Time-resolved spectroscopy at 19,000 lines per second using a CMOS SPAD line array enables advanced biophotonics applications

Citation for published version:

Digital Object Identifier (DOI):
10.1364/OE.25.011103

Link:
Link to publication record in Heriot-Watt Research Portal

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Optics Express

General rights
Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Time-resolved spectroscopy at 19,000 lines per second using a CMOS SPAD line array enables advanced biophotonics applications

A. KUFCSÁK,1,5 A. ERDOGAN,1 R. WALKER,1 K. EHRLICH,2,3 M. TANNER,2,3 A. MEGIA-FERNANDEZ,3,4 E. SCHOLEFIELD,3 P. EMANUEL,3 K. DHALIWAL,3 M. BRADLEY,3,4 R. K. HENDERSON,1 AND N. KRSTAJIĆ1,3,*

1Institute for Integrated Micro and Nano Systems, School of Engineering, University of Edinburgh, King's Buildings, Alexander Crum Brown Road, Edinburgh EH9 3FF, UK
2Scottish Universities Physics Alliance (SUPA), Institute of Photonics and Quantum Science, Heriot-Watt University, Edinburgh EH14 4AS, UK
3EPSRC IRC Hub in Optical Molecular Sensing & Imaging, Centre for Inflammation Research, Queen's Medical Research Institute, University of Edinburgh, 47 Little France Crescent, Edinburgh EH16 4TJ, UK
4School of Chemistry, EaStChem, University of Edinburgh, Joseph Black Building, West Mains Road, Edinburgh EH9 3FJ, UK
5a.kufcsak@ed.ac.uk
*krstajic@physics.org

Abstract: A SPAD-based line sensor fabricated in 130 nm CMOS technology capable of acquiring time-resolved fluorescence spectra (TRFS) in 8.3 milliseconds is presented. To the best of our knowledge, this is the fastest time correlated single photon counting (TCSPC) TRFS acquisition reported to date. The line sensor is an upgrade to our prior work and incorporates: i) parallelized interface from sensor to surrounding circuitry enabling high line rate to the PC (19,000 lines/s) and ii) novel time-gating architecture where detected photons in the OFF region are rejected digitally after the output stage of the SPAD. The time-gating architecture was chosen to avoid electrical transients on the SPAD high voltage supplies when gating is achieved by excess bias modulation. The time-gate has an adjustable location and time window width allowing the user to focus on time-events of interest. On-chip integrated center-of-mass (CMM) calculations provide efficient acquisition of photon arrivals and direct lifetime estimation of fluorescence decays. Furthermore, any of the SPC, TCSPC and on-chip CMM modes can be used in conjunction with the time-gating. The higher readout rate and versatile architecture greatly empower the user and will allow widespread applications across many techniques and disciplines. Here we focused on 3 examples of TRFS and time-gated Raman spectroscopy: i) kinetics of chlorophyll A fluorescence from an intact leaf; ii) kinetics of a thrombin biosensor FRET probe from quenched to fluorescence states; iii) ex vivo mouse lung tissue autofluorescence TRFS; iv) time-gated Raman spectroscopy of toluene at 3056 cm\(^{-1}\) peak. To the best of our knowledge, we detect spectrally for the first time the fast rise in fluorescence lifetime of chlorophyll A in a measurement over single fluorescent transient.

© 2017 Optical Society of America

OCIS codes: (040.1240) Arrays; (120.6200) Spectrometers and spectroscopic instrumentation; (300.2530) Fluorescence, laser-induced; (300.6450) Spectroscopy, Raman; (300.6500) Spectroscopy, time-resolved.

References and Links


1. Introduction

The unique advantage of single photon detection is that each photon can be treated as an individual, digital event. As the early pioneers of time-resolved detection Bollinger and Thomas [1] realised, single photon detection gives an experimenter immense flexibility. For example, as first shown in [1], it is possible to measure the time between the light pulse interacting with the sample and the detected luminescence (or scattered) photon arrival time. When each photon is time-correlated with the respective triggered laser and sorted into a histogram the technique is known as time-correlated single photon counting (TCSPC, see Table 1 for all abbreviations).

Typically, TCSPC is a slow technique and is limited only to low light scenarios. Imposition of the detector dead-time and photon statistics constraints (pile-up) usually do not affect the inherently photon-starved measurements (see Chapter 1 in [2]), but they do slow down techniques where photon budgets are not so constrained. The slowness of TCSPC is mostly due to the fact that detectors in most TCSPC environments are photo-multiplier tubes (PMTs) which are hard to array, resulting in most setups relying on just one or two detectors.

### Table 1. Abbreviations used in the article

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMM</td>
<td>Center-of-mass method, often called method of moments</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary metal-oxide-semiconductor</td>
</tr>
<tr>
<td>DCR</td>
<td>Dark count rate</td>
</tr>
<tr>
<td>FAM</td>
<td>5-carboxyfluorescein</td>
</tr>
<tr>
<td>FLIM</td>
<td>Fluorescence lifetime imaging</td>
</tr>
<tr>
<td>FPGA</td>
<td>Field programmable gate array</td>
</tr>
<tr>
<td>FRET</td>
<td>Förster resonance energy transfer</td>
</tr>
<tr>
<td>FWHM</td>
<td>Full width half maximum</td>
</tr>
<tr>
<td>IIR</td>
<td>Infinite impulse response (filter)</td>
</tr>
<tr>
<td>IRF</td>
<td>Instrument response function</td>
</tr>
<tr>
<td>LLTF</td>
<td>Laser line tunable filter</td>
</tr>
<tr>
<td>LRRS</td>
<td>Low resolution Raman Spectroscopy</td>
</tr>
<tr>
<td>MMP</td>
<td>Matrix metalloproteinase</td>
</tr>
<tr>
<td>NIM</td>
<td>Nuclear instrumentation module (electrical signaling standard)</td>
</tr>
<tr>
<td>PCB</td>
<td>Printed circuit board</td>
</tr>
<tr>
<td>SPAD</td>
<td>Single photon avalanche diodes</td>
</tr>
<tr>
<td>SPC</td>
<td>Single photon counting</td>
</tr>
<tr>
<td>TCSPC</td>
<td>Time correlated single photon counting</td>
</tr>
<tr>
<td>TDC</td>
<td>Time to digital converter</td>
</tr>
<tr>
<td>TRFS</td>
<td>Time-resolved fluorescence spectroscopy</td>
</tr>
<tr>
<td>TTL</td>
<td>Transistor-transistor logic (electrical signaling standard)</td>
</tr>
</tbody>
</table>

Our aim in this paper is to show that recent advances in complementary metal-oxide-semiconductor (CMOS) technology and single photon avalanche diodes (SPADs) allow massively-parallelized photon sorting [3–5] enabling both low light and light bursts to be efficiently time-stamped spectrally in a single fluorescence transient experiment. We use the term fluorescence transient to mean a change in fluorescence intensity [6,7] unrelated to excited state (fluorescence or phosphorescence) lifetime.

While most recent applications of CMOS SPAD arrays have focused on fast fluorescence lifetime imaging (FLIM) [8] using TCSPC [4,9–11], here we build on our prior work in time-resolved CMOS SPAD line sensors [12,13] to show that fluorescence transients can be measured both in the millisecond and nanoseconds regime (“double kinetics” [14,15]) in a single experimental sequence and in the spectral domain. Previously, to study fast fluorescent transients using TCSPC in the millisecond regime the transients needed to be induced many times in order to acquire time-resolved fluorescence kinetics [14,16]. Direct decay measurements have been achieved by a pulse sampling technique [17] and applied in tissue autofluorescence studies [18,19] and protein folding [14,20,21]. Pulse sampling achieves nanosecond resolution, but requires a high-end oscilloscope (2 GHz or more), while the pulsed lasers deployed tend to have high peak power to enable fast and direct detection of the fluorescence decay. Therefore, one of the significant advantages of the work presented here is that we demonstrate time-resolved spectroscopy using high repetition rate picosecond laser diodes [22] with low average power (<0.5 mW) making it immediately more attractive for clinical use [23] as well as opening up a wider range of other applications due to the lower cost and reduced number of safety issues.

Fig. 1. Fluorescence transients can easily be measured using a photodiode in the microsecond to millisecond regime (top graph (I)). Recent developments in compact spectrometers have advanced this further allowing a spectral view into fluorescence transient evolution (middle graph (II)). In this paper we are able to show spectral “double kinetics” in nanosecond and millisecond regime whereby the transient is induced only once. For each time point ($t_1$, $t_2$, ...) a TRFS data cube is acquired during the single transient.

Figure 1 shows the outline of our proposed luminescence transient analysis. The top graph (I) shows a fluorescence transient (“single kinetics”) from an analogue photo-detector such as a photodiode. This is easily achievable and the next step is either to have a fluorescence decay
for each time point [14–16] or spectrum for each time point as shown in middle graph (II). Our work advances the spectral lifetime acquisition or time-resolved fluorescence spectroscopy (TRFS) and we demonstrate “double kinetics” in the spectral domain potentially exploitable in a large number of phenomena such as chlorophyll A fluorescence (plant fluorescence), fluorescence quenching, photo-bleaching, Förster resonance energy transfer (FRET) and protein folding dynamics to name but a few. We particularly pay attention to plant fluorescence transients as an important example of fluorescence kinetics. As described in numerous introductory reviews [7,24,25], most of the plant fluorescence is attributed to chlorophyll A. Plant fluorescence is also an important indicator of plant stress and photosynthesis efficiency. Under controlled illumination fluorescence increases rapidly (during the first second, often 100 ms) which is followed by a slow fall (several minutes depending on plant species). This transient is usually referred to as the Kautsky effect. While the fluorescence properties of the slow fall have been studied in detail we demonstrate our sensor can be applied to studying the rapid rise in terms of spectral lifetime.

Fig. 2. Demonstration of spectral decays in two simulated scenarios. The first scenario is photon rich: (a) defined here as most pixels having more than 2000 counts in decay. Plot (b) shows the difference between the decay in single pixel and binning 80 pixels which results in a better looking decay. The second scenario is photon starved: (c) where the decay derived from a single pixel in (d) is too noisy. However, the decay from 80 binned pixels is good enough to allow the user to manipulate the data cube in an optimal way given the amount of light detected. Spectral decays were simulated in Matlab 2014a (Mathworks, USA), shot noise limited, visible to near-infrared TRFS data cubes. Spectral decays have 0.4 ns time resolution and 1.6 nm spectral resolution. Representing broadly the properties of the CMOS SPAD line sensor presented here.

Lastly, by binning the TRFS data cube either in the time or wavelength domain it can be adapted for time-resolved imaging. To illustrate this and similar scenarios, Fig. 2 shows two simulated TRFS data cubes in Figs. 2(a) and 2(c) and sample decays in Figs. 2(b) and 2(d) taken from the 650 nm - 750 nm region. Three emission peaks are simulated to have different decays to demonstrate a scenario where three fluorophore emissions are excited at 488 nm, 625 nm and 780 nm and studied simultaneously in a broadband visible to near-infrared.
spectrometer. When TRFS at high spectral resolution is needed then binning can be done along the time axis. Similarly, if the time resolution needs to be kept, more counts can be gathered by binning in the spectral domain. In fact, spectral binning can be applied until one gets a single TCSPC decay curve. The compromise is that binning along time or wavelength axis entails loss of time or spectral resolution respectively, even though advanced strategies have been deployed recently [26,27]. The key aim is always to gather more counts in a decay, which in turn could enable better exponential fitting. We apply this methodology throughout our study of TRFS below.

Solid-state single photon detectors have a long history dating back to the 1960’s [28]. Since the first integrated SPAD array was created using industrial CMOS technology [29], a variety of CMOS fabricated SPAD sensors have been designed. CMOS technology allows enhanced detection with additional logic on the detector device and enables cost-effective mass production of such sensors. Recent systems apply SPAD-based line sensors, where light from a point of the specimen is dispersed over several pixels. Timing information is either collected by time-gated photon counting [11,30,31], or by applying time-to-digital converters (TDCs) for each pixel, realizing a multi-channel TCSPC system [12]. Spectrally resolved lifetime images can be acquired by combining time-resolved line sensors with laser scanning microscopy [11].

SPAD line sensors have been reported targeting other applications, such as near-infrared optical tomography [32], 3D imaging based on time-of-flight measurements [33], laser-induced breakdown spectroscopy [34] and time-resolved Raman spectroscopy [35–39]. Line arrangements of pixels are in general favorable for increasing fill-factor of the sensor, as any additional logic can be placed next to the photon-sensitive area of the pixels. A line sensor implemented with 3D multi-wafer stacking CMOS technology has been demonstrated recently [32]. A novel approach in integrating flexible TDC architectures onto SPADs is to connect field programmable gate array (FPGA) based TDCs to SPADs, instead of integrating more functionality into the CMOS sensor (8.5 giga-events/s achieved) [33]. This approach separates analog SPAD circuits from digital FPGA side and allows users to benefit from recent FPGA advances.

TCSPC is restricted to process only one photon arrival per pre-defined exposure time [2]. The sensor we present enhances this by applying a separate TDC channel for each of its 256 pixels. To further increase the efficiency of time-stamping, time-resolved detection can also be performed in a center-of-mass (CMM) mode. This allows photon efficient processing of photon arrival times, because all processing is done on the sensor itself. This technique provides fast fluorescence decay analysis [40–43], while each operating mode of the sensor can be amended with an adjustable ON (allowing recording of photon events only during the time-gate) or OFF time-gate (register all photon events but the ones during the time-gate).

To demonstrate fast time-resolved spectroscopy using our 256 × 2 line sensor we studied: i) fluorescence kinetics of chlorophyll A in a Kautsky transient from an intact leaf [7,24,25] of *Viburnum rhytidophyllum*; ii) fluorescence kinetics of a linear FRET probe; iii) mouse lung tissue slide autofluorescence; iv) time-gated Raman spectroscopy of toluene. Our ambition was to demonstrate utility in a broad range of applications to stress-test our engineering implementation, but also to demonstrate how time-resolved CMOS SPAD line arrays lend themselves naturally to numerous applications in biomedical optics.
2. Methods

2.1 CMOS SPAD line sensor architecture

![CMOS SPAD line sensor diagram]

The common aspects with our prior work are as follows [12], the CMOS SPAD line array sensor comprises 2 rows of 256 pixels aiming for efficient photon detection at different wavelengths. Each pixel incorporates 4 SPADs optimized for blue spectral range and 4 SPADs optimized for red spectral range (referred to henceforth as blue SPADs and red SPADs) of 23.78 µm pitch yielding a 43.7% fill-factor. 256 × 26 bit TDC channels integrated on the sensor underpins TCSPC functionality. Detailed analysis of the instrument response function (IRF) and dark count rate results for blue and red SPADs are similar to our prior work [12] and are given in supplementary notes [44]. Three significant improvements were introduced in the work presented here: i) access to data lines coming from pixels was parallelized to increase the readout speed to 19,000 lines per second as opposed to 200 lines per second previously; ii) the time-gating circuit was optimized to work with the SPAD turned ON all the time, i.e. even outside the time-gate window; iii) a new PCB was designed to cater for the improved CMOS SPAD sensor, the firmware was rewritten to take advantage of both improved time-gating and parallelized access and finally, the front-end software was also completely re-written in Python (version 2.7.2) with both a graphical user interface and scripting options. Scripting allows the customization of the data acquisition protocol for specific requirements of any given experiment. For example, plant fluorescence displays particular kinetics where the rise in fluorescence intensity is fast (ms), while the fall is slow. Customizing the script allows combining TCSPC and CMM acquisitions and tailoring the parameters to obtain maximum information from the specimen. Compared to other TDC based CMOS SPAD line sensors, it has the highest line rate to the PC.
The time-gating circuit is illustrated on Fig. 3. TIME_GATE_START and TIME_GATE_STOP are fed into clock pulse (CP) inputs of two D-flip-flops thus defining the time region during which TDC counter is enabled. As TDC counter acts as a reverse start-stop timing clock in TCSPC mode, it is started by photon detection and stopped by laser synchronization STOP signal (electrical TTL or NIM prior to reaching the PCB). The feedback into clear direct input (CD) reduces the risk of glitches. In SPC mode, TDC counter acts as a photon counter so the same time-gating circuit defines the regions during which the photon counting is performed. The timing diagram of the time-gating circuit is shown on Fig. 4 where LASER indicates a laser optical pulse, TIME-GATE is the time-gated region defined by TIME-GATE-START and TIME-GATE-STOP. STOP is the electrical timing pulse (TTL or NIM) received from the optical setup (for pulsed laser diodes STOP comes from the laser usually). STOP can be delayed on-chip and the time-gated region is defined with respect to this delayed version of STOP signal.

The time-gating can be applied in all modes: i) on-chip photon-efficient CMM; ii) SPC mode; or iii) TCSPC mode. In CMM, the time-gated region defines FIRST and LAST time bins required for accurate CMM calculation [42,43]. In time-gated photon counting the designated bin can potentially detect more photon events than in TCSPC mode where each exposure time allows only for one time-stamped photon event to be detected. Higher readout speed of 19,000 lines per second is the key enabler of new applications such as high speed TRFS presented below.
2.2 Sensor characterization – delay line and time-gating

The CMOS SPAD line sensor was placed in the focus of the spectrograph optics. Volume phase holographic grating is used to optimize light throughput. See Methods section for details.

The time-gating properties of the sensor were tested in ambient random light. The light was attenuated to avoid saturation of the pixels and early photon pile-up on the time histograms (photon arrival rate was less than 1% of the laser repetition rate). The 1 MHz laser sync output of the PLP10 laser driver was used as STOP for TCSPC. The position and width of the time-gate is based on a 128 step, on-chip delay line that is separate from the TDC circuits. The sensor allowed programmable delays for the rising and falling edge of the time-gate window with respect to the STOP signal using this delay line. The shortest applicable time-gate window was examined for both ON and OFF operations. Histograms of the gated photon arrival times in the ambient light were created with different gate positions. In both cases (time-gate ON and OFF) the full-width-half-maximum (FWHM) of the peaks with short delay time-gate (10 step delay for rising edge, 11 step delay for falling edge) were measured.

Time-gating was also tested in time-gated SPC mode. A 5 MHz triggering signal generated by the DG645 digital delay generator was connected both to the PLP10 laser driver and the CMOS SPAD line sensor using two separate channels. The first provided triggering of the laser pulses while the latter was used as STOP sync signal for gating. The delay between the two channels was controlled by the delay generator with 5 ps resolution. Scanning through a range of delays between the channels allowed us to move the specified time-gating window across the laser pulse.

2.3 TRFS from leaf, lung tissue, linear FRET probes and time-gated Raman spectroscopy

Figure 5 shows the CMOS SPAD line sensor characterization setup consisting of i) epifluorescence setup which collects fluorescence from the specimens and focuses it into multimode fiber (50 µm for leaf and FRET biosensor and 105 µm for lung slide, 0.22 NA multimode fiber, M14L01, M15L01, Thorlabs Ltd, UK) and ii) spectrograph consisting of dispersive optics with CMOS SPAD line sensor in focus. The specimens were placed in the focus of × 10 microscope (RMS10X, Thorlabs Ltd, UK) objective. Multiple pulsed laser sources were used. Leaf was illuminated by a 402 nm pulsed laser at 20 MHz repetition rate (PLP10, Hamamatsu, Japan). Peak power was 650 mW, pulse length was 54 ps resulting in an optimal 0.7 mW average output (assuming 20 MHz repetition rate). The illumination on the specimen was 0.4 mW due to the efficiency of the epi-fluorescence illumination path. The linear FRET probe were illuminated by a 483 nm pulsed laser diode at 20 MHz repetition rate. Peak power of 483 nm pulsed laser was 349 mW, pulse length was 60 ps resulting in a 0.42 mW average output. The illumination on the specimen was 0.27 mW. For alignment and
spectrograph resolutions tests a WhiteLaseMicro supercontinuum (NKT Photonics-Fianium, UK) was used. Supercontinuum was also used to illuminate the lung slide. The illumination on the specimen was 1.9 mW. Dichroic filters (FF414-DI01-25X36, Semrock, USA, for leaf and custom three color design from Chroma, USA for lung slide) were used to separate illumination from fluorescence. The spectrograph diffraction grating was volume phase holographic grating (600 lp/mm at 600nm, Wasatch Photonics) with collimating and focusing lens optimized for efficiency and spectral resolution [45–47]. Efficiency and wavelength span of the spectrograph were measured using a WhiteLaseMicro supercontinuum and a laser line tunable filter (LLTF Contrast, NKT Photonics, UK). For spectrograph efficiency measurement, the LLTF filter was set to 520 nm with a linewidth of 2.5 nm (verified using an Ocean Optics usb2000 spectrometer) and the optical power measured after the fiber input and after the focusing lens. The efficiency was measured to be 68%. Wavelength span and spectral resolution were verified by taking a spectrum of the fluorescent lamp using a custom spectrograph and an off-the-shelf spectrometer usb2000 (Ocean Optics, USA). To cover the 130 nm wavelength span across the 256 pixels of the line sensor the collimating lens used was a 50 mm focal length achromat (AC254-050-A-ML, Thorlabs, UK), while the focusing lens was 75 mm focal length achromat (AC508-075-A-ML). Optical spectral resolution was measured using fluorescence lamp and measured to be at least 2.5 nm. Fluorescence decays were fitted using non-linear least square fit routines extracted from DecayFit software [48,49].

The specimen for the leaf study was an intact leaf of *Viburnum rhytidophyllum* picked in December 2016 (location Edinburgh, UK) and illuminated for TRFS on the same day. The leaf was dark adapted for 1 hour in a closed box at room temperature upon removal from the tree. The fluorescence kinetics of the leaf was measured both in TCSPC and CMM modes. In TCSPC mode the applied exposure time was 8.3 µs which was followed by 42.9 µs gap due to firmware limitations. TRFS cubes were built from 1000 lines with TCSPC time-events with 8.3 ms overall exposure time over 51.2 ms with an additional gap of approximately 3 ms due to memory management on host computer. After the first 500 TRFS cubes data was collected with additional gaps of 100 ms between TRFS cubes to adapt for the kinetics of leaf fluorescence transient. In CMM mode the applied exposure time was 83.3 µs with an additional 0.25 µs gap due to firmware. Gap introduced by host computer was ~3 ms after every 1000 lines of CMM values. Accurate time point of each line was embedded into a header by the firmware.

The lung specimen slide originated from murine lungs (euthanized) embedded in paraffin wax following a standard protocol and sectioned at 5µm thickness on a microtome. The slides were then re-hydrated by immersing in xylene, 100%, 95%, 80%, 70% EtOH and dH2O in that order. 10 µL of red Inspeck beads 1% (I-7224, Thermofisher Scientific, USA) were also inserted to allow for increased red fluorescence. They were then mounted under a coverslip in water soluble mountant. The exposure time in TCSPC mode was 8.3 µs.

A custom designed FRET molecular biosensor for thrombin [50] was used to study fluorescence kinetics. The compound consists of a linear peptide as a thrombin substrate with 5-carboxyfluorescein as a fluorophore on one end and methyl red as a quencher on the other, see Fig. 6. In the absence of thrombin, methyl red quenches 5-carboxyfluorescein fluorescence, but in presence of thrombin the enzymatic cleavage of the peptide releases the 5-carboxyfluorescein and the fluorescent signal increases. A solution containing the probe (5 µM) in matrix metalloproteinase (MMP) buffer (10mM CaCl2, 6.1 g Tris-HCl, 8.6 g NaCl per litre, pH 7.5) was incubated with thrombin (T9326-150U, Sigma Aldrich, UK) (at a final concentration of 5 U/mL) in 5 × 2 mm microcuvette with a final volume of 300 µL (or in disposable plastic cuvettes with a final volume of 2 mL). Enzyme free reactions were used as a control. Where appropriate thrombin inhibitor Anti-thrombin III (AT3, Sigma Aldrich, UK) was pre-incubated at a final concentration of 0.4 µM. The enzyme was pipetted into cuvette 8...
s after TRFS acquisition was started. The total TRFS scan was performed over 3 minutes 30 seconds. Firstly, 1000 TRFS cubes each with 8.3 ms exposure time over 54 ms elapsed time were acquired followed by 500 TRFS cubes each with 8.3 ms exposure time over 54 ms with an additional 100 ms gap between TRFS acquisition. Data acquisition parameters were selected to cater for fast transients in the beginning and slowed down for slower transients. This allowed us to study kinetics at short and long time scales. The initial delay of 8 s was used to acquire background fluorescence prior to enzyme activation. Diffusion effects were observed inside the cuvette.

![FRET biosensor](image)

Fig. 6. FRET biosensor works by thrombin cleaving the peptide sequence separating the methyl red and 5-carboxyfluorescein (FAM) thus allowing FAM to fluoresce.

The TRFS data cubes for lung slide and custom FRET probe were post-processed by digital bandstop infinite impulse response (IIR) filter to remove the etaloning fringe observed during the experiment. As bare leaf experiments did not display etaloning, we believe this was induced by slide and cuvette. The bandstop frequency was observed in the power spectrum of the spectral curves and the IIR filter parameters were adjusted accordingly (filtering was implemented in Matlab 2014a). Decay data was used from original non-filtered data.

Time-gated Raman spectroscopy was achieved using the same configuration as for lung slide study above. A 10 mm cuvette filled with toluene was placed in front of the microscope objective. As the pulsed laser source had broad linewidth (7 nm), our implementation is in the low resolution Raman spectroscopy (LRSS) regime [51]. We use toluene as a strong Raman scatterer to test the time-gating operation of the circuit. We focus on 3056 cm$^{-1}$ band (~567 nm for 483 nm excitation) to test the time-gating operation. An experiment was performed to acquire Raman spectra with and without time-gating (5.6 ns ON time-gate coinciding with the laser pulse) exposure time was set to 5 s and the laser repetition rate was set to 50 MHz. Toluene sample was illuminated with 0.68 mW.

Due to firmware limitations, our minimum TCSPC exposure time was 8.3 µs. For each experiment, we maximized the number of TCSPC events detected by optimizing respective setups while taking into account both pile-up conditions (1%-10% of repetition rate limit for the number of TCSPC events detected) and counting loss distortions in TCSPC (see page 332 in [2]). To reduce the counting loss distortion the total number of TCSPC events detected on each pixel was set to be less than 50% of the maximum possible defined by the TCSPC exposure time.
3 Results

3.1 CMOS SPAD line sensor characterization - delay line and time-gating

Fig. 7. Time-gated TCSPC histograms of photon arrival times in ambient light. Time-gates were positioned with an on-chip, 128 step delay line with respect to the STOP signal. The width of each time-gate was 1 step of the delay line. The average FWHM for a short delay (10 steps) ON time-gate was 1.44 ns (a). For the OFF time-gate it was 1.56 ns (b).

Figure 7(a) shows the shortest ON time-gate and Fig. 7(b) shows the shortest OFF time-gate for a representative pixel (pixel 100), both at different time positions on the histogram. The rising and falling edges of the time-gate window were adjusted using a 128 step on-chip delay line, with the zero delay on the rising edge causing this edge to be closest to the STOP signal. Longer delays in the window position resulted in wider time-gated regions, as a consequence of accumulated jitter of the steps in the delay line. The same accumulated jitter was responsible for having photons acquired even in the OFF region on the ‘time-gate OFF’ histograms. This can be avoided either by choosing a wider time-gate window, or by applying an external delay. The external delay can shift the histogram to a position where a certain time interval can be masked with a time-gate window that is closer to the STOP (i.e. lower on-chip delay). The average FWHM of the ON peak for 10 step delay is 1.44 ns. The average FWHM of the OFF peak for 10 step delay was 1.56 ns.

Fig. 8. Sweeping an external delay generator over a fixed time-gate window covering a laser pulse generates a more detailed picture of the time-gate (a) convolved with the asymmetric shapes of the laser pulse and red SPAD IRF (b). Blue SPADs were time-gated in (a) as its IRF does not have the diffuse tail present in red SPADs (b) (see supplementary notes [44]).

The resolution of the delay line was measured from the TDC resolution. The distance was calculated between the peaks of the histograms of 10 step and the 110 step delays. The number of TDC bins between the peaks multiplied with their resolution and divided by the
number of steps gives an estimated delay resolution, which was measured to be 1.4 ns. A finer delay can be applied in combination with an external delay generator.

Figure 8 shows the number of photons acquired in time-gated SPC mode with the minimum time-gate as a function of the applied delay with 5 ps resolution between the trigger and sync channels. The asymmetric shape of the resulting plot is due to the asymmetric shape of the detected laser pulses, and the fact that the described measurement results in a convolution of the laser peak with the time-gate window.

3.2 TRFS of leaf

Chlorophyll A transient is shown in Fig. 9 with a fast rise reached in less than 200 ms, while the slow fall in fluorescence intensity takes longer (tens of seconds). Figure 10(a) shows TRFS in the fast rise at 0.03 s, and at several time points in the slow fall, namely, 2.04 s in Fig. 10(b) and 120.04 s in Fig. 10(c). The TRFS curves at the beginning show a slight increase in lifetime during the fast rise while the snapshot at 120.04 s clearly demonstrates a decay in peak values of decays, but also a reduction in lifetime due to quenching. This is further demonstrated in Fig. 11(a), where an increase in fluorescence lifetime is visible during the initial rise. A clearer decrease in lifetime was detected during the slow fall in Fig. 11(b). The TRFS during the fast rise is available in Visualization 1 video, while the slow fall is in Visualization 2 video.

The time-resolved spectra shown in Fig. 10 have two peaks corresponding to 680 nm peak of Photosystem II and 740 nm peak of Photosystem I. The data obtained broadly correspond to prior data obtained using steady state spectrometers [52]. We also verified the spectral shape of leaf fluorescence with off-the-shelf steady-state usb2000 spectrometer (not shown).
Fig. 10. Fluorescent transient of the leaf during fast rise and slow fall. Three TRFS time points at: (a). 0.03 s; (b). 2.04 s; and 120.04 s (c). See Visualization 1 and Visualization 2 for fast rise and slow fall videos respectively. 10 pixels were binned for each 3D plot in (a-b) resulting in 5 nm spectral coverage.

Fig. 11. Fluorescence decays extracted from TRFS data cubes taken during the fast rise (a,c) and the slow fall (b,d). Increase in lifetime on (a) is more obvious with decays fitted for 26.275 ms and 130.188 ms time points (c). Fluorescence quenching is indicative by reduction of lifetime in (b,d) as expected for the chlorophyll A transient. The decays in (a,b) were taken from single pixel (0.5 nm spectral coverage) and from 1 TRFS data cube (8.3 ms exposure time over 51.25 ms) at each time point during fast rise (a), 3 TRFS data cubes (25 ms exposure time over 153.75 ms) around each time point during slow fall (b).
In order to detect the lifetime increase during the fast rise, decays were binned over 150 lines acquired and over all pixels. This provides less noisy decays (Fig. 12) while maintaining reasonably good sampling in time. Spectral information is, however, lost. The lifetime increase during the fast fluorescence rise was also observed previously [16], but that measurement was done by periodic stimulation of chlorophyll A fluorescence kinetics, whereas we measure it in a single chlorophyll A fluorescent transient. Spectral lifetime information with fine sampling in time requires more efficient photon collection for which CMM mode of the chip was applied. CMM can provide lifetime estimates spectrally with much shorter exposure times and hence finer sampling in time. Figure 13 shows the lifetime rising over the first 40 ms of fast rise on the 680 nm and 740 nm peaks, using CMM mode. Several lines of CMM values from the sensor can be processed to have more counts over longer time and hence coarser sampling in time, or binning can be done over pixels reducing spectral resolution.

Fig. 12. Fluorescence decays during the fast rise (a). Lifetime change is more obvious when binning timestamps of all pixels, but no spectral information is available from decays in (a). The decays were created by binning 150 lines of TCSPC time-events of 1.3 ms total exposure time over 7.7 ms around each time point. The decay fits are shown in (b) for two time points.

Fig. 13. Increasing fluorescence lifetime over the first 40 ms of fast rise on the 680 nm and 740 nm peaks. CMM estimates of single pixel with photons captured over 166.6 µs shown on (a) broadly match lifetimes calculated by fitting decays to TCSPC data. Photons captured over 1.67 ms and 30 pixels result in smoother transients.

3.3 TRFS of thrombin FRET biosensor

Continuous TRFS acquisition before, during and after enzyme activation is illustrated in the figures below. Fluorescence intensity variation with time (extracted from TRFS) is shown in Fig. 14. The laser is switched on in the first second and the dip at 8 s indicates the time when the enzyme thrombin was pipetted into cuvette. The short dip in fluorescence intensity is likely to be due to movement of fluid during pipetting.
Fig. 14. Enzyme kinetics curve of a thrombin FRET biosensor. At 8 s thrombin was pipetted into a cuvette initiating a rise in fluorescence intensity. TRFS was acquired in 8.3 ms every 50 ms during the first 50 s (every 150 ms after 50 s) and the kinetics curve was derived from underlying spectral double kinetics data.

Fig. 15. Sample TRFS from the time point at 130 s. Spectral resolution is 0.4 nm and no binning was applied in the spectral domain.

The choice of acquisition parameters was led by memory requirements of 130 s acquisition. Sample TRFS data cube at full spectral resolution is shown on Fig. 15. Figure 16 shows decays extracted from the spectral peak (20 pixels, spectrum covered 8 nm) before and after enzyme activation. The lifetime change is minor indicating that background fluorescence is mostly due to unquenched 5-carboxyfluorescein. Increased time resolution and improved jitter will allow better decay analysis. The increase in fluorescence is due to 5-carboxyfluorescein being cleaved from the FRET biosensor. A similar effect has been observed for calcium dye Fluo-4am [16,53].
3.4 TRFS of ex vivo lung

TRFS of ex vivo lung slide acquired in 258 ms is shown in Fig. 17. 10 pixels were binned for the TRFS in Fig. 17 resulting in 6 nm spectral resolution. The TRFS acquisition speed allows for point based analysis in vivo. For display purposes, etaloning fringes were removed in the same way as with FRET biosensor. We detect sharp broadband spectra at 15 ns in Fig. 17, which requires further investigation. It does not affect the spectral decays studied. The reduction in lifetime from 3.36 ns to 2.12 ns shown on Fig. 18(a) is due to presence of 1% Inspeck beads. This is also shown in lifetime change with wavelength on Fig. 18(b), top curve. Reduced $\chi^2$ is above 3 across the spectrum (Fig. 18(b), bottom curve) indicating multi-exponential nature of the decay.

Fig. 16. Decays extracted from the fluorescence peak before enzyme activation (2 s time point) and after enzyme activation (130 s time point). 20 pixels were binned for both decays covering 520-528 nm.

Fig. 17. TRFS of ex vivo lung acquired in 258 ms. 10 pixels were binned to obtain spectral resolution of 6 nm. Autofluorescence did not vary over 50 s measurements so no transient was observable.
However, if we use CMM for lifetime estimate, then spectral CMM can be acquired faster. The total number of photon events detected in 10 µs is ~1000 for the band 500 nm to 550 nm (CMM lung slide experiment data not shown). This indicates that a single binned CMM value for the whole green band is obtainable in 2 µs assuming CMM value is correct for 200 photon events. This is relevant when considering photon budgets required for laser scanning confocal FLIM, because it is indicative of 500000 lifetime estimates per second which is equivalent ~8 frames per second with 256 × 256 frames. The mouse lung slide used did not contain any fluorescent dye and it was 5 µm thick, so in vivo tissue imaging is likely to produce more autofluorescence. Furthermore, by deploying molecular imaging smartprobes [54,55] across the visible-NIR spectrum will allow us to use the CMOS SPAD line array for fast FLIM applications.

Fig. 18. Decays (from ex vivo lung tissue) with fits for green band (524 nm) and red band (697 nm) (a). Lifetime reduction in red band is shown in (b), top curve. Lifetime reduction is due to red Inspeck beads. Spectral resolution of each decay was 6 nm.

3.5 Time-gated Raman spectroscopy of toluene

Figure 19 demonstrates the time-gating performance of the CMOS SPAD line sensor. Time-gating was done in SPC mode. Exposure time was 5 s. The spectrum shown is from toluene focusing on expected Raman peak at 3056 cm⁻¹. Due to LRRS regime (linewidth of the laser is ~250 cm⁻¹) the peak is broad, but this still allows us to study the time-gating circuit performance. The background noise is suppressed by a factor of 3.6 by reducing the window for which the detector is on, thus reducing DCR. This is close to expected value of 3.57 (20 ns / 5.6 ns). Further demonstration of the application of this detector for removal of fluorescent background signals (in addition to DCR) has been shown elsewhere [13].

The background noise for both curves was estimated as standard deviation of values between 2000 cm⁻¹ and 2500 cm⁻¹ which are expected to be non-fluorescent. The main component of the background noise is DCR. However, besides DCR suppression, we can also observe that the main 3056 cm⁻¹ signal peak is reduced from 3600 without time-gating to 1905 photon counts with time-gating. This needs further investigation and testing, but this is at least partially due to the time-gate missing a part of the diffuse tail of the red SPADs (see Fig. 8(b)).
4. Discussion

We demonstrated the versatility of the CMOS SPAD line array in several applications, but detailed work remains to be done. The main issues observed have been the coarse time resolution (0.43 ns), coarse jitter (at least 0.62 ns, but often > 1 ns) and etaloning. Coarse timing limits lifetime analysis in the plant specimen where lifetime values are <1 ns. Better jitter and time resolution have been designed previously for time-resolved CMOS SPAD cameras [56], where time bin resolution was ~50 ps and the jitter was less than 200 ps. The aim for forthcoming CMOS SPAD line sensors [57] will be to introduce improved timing performance which will significantly enhance TCSPC TRFS and time-gating capability (e.g. DCR suppression). Etaloning was observed during lung slide and custom FRET probe experiments, but it was not observed during leaf experiments. This needs further investigation, as sensor induced etaloning expected from front illuminated light sensors should be observable in all specimens. Our current understanding is that the etaloning observed is induced by cuvette or slide.

Another important point is the multiplexing capability of pixels containing multiple SPADs (4 SPADs per pixel in our case), because this decides the minimum time required to obtain enough photons for lifetime estimation. While the number of photons required differs for single exponential and multi exponential fits, many applications can deal with lifetime derived from single exponential fits. Assuming 50 ns dead-time per SPAD, the number of photons that can be processed by a single TDC is 20 million photon events per second. This applies to TDC architectures with negligible conversion dead-time as is the case for gated ring oscillator architectures deployed in the CMOS SPAD line sensor presented. This fits nicely with using a 20 MHz pulsed laser. However, the situation changes if there are multiple SPADs per TDC per pixel. As soon as the TDC receives a STOP from the laser trigger (see Fig. 4) it is ready to be started by any of the remaining 3 SPADs while the SPAD used immediately prior to STOP is in dead-time. Therefore, in situations when there is a photon burst or samples emit many photons, the multiplexing allows more events to be processed. This is especially true if the repetition rate of the laser is increased. The key advantage brought by massively parallel CMOS SPAD timing circuits is that they can beat 1% pile-up limits by a wide margin [58] and exploit a variety of photon bursts to the full. The best results
we achieved are in chlorophyll A fluorescence transient analysis (section 3.2 above). We plan to expand this to spectral lifetime analysis of other fluorescence effects such as quenching and photobleaching and any transients observable in tissue autofluorescence and molecular imaging. A natural progression for our sensor is to implement it in confocal FLIM setup and demonstrate its performance in spectral imaging [11]. Frequency domain FLIM has already been deployed on plant cell FLIM at impressive speeds (26 fps) [59,60]. The advantage of massively parallelized CMOS SPAD sensors such as ours is that it can shift the study from measuring static fluorescence lifetime in either imaging or spectroscopy regime to measuring rapid changes in lifetime, intensity, spectra and other variables thus offering unique perspective in dynamic optical phenomena.

5. Conclusion

We demonstrate fast and flexible TRFS acquisition using TCSPC. Parallelization of TDC and time-gating speeds up the acquisition significantly for a range of applications and will enable advanced metrology such as spectral fluorescence kinetics in the ns and ms domains. Undoubtedly, 3D wafer stacking and process optimization [61] will enable faster, lower DCR and more integrated time-resolved sensors.

Funding

We would like to thank the Engineering and Physical Sciences Research Council (EPSRC, United Kingdom) Interdisciplinary Research Collaboration (grant number EP/K03197X/1) for funding this work. We would also like to thank EPSRC and MRC Centre for Doctoral Training in Optical Medical Imaging, OPTIMA, (grant number EP/L016559/1) for access to the supercontinuum laser. We would also like to thank ST Microelectronics, Imaging Division, Edinburgh, for their generous support in manufacturing of the CMOS SPAD line sensors.

Acknowledgments

We appreciate assistance from Dr Neil Finlayson, University of Edinburgh, with LRRS. Link to data sets can be found in [44].