paramount importance to efficiently use these methods [128].

**d) Detectors used for wide field single-molecule imaging**

Standard wide field experiments, for the most part, use cameras acquiring successive frames exposed for a finite period of time. Different sensor technologies and readout designs exist, most of which are compatible with single-molecule detection, provided they meet a few performance requirements [9].

Back-thinned CCD cameras have the best sensitivity (figure 1b) but relatively low readout rate and suffer from increasing readout noise at higher frame rate. Intensified CCD cameras use an image intensifier converting each photon incident on the front-most photocathode into a shower of thousands of photons detected by a standard CCD [129]. Signal amplification reduces the contribution of readout noise to the overall SNR. There are a few problems with ICCDs but also some advantages. The first problem is the use of high voltage for the intensifier and the resulting risk of damage by excessively bright illumination. The second drawback is that the sensitivity of standard intensifier photocathodes (GaAsP) is at most half that of a good CCD camera (figure 1b), therefore the gain in SNR appears only at very low light levels. Finally, the gain mechanism introduces additional variance that can affect signal quantification. On the other hand, an advantage of intensifiers is that they can be gated very rapidly (in less than 1 ns) or modulated at high frequency, allowing time-resolved studies.

The concept of intensification of the incoming photon signal before readout has been implemented with many variations. An image intensifier can, for instance, be coupled with a complementary metal oxide semiconductor (CMOS) camera instead of a CCD camera, in order to take advantage of its faster frame rate [130]. Electron-bombarded (EB) cameras use a different amplification mechanism than intensified camera by accelerating each photoelectron generated by the front photocathode in order to generate thousands of secondary electrons in the impact material (CCD or CMOS pixel) [131], with the advantage of a better spatial resolution and a smaller ENF than ICCD [132].

The latest amplification concept to be implemented and commercialized is that of electron multiplication (EM) before analogue-to-digital conversion used in the EMCCD camera [133]. Here, the sensor is a CCD with its ultimate sensitivity (at least for back-thinned CCD cameras; figure 1b), but the readout process is modified in order to amplify the photoelectrons stored in each pixel several hundred times prior to digitization, helping overcoming readout noise. As in all integrating amplifying detectors, there is an ENF cost due to this process depending on the applied gain.

The first requirement for a good single-molecule imaging detector is identical to that encountered in point detection: only single molecules can be detected, provided SNR and SBR are sufficient. However, although for single-photon-counting point-detectors this criterion was equivalent to maximizing the QE, it is not sufficient in the case of some cameras or wide field detectors, some of which may have small fill-factors (ratio of the pixel sensitive area to the square of the pixel pitch), which can result in a PDE smaller than the QE by one or more orders of magnitude. Large readout noise or added noise owing to gain mechanisms will also reduce the SBR and SNR, respectively, compared with a detector with identical QE but no readout noise or gain. Low fill factors can, in principle, be compensated by microlens arrays, whereas readout noise can be compensated by amplification (gain) before readout. The latter process comes at the expense of additional signal variance, and this can reduce SNR by up to 40 per cent or more compared with the ideal case of a photon-counting detector. The result is that ENF can cancel out any advantage provided by a larger QE. Luckily, CCDs, scientific CMOS cameras, but also intensified cameras (e.g. EMCCDs) are generally adequate for this task [134].

The second requirement of single-molecule separability has been briefly discussed previously: owing to the diffraction limit of optical microscopy, the images of single-molecules need to be distinguished either spatially (or temporally). Typically, proper localization of a single molecule requires slight over-sampling of the PSF image [117], and non-overlap between nearby molecules requires them to be separated by at least a few PSF diameters. This, however, is easily achieved by a proper choice of imaging magnification.

Finally, because most wide field detectors are integrating detectors, their temporal resolution (frame rate) needs to be sufficient to resolve the phenomenon to be studied, while preserving SNR and SBR. The optimal resolution is application-dependent: single-molecule localization may not need as good a temporal resolution as for studying single-molecule conformational dynamic using FRET, for instance. In fact, the optimal choice of temporal resolution may sometimes be counterintuitive, as in the case of diffusion coefficient measurements, where longer integration times are in general always preferable [135].

With cameras, the minimum temporal resolution increases with the total number of pixels owing to the way pixel values are digitized and/or transferred to memory. The exact value depends on the technology (full frame, interline or frame transfer readout for CCD technology, or serial digitization for CCD versus parallel digitization for CMOS cameras). For most devices, the values are in the millisecond per frame range or higher. As a reminder, because of readout noise, very high theoretical frame rates are in practice useless for single-molecule detection owing to their low emitted signal. In other words, as soon as the readout rate reaches a value for which each pixel collects at most a few photons per frame, readout noise (or the effective noise factor associated with gain) will result in SNR too low for practical use, unless strategies equivalent to using these detectors as photon-counting detectors are used [136].

Needless to say, time-correlated measurements on the nanosecond timescale discussed in §2 are impossible with such devices lacking subnanosecond timing capabilities. Moreover, the use of time-gating [137] or frequency-modulation [138] with intensifier-equipped cameras used in fluorescence lifetime imaging microscopy (FLIM) significantly reduces the effective sensitivity of these detectors, making those approaches inappropriate for single-molecule detection [139].

Finally, it is worth considering a serious drawback of standard wide field detectors at high frame rates: because single-molecule imaging is sparse imaging (owing to the requirement to be able to distinguish each individual PSF), most of the pixels contain information unrelated to any single molecule (and for the most part just readout noise) and are thus wasted bandwidth and disk space. We will
4. New strategies for high-throughput single-molecule spectroscopy

Multispot parallelization is in theory the simplest way to address the throughput limitations of point-like detection. The same need to speed up acquisition was in fact initially encountered in confocal imaging and has received a number of technical solutions in the past. However, single-molecule detection introduces particular constraints that render some of these solutions inadequate. Parallelization is challenging because it needs to address four separate but inter-related issues:

- Sample excitation parallelization.
- Signal detection parallelization.
- Excitation and detection alignment.
- High-throughput data processing.

We will briefly discuss these points, before presenting some experimental solutions.

(a) Sample excitation parallelization

Multiple diffraction-limited spot excitation can be obtained using different approaches among which cascaded beamsplitters [140], microlens arrays [97], digital micromirror devices (DMDs) [100], diffractive optics element [141], spatial light modulators (SLMs) [142] or ZMWGs [16] are a few proved solutions. A first practical requirement is that each spot is sufficiently far away from its neighbours to avoid cross-excitation of the same molecule by adjacent spots. A simple rule of thumb to avoid this problem is to ensure that the inter-spot distance, \( l \), is at least a few times its diameter, \( d \) (see the electronic supplementary material, figure S3) [57]. The spot size itself is defined based on the excitation PSF, which to first order approximation can be modelled by a slightly elongated three-dimensional Gaussian with transverse standard deviation \( \sigma \) (see above). The separation criterion can thus be expressed as \( l > k \cdot d \sim 6k \cdot \sigma \), where \( k \) is a constant much larger than 1, and the diameter of the excitation spot \( d \) is that of the Airy spot. In practice, larger separation may be needed to account for imperfect (non-diffraction-limited) PSF-generating excitation away from the PSF centre, an effect that will be compounded by the presence of many neighbouring spots [143]. Note that this does not eliminate long-term correlation between signals from neighbouring spots, because a single molecule can diffuse from one spot to the next-nearest neighbour in a typical time \( \tau_{\text{diff}} \sim \lambda^2 / 4D \), where \( D \) is the diffusion coefficient of the molecule. For \( k = 10 \), \( D = 414 \, \mu m^2 \, s^{-1} \) (R6G dye) [144], \( \lambda = 532 \, nm \) and NA = 1.2, we find \( \tau_{\text{diff}} > 18.8 \, ms \). This timescale is sufficiently well separated from the diffusion time across an individual spot (\( \tau_{\text{diff}} \sim \sigma^2 / D \sim \tau_{\text{diff}} / 9k^2 \sim 21 \, \mu s \)) to not affect individual ACF curves or perturb single-molecule burst measurements [45].

(b) Signal detection parallelization

Detectors used in multispot experiments need to be able to collect light from each single spot, with minimum contamination from other spot emissions. Although one could imagine multiplexing schemes using a single detector to collect and disentangle signals originating from different locations [145], it is simpler to use detectors with distinct sensitive element(s) (or pixel(s)) associated with each individual spot. In this case, the detector’s geometry should reflect that of the excitation spots, scaled up by the optical magnification, \( M \). In particular, the detector pitch (distance between sensitive areas collecting signal from nearby spots), \( l \), needs to match the spot separation, \( l \), times the magnification: \( L = M 	imes l \). Except in detectors with 100 per cent fill factor such as cameras (where an arbitrary pixel or a group of pixels could be dedicated to the collection of signal from a specific spot) [146], detector pitch (or pixel separation) is in general not adjustable and becomes a constraint in the experiment.

On the other hand, because the (minimal) excitation spot size \( d \) is determined by the focusing optics (see the electronic supplementary material, equation S5), the ratio \( \delta = S / Md \) of the detector dimension \( S \) and spot image dimension \( M \times d \) depends mostly on the magnification, \( M \). As discussed in §2, this is an important parameter in both single-molecule burst detection and FCS applications, for which the optimal values of \( \delta \) are different (but of the order of 1). Choosing the optimal \( \delta \) for an application therefore fixes \( M \), which then determines the distance \( l \) between excitation spots in the sample:

\[
L = \alpha \delta \lambda \, (4.1)
\]

where we have introduced the detector aspect ratio \( \alpha = L / S \). For objective lenses with large NA, this condition is equivalent to \( l > \delta \lambda \). It is worth noting that the condition \( l > d \) (i.e. \( L > S \)) required to avoid optical cross-talk between neighbouring spots means that the fill factor of an ideal detector for single-molecule spectroscopy, \( \pi \delta^2 / 4L^2 \sim 1 / \alpha^2 \ll 1 \), a requirement that distinguishes these applications from traditional imaging applications, where a fill factor as close to 1 as possible is generally sought (whether this achieved with or without abutted microlens arrays) [147,148].

It is clear from the previous discussion that the larger the separation between spots (and hence between detector pixels), the lower the cross-talk and/or correlations between neighbouring spots/pixels. However, there is a limit to the extension of both the pattern of spots in the sample and the detector in the image plane. The relevant parameters are discussed in detail in the electronic supplementary material.

(c) Excitation and detection alignment

Once the optimal parameters needed to match excitation spot size/pitch and detector geometry have been selected, a few more challenging steps are needed to ensure that single-molecule signals will be efficiently collected. In particular, aligning all excitation volumes with their respective detector pixels happens to be a non-trivial task, as anyone having aligned a single-spot single-molecule set-up will easily understand. With detector pixel sizes in the range of 10–100 \( \mu m \), a task that is relatively easy to accomplish at the single-pixel level with standard micrometre-resolution translation stages is rendered more challenging by the introduction of one or more additional degrees of freedom for multipixel detectors, corresponding to the orientations of the detector and its pitch. This additional alignment
complexity can be easily solved using a programmable pattern generator allowing complete control on the position, scale and orientation of the excitation pattern. As described below, a liquid crystal on silicon-spatial light modulator (LCOS-SLM) used in a direct space pattern generation mode (rather than a Fourier space or holographic approach as commonly used) makes it straightforward to either interactively or automatically orient and shift, as well as adjust the pitch of simple patterns, as we and others have recently demonstrated with one- and two-dimensional patterns [142,149,150]. However, the alignment of more than one multipixel detector to a common excitation pattern, as needed for multicolour detection used in smFRET experiments, remains challenging.

(d) Parallel data processing

Data from single-photon detectors consist of digital pulses, which can be time-tagged, binned or counted, processed to detect bursts and extract different related quantities (e.g. FRET efficiency). In FCS, the intensity trace or the raw stream of photon time-stamps is auto-correlated (or cross-correlated with other signals), while in time-correlated applications, precise timing information needs first to be extracted using TDCs and then histogrammed and fitted to decay models [21]. The computational cost of these tasks grows linearly with the number of pixels and can become rapidly taxing for a personal computer in terms of central processing unit (CPU) load, memory utilization and disk space. Efforts to offload some or all of these tasks to digital signal processors, field programmable gate arrays (FPGAs) [149,151] or graphics processor units (GPUs) will be needed to allow real-time data analysis and representation for more than a few channels.

(e) Multispot excitation

We have tested two different approaches in order to generate square arrays of diffraction-limited excitation spots:

- Microlens array.
- LCOS-SLM.

(ii) Liquid crystal on silicon-spatial light modulator

Using a LCOS-SLM to perform the same task is a more expensive solution but definitely easier for fine alignment of the set-up. Figure S4B in the electronic supplementary material shows a schematic of a typical optical arrangement used for this purpose. The pattern is generated by directing a collimated and expanded beam onto a LCOS at a small angle from its normal axis, in order to form a pattern of spots in close proximity to the LCOS. This pattern is relayed into the sample with a simple recollimating lens placed before the objective lens. The pattern pitch can be controlled programmatically and adjusted using the LCOS. Finally, the detection path is designed to have the desired magnification to obtain the correct detector to spot image ratio. Alignment of the spots and the detector can be performed coarsely by moving the detector and refined (including rotation) by adjusting the excitation pattern by software. The obtained excitation pattern (see the electronic supplementary material, figure S4D) is of similar if not better quality than that obtained with the microlens array and can be improved by further optimization of the programmed LCOS pattern. All experiments reported below were carried out using an 800 × 600 pixel LCOS-SLM (Model X10468-01, Hamamatsu Photonics). There is plenty of real estate on such a device to create multiple patterns for different wavelengths if needed, as shown in the electronic supplementary material, figure S4D, where the pattern used to generate the 8 × 8 spot array shown occupies less than 10 per cent of the total surface area of the SLM. It must be emphasized that laser power required for these experiments is significantly larger than is customary for standard single-molecule experiments (we used a 1 W laser in the experiments described later). This is owing to the fact that the laser needs to be expanded in order to provide a relatively uniform intensity over all pixels used to generate the power, and only part of the energy ends up in the spots focused in the sample.

Other approaches could be used. For instance, diffraction optical elements (DOEs) [141] can be designed to transform a collimated beam in the desired pattern of spots or lines, but they can be costly and cannot be modified, therefore preliminary testing needs to be performed by some other means before the adequate DOE can be designed. Also, their property is wavelength-dependent, which makes multicolour excitation techniques such as ALEX more difficult to implement. Another possible technical solution would be using DMDs, as used for instance in the PAM [100]. Also programmable, they are less flexible than LCOS-SLM, because a spot in the sample is generated by re-imaging a single ‘on’ pixel of the device: pitch and rotation possibilities are therefore quantized in single spot size units. Finally, for sparse patterns as needed for multispot excitation, most of the laser energy is lost in the ‘off’ pixels (i.e. deflecting light off the excitation path).

(f) Multipixel detectors

We have used three different detector technologies for multispot single-molecule detection:
— multipixel HPD array (Hamamatsu),
— multipixel epitaxial SPAD array (Politecnico di Milano) and
— multipixel CMOS SPAD camera (Politecnico di Milano),
which we will describe in the order they were tested in our laboratory.

(i) Multipixel hybrid photodetectors array
We first tested a multipixel (8 × 8) version of the GaAsP HPD developed by Hamamatsu [153]. The single-pixel version of this technology had proved worthy of consideration for single-molecule spectroscopy in our early experiments [67], despite the potential fragility (and danger for the user) of a high-voltage device. The multipixel version of the device is composed of 64 adjacent ‘pixels’ 2 × 2 mm² each, arranged in a square pattern. In this respect, it has a fill-factor of close to 100 per cent and can be viewed as a true ‘imager’ albeit one with very few pixels. Unfortunately, this is a very complex device to use, as each individual APD functioning in linear mode outputs a very small and brief signal upon conversion of each detected photon: the typical gain of this device is the product of the gain of the first electron-bombardment stage (approx. 1000) and that of the APD avalanche (approx. 500), i.e. about one order of magnitude smaller than that of a PMT [76]. For this reason, the signal of each individual APD needs to be preamplified before it can be detected by a discriminator circuit. Because of the very short duration of the APD output signal (approx. 1–2 ns), the preamplifier circuit needs to be extremely fast, which makes it prone to picking up small noise fluctuations. In other words, the design of fast multichannel analogue electronics for this kind of device presents significant challenges. Next, a second electronic stage (discriminator) needs to pick up the amplified pulses and convert them into digital signals acceptable by standard readout electronics. We used a 16 channel ASIC board (Hamamatsu) designed for multianode PMTs, providing additional preamplification, adjustable threshold discrimination and pulse shaping. Four such boards were used and outputted 4 × 16 low voltage differential signal pulse trains that were converted into 50 ns long transistor–transistor logic (TTL) pulses by a custom-designed board. The TTL pulses were fed to a reconfigurable multichannel counting board (PXI-7813R, National Instruments) programmed in LabVIEW. As the detector output was hampered by random oscillations and frequent ‘spikes’, we eventually stopped our efforts and switched to SPAD arrays as discussed next. Note that even in the case of successful development of a reliable electronics, this detector is by design far from ideal for multispot detection owing to its large fill factor, which would require the equivalent of a pinhole array to be used in order to ensure out-of-focus signal rejection and prevent optical cross-talk between adjacent pixels.

(ii) Eight-pixel epitaxial SPAD array
A prototype of a linear array of thin junction SPAD designed by the Politecnico di Milano group was tested for FCS and smFRET applications. The detector consists of eight individual SPADs (diameter d = 50 μm, pitch l = 250 μm) [154]. Their QE is identical to that of single-SPAD detectors manufactured by Micro Photon Devices (figure 1a) and peaks at 580 nm (QE ~ 50%), reaching approximately 45 per cent at 580 nm, the emission peak of Rhodamine 6G (R6G) or Cy3B, and approximately 30 per cent only at 700 nm, the emission centre of Alexa 647, dyes used in the experiments described below. A linear four- or eight-spot excitation pattern was created using a high power 323 nm pulsed laser (picoTrain, High Q Lasers) and a LCoS-SLM as described previously.

TTL signals generated by each SPAD upon photon detection were fed to individual input channels of the reconfigurable counting board mentioned earlier and time-tagged with 12.5 ns resolution. Data were transferred asynchronously to a computer, saved to disk and processed in real-time (time trace binning) or offline (ACF calculation, FRET histogram).

We first successfully demonstrated parallel FCS measurements from eight quasi-diffraction-limited spots separated by approximately 5 μm using a sample of R6G molecules diluted in various concentrations of sucrose in aqueous buffer [142,152]. Increasing sucrose concentration results in higher solution viscosity, hence in larger diffusion times. Owing to the difference in PSF and alignment of each pair of spot/SPAD, calibration of individual FCS curves was necessary. This was done using the 0 per cent sucrose R6G sample with known diffusion coefficient and concentration as a reference. After renormalization of the ACF curves obtained in different sucrose concentrations, curves from all eight sources overlapped satisfactorily (figure 3a), as indicated quantitatively by the narrow dispersion of fitted parameter obtained from individual curves [142]. Note that the SPAD afterpulsing contribution is significant at timescales below 10 μs, and needs to be included in the fits in order to extract reliable ACF amplitudes.

These experiments illustrate several potential applications of these (and future larger) arrays:

— Higher throughput single sample data acquisition: by averaging the fitted parameters of n measurements obtained in parallel, the standard deviations of the fitted parameters are reduced by a factor 1/√n. Note that this is equivalent to increasing the total measurement duration by a factor n, which, as mentioned in §2c, increases the SNR of FCS measurements by a factor √n.
— Alternatively, a single measurement could be performed in n samples in parallel. Although this is an exciting prospect for high-throughput screening applications, one needs to keep in mind that it requires the samples to be located at a distance l ~ ad from one another in the object plane. For d ~ 0.5 μm and a typical detector aspect ratio α ~ 10, this represents a 5 μm distance, within reach of current microfluidic technology [155].
— The different afterpulsing characteristics of each SPAD in the array require a careful analysis for reliable parameter extraction, but prevent any phenomena with timescale much shorter than 100 μs to be studied. However, as for single-confocal spot FCS analysis [45], this problem can be solved by cross-correlation function analysis of two independent SPAD detectors within a single array (using a dual-view geometry, as described below) or two separate SPAD arrays.

During the course of these experiments, study of a much less concentrated sample of R6G (100 pm) yielded intensity time traces, which exhibited clear single-molecule FRET
experiments. These experiments are more challenging for two different reasons:

— Because our 532 nm laser needs to be used for donor excitation, a red-shifted acceptor had to be chosen. Unfortunately, the QE of the epitaxial thin SPAD is significantly lower than thick SPADs in the far red region of the spectrum (figure 1a). This means that low FRET efficiency molecules, emitting less red (acceptor) photons than donor photons, would be difficult to distinguish from donor-only molecules. For this reason, we focused our preliminary experiments on high FRET samples to maximize our chances to detect doubly labelled molecules.

— Two different images of the excitation spots need to be aligned onto two different sets of SPADs: the donor emission pattern and the acceptor emission pattern. Moving the excitation spot pattern programmatically (using the LCOS software) can help only with alignment on one set of SPADs; therefore, a more tedious alignment is expected involving some mechanical adjustment on the detector side.

To simplify our task, we used a ‘dual-view’ emission path approach [156] in which the donor and acceptor signals are split using a dichroic mirror and bandpass filters but are imaged onto the same detector. Because one-half of the detector is used for each colour, only four excitation spots can be used. Using a commercial unit (OptoSplit II, Cairn Research) for easier alignment, we developed an iterative and semi-automated strategy for precise alignment of the linear array of four excitation spots onto the donor and acceptor region of the sensor using a solution of dye (Alexa 546) with significant signal in both channels [149]. As illustrated in figure 4, clear uncorrelated bursts can be detected in all four pairs of channels when a high FRET sample is observed. More importantly, the uncorrected FRET efficiencies (proximity ratios) calculated for each channel are in agreement with each other (but smaller than the correct value due to the lack of proper corrections), confirming the excellent alignment of all SPAD pairs. Because this experiment was performed using a single laser, the histograms also include donor-only labelled molecules (the ‘zero’ peaks to the left).

Future work will use an ALEX scheme in order to be able to perform measurements on lower FRET samples as well as distinguish between singly and doubly labelled species, and use one SPAD array per channel in order to be able to record data from more spots in parallel.

(iii) 32 × 32 pixel CMOS SPAD camera

We tested a CMOS SPAD array developed by Politecnico di Milano, comprising 32 × 32 SPADs fabricated in standard 0.35 μm high-voltage CMOS technology (diameter $d = 20 \mu m$, pitch $l = 100 \mu m$) [157,158]. This kind of detector, or even more advanced ones, has been presented in the recent past but had never been tested for single-molecule or FCS experiments [159–162]. Because the fabrication process is not optimized for photon detection (but instead for integrated circuit performance), CMOS SPAD QE peaks at 460 nm (40%) and reaches approximately 20 per cent at 580 nm, the emission peak of R6G, the dye used in our experiments (figure 1a). This lesser performance is compensated by the very large number of SPADs contained in a single detector chip.

Each pixel is equipped with its own quenching electronics, eight-bit counter and latch memory. The pixel memory allows counting to proceed continuously while the array is read out, up to every 10 μs. This is one of the major differences of this kind of device from the photon-counting devices discussed so far, in which each photon-detection event was transmitted as an electrical pulse to an external acquisition electronics, which could precisely time-stamp each photon (counting being performed in software). CMOS arrays lose the absolute time information of individual photons, and replace it with the coarser ‘frame’ time information (up to 10 μs resolution in this case) plus the number of successive photons detected during this short period of time. This temporal resolution is sufficient for single-molecule burst detection (each burst being at least approx. 100 μs long) and for single-molecule diffusion studies by FCS, making these devices potentially interesting.

The array function is controlled by an FPGA board containing 32 megabytes (MB) of memory and a universal serial bus (USB) communication module for data transfer to...
the host computer. The 1 kilobyte content of the full array can be read in 10 ms into an onboard memory, but reducing the bit depth of each counter or the size of the region of interest allows faster readout rates. An additional bandwidth limitation exists owing to USB transfer, effectively limiting continuous acquisition of data with lossless transfer at approximately 20 MB s$^{-1}$ to the computer.

Owing to the lower sensitivity and larger dark count rate of this detector (greater than 1 kHz per SPAD, with some SPADs reaching greater than 100 kHz) [158], detection of single-molecule bursts turned out to be impossible [143]. Moreover, although we could generate up to $32 \times 32$ excitation spots in the sample using a 40× oil immersion objective lens (Plan-Neofluar 40×, Zeiss, NA = 1.3), the out-of-focus light generated by this large number of spots resulted in significant background signal in FCS measurement, leading to very low amplitude ACF curves. For this reason, we limited our later experiments to $8 \times 8$ spots FCS analysis of bright 100 nm diameter fluorescent beads. After correction for channel differences (as explained earlier), the 64 calibrated ACF curves collapsed to a single curve, demonstrating the excellent high-throughput capabilities of such a device for bright enough samples (figure 3b). Note that for this particular array the influence of any afterpulsing taking place within the 10 μs resolution is summed up in the first ACF bin, which is therefore showing significant deviation from the asymptotic ACF amplitude.

This series of experiments is instructive for both its positive and negative results. In particular, using a large number of spots seems to necessitate large spot separation to limit background signal due to out-of-focus excitation created by the spot pattern. Wide field three-dimensional sectioning illumination approaches mentioned in §3 may for this reason be preferable for some applications (see Buchholz et al. [151] for a recent example). Large spot separation requires correspondingly large detector pitch to diameter aspect ratio (i.e. smaller fill factor), which fortunately is the norm in SPAD arrays.

Finally, it is worth mentioning again that generating a large number of excitation spots requires significantly higher laser power than is customary in single-spot smFRET or FCS experiments. A 1024 spot excitation pattern will typically require $1024 \times 100 \mu W \sim 100$ mW at the sample. In our LCOS-SLM approach, which requires expanding the laser beam in

---

**Figure 4.** Multispot single-molecule FRET. (a–d) 1 s excerpt from donor (green) and acceptor (red) time traces corresponding to doubly labelled dsDNA molecules diffusing through four confocal spots. A linear eight-pixel SPAD array was used, with one-half of the array collecting donor signals, while the other half collected the acceptor signals. (e–g) FRET histograms corresponding to the time traces shown in the left panels.
order to uniformly cover the LCOS array, this necessitates a significantly larger laser output power.

Although CMOS SPAD arrays do not provide time-tagging information for each detected photon, the amount of transferred information is still significant due to the large number of SPADs per array. In our approach, ACF curves were computed offline using efficient code, but sequentially, resulting in lengthy computation times and requiring storage of large amount of data prior to processing. Parallelization afforded by modern multicore, multithread CPUs or GPUs is an obvious option that could speed up analysis and make it possible to perform in real time. Another very promising approach has been recently demonstrated using FPGA to compute $32 \times 32$ ACFs obtained from a similar CMOS device as the one described here [151]. Similar approaches can obviously be used with single-photon time-tagging approaches.

5. New detectors for single-molecule imaging

Modern cameras are extremely sensitive and some are even capable of frame rates up to several thousand per second [163]. However, high frame rates are useless at very low photon count rates because of the readout noise or ENF of these cameras.

To address this temporal resolution barrier, detectors working in photon-counting mode are most adequate. With some efforts and ingenuity, some cameras can be used in an approximate photon-counting mode [130,136,164] but this is a rather inefficient use of data bandwidth. In order to achieve better photon-counting frame rates, dedicated CMOS sensors have been developed to work with EB tubes [165] or microchannel plate amplification of photocathode-generated electrons [166]. Here as well, the maximum global photon count rate is equal to a few hundred times the frame rate (the number of detected emitters, supposed to emit only a few photons per frame, times the readout frame rate), whereas the raw data bandwidth is equal to the number of pixels times the bit depth of each pixel. Although data reduction (i.e. individual photon localization) can be performed on a frame-by-frame basis prior to transfer, the strategy becomes extremely costly as the number of pixels and/or the frame increase. An event-driven photon-counting detector is a more natural solution. Additionally, these approaches at best provide coarse time-tagging for each photon (with a precision given by the inverse of the frame rate), and are not capable of time-correlated measurement with a pulsed-laser source as used in fluorescence lifetime studies. Using wide field sensors for single-molecule studies is thus paid for by a significant loss of capabilities compared with point-detection geometries discussed earlier.

Faced with this limited set of options, we have looked for technologies capable of bridging the gap between single-photon-counting detectors and cameras. Inspired by the pioneering (and heroic) work of Hübner et al. [167], who used a low sensitivity, position-sensitive wide field photon-counting detector for single-molecule fluorescence lifetime analysis, we have developed a series of improved versions of this type of detector [139,168–173]. It should be pointed out that commercial detectors based on a similar principle have been available for some time but their performance is somewhat limited for single-molecule imaging (see references in Michalet et al. [9]).

The series of prototypes was dubbed H33D (pronounced ‘heed’), for high-throughput, high-spatial, high-temporal resolution, two spatial, one temporal dimensions detector (table 1; electronic supplementary material, figure S5). Their general design is based on a large area (18 or 27 mm diameter) microchannel plate photomultiplier (MCP-PMT) head followed by a position-sensitive anode (PSA) providing approximately $100 \mu m$ or better spatial resolution, measured as the FWHM of the distribution of detected photon localization emitted by a point source (few $\mu m$ wide) placed closed to the MCP input face. Note that the location of each photon (in reality, the location of the centre of each electron cloud generated after photon detection) can be reported with a better resolution by the readout electronics (10–12 bits for the whole sensor diameter). However, the imaging resolution is ultimately limited by the statistical spread of cloud centres locations.

Like for a PMT, the detection efficiency is mainly defined by the photocathode’s QE. Our first generation prototype (H33D Gen 1) used a multialkali (S20) photocathode with a maximum QE of 10 per cent at 400 nm, whereas our more recent SuperGen 2 photocathode device (H33D Gen 2) achieves a QE of approximately 15 per cent in the visible range, and a future version will use a GaAs photocathode with QE approximately 30 per cent (figure 1b).

In addition, like a MCP-PMT, the detector is capable of better than approximately 100 ps temporal resolution (the actual resolution depending on the readout design). Unlike a PMT, however, the anode is designed to locate the secondary electron cloud exiting the back of the MCP stack. Initial position-sensing anodes in previous detectors designed by the Space Sciences Laboratory used a resistive quadrant design [174] with limited readout rate capability. Our first prototype (H33D Gen 1) used a cross-delay line (XDL) anode design [175] allowing up to approximately 1 MHz global counting rates [168,169] and better than $100 \mu m$ FWHM resolution for each photon and approximately 150 ps time-correlated resolution (table 1).

Despite the advantage of photon-counting, there were some limitations to this prototype:

— The dead-time of approximately 500 ns, due to readout electronics (based on a TDC), resulted in limited global count rate $f_{max} \sim 500 \text{ KHz}$.
— The large gain, needed to ensure a large enough charge generation by the MCP to allow accurate localization of each photon, limited the maximum local count rate $f_{max}$ to approximately 10 kHz due to the time constant of the microchannel recharge.

Although we managed to detect single QDs [172] and perform FLIM on both live cells and large fluorescent objects such as beads or QD clusters [139] with this prototype (figure 5), the QE was too low for conventional (organic dye) single-molecule analysis. However, this prototype showcased three advantages of this technique for future single-molecule imaging studies:

— The absence of readout noise makes it possible to study dynamics on timescales limited only by the maximum local count rate $f_{max}$ and shot noise [172].
— Single-molecule fluorescence lifetime studies can be performed at frame rates of the order of $f_{max}/100$ [139].
— The number of single molecules that can be simultaneously studied with optimal temporal resolution is limited only by the ratio $F_{\text{max}}/f_{\text{max}}$.

The absence of readout noise and very low dark count noise of the H33D detector (less than 1 kHz over the whole sensor for the Gen 1 prototype and less than 15 kHz for the Gen 2A prototype), leaves only sample background as a possible source of additional noise on top of fundamental shot noise. For localization, this means for instance that a mere computation of the barycentre of photon locations can provide a localization precision of approximately $s_p/N$, where $N$ is the number of photons collected from a single molecule. Assuming that the single-molecule signal reaches the maximum local count rate $f_{\text{max}}$, this number can be obtained in a time approximately $N f_{\text{max}}$. For $s_p = 100$ nm, $N = 100$, $f_{\text{max}} = 10$ kHz, we obtain a potential precision of 10 nm in 10 ms. Larger maximum local count rates provided by the next generation prototypes (up to 10 times larger) will further increase the ‘frame rate’ at which this level of precision can be achieved. Alternatively, increasing the magnification in order for the PSF to cover a larger area of the detector (so that the maximum count rate $f_{\text{max}}$ is not reached) would further increase this performance.

A similar performance increase can be achieved in fluorescence lifetime measurements, and in a particularly effective way when using the phasor representation of lifetime [139,171]. Indeed, in the same way it is possible to determine the fluorescence lifetime of a single molecule with $N \approx 100$ photons, using a maximum-likelihood estimation approach [18,176]; such a small number of photons is sufficient to obtain a phasor value with reasonably small variance [139,177]. Therefore, not only does it appear feasible to track individual molecules with approximately 10 nm resolution and kHz frame rate, but it is simultaneously possible to measure their lifetime at the same rate, thus monitoring their electronic environment as they explore it (e.g. live cell membrane or cytoplasm).

Finally, the ratio between maximum global count rate and maximum local count rate ($F_{\text{max}}/f_{\text{max}} \approx 50$–200 depending on the H33D prototype) matches the typical number of single molecules that can conveniently and simultaneously be detected and tracked in a microscope field of view.

To improve the local and global count rate limitation of our first prototype, we used a different position-sensing anode based on a cross-strip design [173] and a more efficient SuperGen 2 photocathode (table 1 and figure 1b). The strips collect the local electron cloud charges, which are then digitized and used to compute a weighted average cloud position. This approach provides similar single-photon localization accuracy to the XDL approach, but using one order of magnitude lower MCP gain. This in turns translates into larger achievable local count rates of up to 100 kHz.

In order to achieve higher global count rates, we dispensed with a TDC to time-correlate each photon with the source-pulsed excitation, using instead the electron cloud charge impulse shape collected by the strips to extract the time of arrival of each photon with approximately 1 ns resolution [173]. Although this resolution is lower than what the detector is intrinsically capable of, it is sufficient to extract quantitative information about fluorescence decay using very few photons [177], provided a phasor analysis is used,

---

**Table 1.** The H33D series of wide field photon-counting detectors. Characteristics of the successive H33D prototypes. Two numbers are provided for the temporal resolution. The macrotime $T$ measures the time of arrival of each photon with respect to the beginning of the measurement, while the nanotime $t$ measures the time of arrival of each photon with respect to the next laser pulse. Note that the Gen 2 prototype uses a different approach for lifetime analysis (phasor analysis), which dispenses with the need for a very accurate nanotime. n.d., not yet determined, n.a., not yet available, XDL, cross-delay line; XS, cross-strip.

<table>
<thead>
<tr>
<th>H33D generation</th>
<th>photocathode (QE at 600 nm)</th>
<th>anode type</th>
<th>readout electronics</th>
<th>spatial resolution (μm/pixel)</th>
<th>pixels per axis</th>
<th>temporal resolution $T$, $t$ (ns)</th>
<th>local count rate (kHz)</th>
<th>global count rate (MHz)</th>
<th>completion date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 S20 (4.5%)</td>
<td>XDL</td>
<td>two dual TDC</td>
<td>100/270</td>
<td>20.15</td>
<td>10</td>
<td>1</td>
<td>100/270</td>
<td>20/0.15</td>
<td>2006</td>
</tr>
<tr>
<td>2A SuperGen 2 (14.2%)</td>
<td>XS</td>
<td>60 MHz 502 XIC</td>
<td>n.d.</td>
<td>n.a.</td>
<td>n.d.</td>
<td>n.a.</td>
<td>n.d.</td>
<td>n.d.</td>
<td>2010</td>
</tr>
<tr>
<td>3 GaAsP (approx. 44%)</td>
<td>XS</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.d.</td>
<td>n.d.</td>
<td>n.d.</td>
<td>n.d.</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

---

instead of the classic photon arrival time histogramming and fitting approach used in most TCSPC studies [139].

A similar device using an even more efficient GaAs photocathode will soon be tested and should provide sufficient sensitivity to perform true single-molecule imaging experiments with increased temporal resolution and lifetime measurements capabilities. Although further global count rate improvements can be achieved using modern electronic design and components (and are planned), there are physical limits to what can be achieved, set by (i) the maximum current (number of electrons per unit time) that can be drawn from a MCP device without reducing its performance and lifetime, and (ii) the minimum gain needed to accurately localize and time each electron cloud, which can also not be reduced beyond some lower limit.

6. Conclusion and perspectives

The results of different collaborative developments presented above hopefully show that significant throughput enhancement of single-molecule techniques can be obtained, and more will be achieved in the near future.

Detectors for single-molecule detection need to have the best possible QE (while keeping dark counts to a level compatible with single-molecule detection). Although the physical limit seems to have been already reached at the single-SPAD level in new thick-junction devices, there is room for improvement as far as arrays are concerned, in particular in the red part of the spectrum. There are no reasons to expect that the spectacular enhancement of sensitivity obtained with the thin SPAD technology will not be extendable to array configurations. The question remains to see how many SPADs such custom-technology arrays may eventually be able to contain without becoming cost-prohibitive, overly complex and in the end, of little use to the single-molecule biophysicist. In particular, larger number of SPADs will require more efficient ways to acquire, process and store data.

The wide field photon-counting H33D detectors we are developing have some intrinsic limitations (maximum local count rate imposed by the MCP gain, maximum global

![Figure 5. Single quantum dot imaging with the H33D Gen 1 detector. (a) Single QD intensity time trace binned with different temporal resolution exhibiting the characteristic blinking pattern of QD. (b) Fluorescence lifetime histograms along the trajectory of a QD cluster. Each curve corresponds to 1 s worth of data and is separated from the next by 10 s. After Michalet et al. [172], reproduced with permission from Bentham Science.](http://rstb.royalsocietypublishing.org/)

of the class in photon arrival time histogramming and fitting approach used in most TCSPC studies [139].

A similar device using an even more efficient GaAs photocathode will soon be tested and should provide sufficient sensitivity to perform true single-molecule imaging experiments with increased temporal resolution and lifetime measurements capabilities. Although further global count rate improvements can be achieved using modern electronic design and components (and are planned), there are physical limits to what can be achieved, set by (i) the maximum current (number of electrons per unit time) that can be drawn from a MCP device without reducing its performance and lifetime, and (ii) the minimum gain needed to accurately localize and time each electron cloud, which can also not be reduced beyond some lower limit.

6. Conclusion and perspectives

The results of different collaborative developments presented above hopefully show that significant throughput enhancement of single-molecule techniques can be obtained, and more will be achieved in the near future.

Detectors for single-molecule detection need to have the best possible QE (while keeping dark counts to a level compatible with single-molecule detection). Although the physical limit seems to have been already reached at the single-SPAD level in new thick-junction devices, there is room for improvement as far as arrays are concerned, in particular in the red part of the spectrum. There are no reasons to expect that the spectacular enhancement of sensitivity obtained with the thin SPAD technology will not be extendable to array configurations. The question remains to see how many SPADs such custom-technology arrays may eventually be able to contain without becoming cost-prohibitive, overly complex and in the end, of little use to the single-molecule biophysicist. In particular, larger number of SPADs will require more efficient ways to acquire, process and store data.

The wide field photon-counting H33D detectors we are developing have some intrinsic limitations (maximum local count rate imposed by the MCP gain, maximum global...
count rate owing to the requirement that not two photon-detection events overlap in time and space and simultaneous event handling capabilities) owing to the laws of physics and electronics design. Nonetheless, there are some obvious possible improvements to the H33D prototype, including:

- Using a GaAsP photocathode, providing up to approximately 45 per cent QE in the visible range (figure 1b).
- Developing faster and low noise strip readout electronics in order to reduce the dead-time of each single-photon measurement and improve the temporal resolution.
- Developing pixilated PSA or novel concepts allowing multiple, simultaneous hits to be handled and increasing the global counting rate [178].

New design principles will be required once the prototype’s performance hits these limits. In particular, it appears likely that large CMOS SPAD arrays of the type discussed here but with improved sensitivity [179–181], coupled to microlens arrays [147,148] or designed to achieve fill factors close to 100 per cent by alternative architectures, will play an important role in time-resolved photon-counting imaging [162,179,182,183].

Single-molecule imaging and spectroscopy raises a number of interesting challenges from a detector point of view, in both its point-like and wide field versions. We have presented our current efforts to take advantage of existing technologies to increase the throughput and information content achievable in these experiments. Clearly, a lot more needs to be done to reach optimal performance. What we have learned from our efforts is that a dialogue between detector developers and single-molecule experimentalists is essential to optimize detector design and performance, and, inversely, knowing the existing potential of detector technologies suggests new types of experiments or data analysis approaches. We believe that the collaborative approach we have used will avoid the development of ‘monster’ detectors with few applications or useless or detrimental characteristics.

At the same time, it is clear that many of the requirements of single-molecule imaging and spectroscopy overlap with a large number of other low light level applications and therefore ideal detectors for our purposes will certainly be used in many others. The potential benefits of these technologies for biomedical, biopharmaceutical and biodetection applications in general will hopefully motivate more research and development of novel detectors, as well as funding for this type of activities.

This work was supported by NIH grant no. R01-GM084327 (UCLA), NIH grant no. R01-EB006353 and NSF grant no. DBI-0552099 (UCLA and UC Berkeley), EC agreement no. 232359 (PARAFLUO) FP7-SME-2008-1 (Politecnico di Milano) and NIH grant no. R01-GM095904 (UCLA and Politecnico di Milano). We thank Dongisik Kim and Dr Taipyo Kim (Nesher Technologies), as well as current and past members of the Weiss laboratory for their contributions to these projects. The help and collaboration of Dr Suyma and co-workers (Hamamatsu Photonics) during our tests of single and multipixel HPD for single-molecule applications is gratefully acknowledged. The help of Drs Paul Hink and Emile Schyns (Photonis) and skilled work of engineers and technicians at Photonis for manufacturing photodetectors and assembling the Gen 2 H33D detector tubes were critical for the successful outcome of the H33D project.

Endnotes

1Note that the magnification of the lens does not matter for this type of experiments.

2This corresponds to an arrangement where the signal from each single molecule is collected by a single pixel and is separated from the next molecule by a one-pixel gap.

3Note that this problem would be somewhat alleviated in an ALEX scheme, where the acceptor molecule would be fully excited (and detected) by the acceptor excitation laser, eliminating this ambiguity (but not the SNR issue).

4SPAD arrays being read as ‘frames’ at a fixed frequency, the maximal temporal resolution with which each detected photon (or group of photons) can be timed is given by the frame interval. TCSPC-capable SPAD arrays may provide accurate timing (< ns) with respect to a laser pulse, but this only constitutes a local resolution, the resolution with respect to the start of the experiment remaining that of the array readout rate.

References

52. Safarrian S, Elson EL. 2003 Statistical analysis of fluorescence correlation spectroscopy: the standard


