
This version is available at https://strathprints.strath.ac.uk/49457/

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Unless otherwise explicitly stated on the manuscript, Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Please check the manuscript for details of any other licences that may have been applied. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (https://strathprints.strath.ac.uk/) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to the Strathprints administrator: strathprints@strath.ac.uk
Abstract: We have successfully demonstrated video-rate CMOS single-photon avalanche diode (SPAD)-based cameras for fluorescence lifetime imaging microscopy (FLIM) by applying innovative FLIM algorithms. We also review and compare several time-domain techniques and solid-state FLIM systems, and adapt the proposed algorithms for massive CMOS SPAD-based arrays and hardware implementations. The theoretical error equations are derived and their performances are demonstrated on the data obtained from 0.13 μm CMOS SPAD arrays and the multiple-decay data obtained from scanning PMT systems. In vivo two photon fluorescence lifetime imaging data of FITC-albumin labeled vasculature of a P22 rat carcinosarcoma (BD9 rat window chamber) are used to test how
different algorithms perform on bi-decay data. The proposed techniques are capable of producing lifetime images with enough contrast.

**Keywords:** fluorescence lifetime imaging microscopy; FLIM; CMOS; single-photon avalanche diode; single-decay; multi-decay; time-domain FLIM; lifetime based sensing

---

1. Introduction

Fluorescence based imaging techniques have become essential tools for a large variety of disciplines, ranging from cell-biology, medical diagnosis, and pharmacological development to physical sciences. Fluorescence lifetime imaging microscopy (FLIM) yields images with each pixel sensing the exponential decay rate instead of the intensity of the fluorescence from a fluorophore. It has the advantages of insensitivity to variations in excitation illumination, photon scattering and probe concentration, and it allows the resolution of different fluorophores (with different lifetimes) fluorescing at the same wavelength. FLIM can be applied to acquire more quantitative information about physiological parameters such as pH, Ca$^{2+}$, pO$_2$, etc. or image intracellular functions with Förster resonance energy transfer (FRET) techniques [1–3].

Both time-domain and frequency-domain instrumentation systems are available to acquire FLIM data. Here, we will focus on the recently proposed time-domain approaches and discuss how they can be applied to the latest solid-state sensor arrays for FLIM applications. For detailed discussions in frequency-domain FLIM systems, please see [4–7]. In a typical time-resolved FLIM experiment, the samples with fluorescent markers are illuminated by a pulsed laser and the time-correlated photons emitted from the markers are collected by detectors. Commercially available FLIM systems usually use photomultiplier tubes (PMT) or multiple channel plate (MCP) PMTs plus time-correlated single-photon counting cards (TCSPC) [8] or gated intensified/electron-multiplying CCDs to measure the lifetimes. The latest multi-channel PMT systems can significantly increase the imaging speed, but they still require image-scanning. For example, to avoid local heating and photobleaching, the pixel dwell time is set to be 15.25 $\mu$s and the sample is scanned hundreds of times, say 200 times, to accumulate enough photon counts. It will take $256 \times 256 \times 200 \times 15.25$ $\mu$s/$N_p \sim 200$ s/$N_p$ ($N_p$ is the number of channels) to generate a $256 \times 256$ lifetime image. CCD based systems usually require coolers and/or high voltage supplies. Compared with CMOS compatible solutions, the above mentioned systems are expensive, bulky, and not user-friendly. Recent developments on CMOS-based FLIM systems can be found in [9–20]. Kawahito applies pin photodiodes whereas Shepard uses active pixel sensors, but both employed the 2-gate rapid lifetime determination (RLD) method [21] to calculate the lifetimes. Li *et al.* proposed several non-gating time-domain lifetime algorithms and demonstrated video-rate lifetime imaging [14,15] on single-photon avalanche diode (SPAD) plus in-pixel TCSPC arrays [22]. Unlike conventional CCD based sensors, a SPAD is a p-n junction reverse biased above the breakdown voltage to sustain the avalanche multiplication process triggered by photogenerated carriers. The transfer gain is so large that the output current from the SPAD can be easily converted into a digital signal without using complex front-end amplifiers deteriorating the signal-to-noise ratio (SNR). With such single photon sensitivity, SPADs are suitable for photon-starved
applications such as single molecule detection [23,24], fluorescence lifetime measurements [13,20], optical range finding, optical fiber fault detection, and portable explosives sensing [25]. Recent developments of CMOS SPADs have shown significant improvements in the dead time [26], dark count [27], technology migration to advanced process [28] and pixel miniaturization [29], and quantum efficiency in the longer wavelength region [30]. It is expected high resolution CMOS SPAD arrays for ranging applications [31] will soon be applied to FLIM applications.

Another bottleneck for high-speed lifetime imaging is lifetime calculation. Common FLIM systems usually use iterative linear or non-linear least square methods (LSM), such as Marquardt-Levenberg algorithms, to extract the lifetimes. Although this approach is accurate and suitable for analyzing multi-exponential decays, it is computationally time consuming which makes it unsuitable for real-time applications. It is desirable, therefore, to develop non-iterative simple algorithms to speed up the lifetime calculations while maintaining enough imaging quality. Compared with the LSM, iterative-free gating methods only require two time bins for single-exponential decays [9,13,21], four bins for bi-exponential decays [32,33] or eight bins for multi-exponential decays [34]. The hardware complexity is greatly reduced and the speed is much higher. There are different acquisition schemes for the gating methods. Figure 1(a) shows the traditional sequential acquisition in a pixel, where at least two sub-images are recorded sequentially at different delayed windows with respect to the excited laser pulses to extract the lifetime. The block ‘counter’ can contain front-end circuits, analog-to-digital converters and accumulators in conventional imaging systems or simply inverters and digital buffers in the latest CMOS SPAD systems. Chang and Mycek applied four time-gates to analyze single-exponential decay data [35]. This approach is slow and sensitive to motion artifacts unless the samples are stationary, and the recorded sub-images are uncorrelated. If a full fluorescence emission histogram is needed for detailed examinations, it will take a significant amount of time to record a large number of sub-images with different delay times [36].

Figure 1. (a) Sequential acquisition in a pixel; (b) parallel acquisition in a super pixel (more than one detector in a pixel) or multi-channel gated optical intensifiers; (c) parallel acquisition with gated counters; (d) parallel acquisition with binned pixels and gate counters; (e) single-channel PMT+TCSPC or in-pixel SPAD+TCSPC acquisition.
Unlike the sequential acquisition scheme, some researchers applied ‘single-shot’ acquisition methods by applying multi-channel segmented gated optical intensifiers (GOI) [37] or focusing optically delayed images on different sections of the detector [38,39] to collect sub-images simultaneously. Video-rate FLIM has been demonstrated in GOI FLIM systems. This trades off spatial resolution for imaging speed. Similar to the sequential acquisition scheme, the detectors enabled by the gating signals only work ‘part-time’. The background level for the two detectors could be different, and the two recorded events are independent (uncorrelated).

Figure 1(c) shows a different acquisition scheme from above. The detector works fully, but with several gated counters collecting photons in different timing windows. The gated counter accumulates only when the leading edge of the detector output is inside its enable period. The gating signals for controlling the counters can be non-overlapping [36–40] or overlapping [41]. Gerritsen et al. applied several non-overlapping gated counters to confocal scanning systems [40]. The system can deal with multiple-decay data. The overlapping gating acquisition, denoted as ORLD, was first proposed by Chan et al. [41] and it was proven to be efficient by running Monte Carlo simulations, but the authors only suggested splitting the PMT signal and sending to two gated counters without actually carrying out the experiments probably because of the complexity and the cost of reconfiguring the existing acquisition systems. Later, Elangovan et al. proposed an overlapping multi-gated algorithm to analyze double-exponential decays and applied Monte Carlo simulations to find suitable experiment settings [32]. For the latest developed CMOS SPAD arrays, however, the ORLD is a good option considering the system complexity and photon efficiency (the photon events collected in the gated counters are correlated if the overlapped proportion is significant). We will for the first time derive the theoretical error equations and test the ORLD methods on the data obtained from a 0.13 $\mu$m CMOS SPAD array.

Stoppa’s research team combined the acquisition schemes (b) and (c) by binning 4 and 25 CMOS SPADs, respectively, in a ‘super’ pixel and using gated counters to collect the photon events [11,12] as in Figure 1(d). Similar to the second gating scheme, this technique trades the spatial resolution for imaging speed. It provides faster acquisition in some sensing applications such as high throughput screening. They used non-overlapping time gates to collect the photons and used nonlinear LSM to extract the lifetimes.

Figure 1(e) shows a pixel with time-correlated single-photon counting (TCSPC) circuitry. The detector works fully, with a time-to-digital converter (TDC) to record the time tag of each photon. Compared with the hardware for time-gating acquisition, the complexity of the TCSPC systems is much higher. The latest commercial multi-channel systems only contain several PMTs and still require scanning. Recently we have demonstrated a 1024-channel, a 20 k-channel and a 100 Mphoton/s multi-channel FLIM systems with 0.13 $\mu$m CMOS SPAD plus in-pixel TDC (10-bit TCSPC) arrays [14–16,20]. The parallelism of the proposed systems can be fully exploited by implementing lifetime calculations in hardware [42,43]. Similar to the above-mentioned ‘single-shot’ systems, the imaging speed of the proposed cameras can operate over 100 f/s.

2. Time-Domain FLIM Data Analysis

For simplicity, we consider a single-exponential decay with a negligible impulse response function (IRF). Such assumptions allow a proper comparison of various fitting algorithms. Moreover, a
single-exponential decay model is still enough to contrast different types of fluorophores. We will discuss the accuracy of FLIM algorithms and compare the performance of various algorithms on bi-exponential decays. Assume that the fluorescence decay function \( f(t) = A \exp(-t/\tau)u(t) \) with \( \tau \) being the lifetime and \( u(t) \) the step function.

2.1. Gating Techniques

Gating algorithms have been widely used in real-time FLIM systems owing to their simplicity and compactness in hardware implementations. The simplest 2-gate RLD (RLD-2) only uses two time gates of the same width. We adopt the \( F \) value introduced in [44] to quantify the performance of a FLIM algorithm. The \( F \) value is defined as \( F = N_{c}^{1/2} \sigma/\tau \), where \( \sigma \) is the standard deviation of repeated measurements of the lifetime value \( \tau \) and \( N_{c} \) the total count in the measurement window from \( t = 0 \) to \( t = T \), the average total count \( EN_{c} = \int_{0}^{T} f(t)dt = \tau A \exp(\tau t) \). The \( F \) value for the RLD-2 can be derived from [21] and [44] as:

\[
F_{RLD} = \sqrt{N_{c}} \frac{\sigma_{\tau}}{\tau} \sim \frac{x + 1}{h \sqrt{x}},
\]

where \( x = \exp(-h/\tau) \) with \( h \) being the gate width. The two gates can be generalized to be overlapping with unequal widths as shown in Figure 2, denoted as GRLD hereafter, where \( N_{1} (= N_{o1} + N_{ov}) \) and \( N_{2} (= N_{ov} + N_{ov2}) \) are the total counts of the first \((0 < t < h)\) and second \((Sh < t < Rh)\) gates respectively. Chan et al. [41] used Monte Carlo simulations to get an optimized setting \((S = 0.25, R = 12.25)\) and mentioned that the PMT signal could be split and sent to two gated integrators, but the experiments were not actually carried out. Their “optimized setting” were not based on the theoretical derivations, and we will present theoretically optimized settings that are different from their suggestions. We will also compare its performance with other time-domain algorithms. From the discussions in Section 1, the GRLD can be applied to different gating schemes shown in Figure 1(a–d). We divide these acquisition schemes into two categories. The first category is uncorrelated, where the overlapping parts are uncorrelated, denoted as UGRLD hereafter, because they are recorded in different snapshots or in different gated detectors. The photon events \( N_{1} \) and \( N_{2} \) are mutually independent and therefore uncorrelated. The second category is correlated, denoted as CGRLD hereafter, because the overlapping counts, \( N_{ov} \), collected by different gated integrators are from the same photon events. The \( F \)-values for the UGRLD and CGRLD can be derived directly from Equation (A5) in the Appendix (using similar derivation methods and notations with [45]) as:

\[
F_{UGRLD} = \frac{\sigma_{\tau}}{\tau} \sqrt{N_{c}} = \left(1 - S\right)x^{R+1} + \left(R - 1\right)x^{R+1} + Sx^{S} - Rx^{R}
\]

\[
k_{1}(x) = (1 - x)(1 - x^{R})(x^{S} - x^{R})(1 - x + x^{S} - x^{R}),
\]

\[
F_{CGRLD} = \frac{\sigma_{\tau}}{\tau} \sqrt{N_{c}} = \left(1 - S\right)x^{R+1} + \left(R - 1\right)x^{R+1} + Sx^{S} - Rx^{R}
\]

\[
k_{2}(x) = k_{1}(x) - 2(1 - x)(1 - x^{R})(x^{S} - x)(x^{S} - x^{R}).
\]
Equation (2) allows us to determine an optimized setting without resorting to Monte Carlo simulations. Moreover, setting \( g(x) = 0 \) in Equation (A2), we can obtain:

\[
\frac{x^S(1-x^{R-S})}{1-x} \Rightarrow \frac{N_2}{N_1} \Rightarrow x^S \sum_{i=0}^{R-S-1} x^i = \frac{N_2}{N_1} \Rightarrow x^{\left(\sum_{i=0}^{R-S-1} x^i\right)^{1/S}} = \left(\frac{N_2}{N_1}\right)^{1/S},
\]

where both \( S^{-1} \) and \( (R-S) \) are integers. It is easy to build a look-up table for Equation (3) on FPGAs and make the calculations much faster as in [15].

Some researchers used multigate RLD (RLD-M) to calculate the lifetimes [35,36,42]:

\[
\tau_{RLD-M} = \left(\frac{\sum_{j=0}^{M-1} t_j}{M}\right)^2 - M \sum_{j=0}^{M-1} t_j^2 \left[ M \sum_{j=0}^{M-1} \ln(N_j) - \sum_{j=0}^{M-1} t_j \sum_{j=0}^{M-1} \ln(N_j) \right],
\]

where \( N_j \) is the count number in the \( j \)th time bin (\( t_j < t < t_{j+1} \)), \( j = 0, \ldots, M-1 \). Similar to Equation (A2), an average lifetime is obtained from Equation (4) if the fluorescent emission is a multi-exponential decay. The theoretical error equation, Equation (4), only exists when the bins are equally spaced, and it was for the first time derived in [42] considering also the impact of timing jitters and later derived in [35] for a special case (\( M = 4 \)) incorporating filtering techniques. Neglecting the impact of the TDC jitters, the \( F \)-value can be re-written from Equation (25) in [42] as:

\[
F_{RLD-M} = \frac{6\tau}{M(M^2-1)h} \sqrt{\frac{1-x^M}{1-x}} G(x), \quad x = \exp(-h/\tau),
\]

\[
G(x) = \left(\frac{1}{x-1}\right)^3 \left[ (M-1)^2 \left( x^{M-2} - 1 \right) + (6-2M^2) \left( x^{M-1} - x^{-1} \right) + (M+1)^2 \left( x^{-M} - x^{-2} \right) \right],
\]

where \( h \) is the bin width. Figure 3(a,b) shows the \( F \)-value curves in terms of the ratio of the lifetime over the measurement window for the previously reported and our center-of-mass method (CMM) [15], integration for lifetime extraction method (IEM) [42], Equations (1), (2) and (5). The 256-gate maximum likelihood estimation (MLE) offers the best precision with a very wide optimized window. RLD-2 has its best performance at \( T \sim 4.8\tau \) (or \( h = 2.4\tau \)) [21], and it usually requires to know the lifetime range of the samples in advance for tuning a proper gate width. Theoretical (solid lines) and Monte Carlo simulation (marked with circles or crosses) results of the UGRLD and CGRLD show good agreement with Equation (2). For extracting single-decay lifetimes, the performance of RLD-M is only better than the other gating techniques in the region \( T \sim \tau \) and increasing \( M \) over eight does not improve the precision any further. This conclusion contradicts the comment stating that the RLD-M is more precise than RLD-2 made in [35]. In their experiments, they applied four time gates on single-exponential decays. The advantage of using the multiple time gate technique, however, is its capability to image multi-exponential decays [11,34] using linear or nonlinear LSMs or to image bi-exponential decays using the fast algorithm proposed by Sharman and Periasamy [32,33]. The performance of UGRLD is worse than that of CGRLD. For the same acquisition time, the gated integrator acquisition schemes shown in Figure 1(c,d) provide a better precision than those in Figure 1(a,b). Chan et al. suggested an optimized setting (\( S = 0.25, R = 12.25 \)) [41], where the best resolving window is roughly from \( T \sim 12 \) to 100\( \tau \) with the duty cycle of their lifetime measurements.
being comparatively low. Figure 3(c,d) show two $F$-value plots with ($R$-$S$) fixed and with $S$ fixed respectively. For $R = S + 3$, a smaller $S$ results in a wider optimized window allowing to resolve smaller lifetimes, but the minimum $F$ value occurs at about $S = 0.5$. It is obvious that a bigger $R$ with a proper $S$ (around 0.2 to 0.7 considering the lifetime resolvability range) allows resolving lifetimes much less than the measurement window. The lower bound of the resolvable lifetime range, however, is limited by the full width at half maximum of the IRF, so choosing a much bigger $R$ is not realistic. Instead of using the setting suggested by Chan et al., we suggest an optimized setting ($S = 0.2$ and $R = 3.2$) from Equation (2) to provide a comparable performance with RLD-2 in $T < 5 \tau$ while maintaining an enough resolvability up to $T \sim 50 \tau$. For a laser repetition rate of 80 MHz, the minimum resolvable lifetime is about $200 \sim 300$ ps, in the same order of the timing jitters of the latest developed 0.13 μm CMOS SPADs [22]. The minimum $F$-value of the CGRLD is about 1.2 and its simplicity in hardware implementation is promising for massive sensor arrays. In Figure 3(b), the IEM-7 (7-gate IEM) [42] and CMM-256 (256-gate CMM) [15] are non-iterative algorithms suitable for TCSPC systems.

**Figure 2.** (a) Generalized two-gate rapid lifetime determination and (b) multi-gate or TCSPC acquisition.

**Figure 3.** $F$-value versus $\tau/T$ plots of (a) previously reported 256 gate MLE, 2-gate RLD, 10-gate RLD (RLD-10), CGRLD/UGRLD with $S = 0.25$, $R = 12.25$, and (b) the proposed 256-gate CMM, 7-gate IEM, CGRLD/UGRLD with $S = 0.2$, $R = 3.2$. 3-D $F$-value plots of (c) $R-S = 3$ and (d) $S = 0.2$. 

![Figure 2](image1.png)

![Figure 3](image2.png)
2.2. Non-Iterative Algorithms Suitable for TCSPC Systems

Although the gating methods introduced above can provide fast lifetime preview and are easy to implement, scientists still prefer to keep the raw timing data for detailed examinations on what really happens in the samples. For gating methods, recording raw timing data is not easy. To obtain a full histogram, a large number of snapshots at different time delays are acquired sequentially. It is very inefficient. To increase the raw data collecting speed, the number of gated integrators can be increased as in [34], but at a cost of system complexity. To enhance the raw data collection rate, we incorporated on-chip high-resolution TDCs into the CMOS SPAD arrays and built 1024-channel and 20k-channel “miniaturized PMT+TCSPC” prototypes [16]. The prototypes can operate in the RAW MODE for collecting the raw data and in the COARSE MODE for high-speed lifetime imaging preview. Due to the slow speed of the widely used iterative based least square methods, several time-domain FLIM algorithms suitable for TCSPC systems (Figure 1(e)) have been proposed to improve the imaging speed [42,43]. We proposed an innovative background-insensitive algorithm suitable for our SPAD prototypes, called IEM. The calculated lifetimes are:

$$\tau_{IEM} = \frac{h \cdot \sum_{j=0}^{M-1} (\bar{C}_j N_j)}{N_0 - N_{M-1}},$$

where $\bar{C} = [1/3, 4/3, 2/3, \ldots, 4/3, 1/3]$ from the Simpson’s integration rule [42], $h$ is the TDC bin width, and $N_j$ is the count number in the $j$th time bin. The precision and accuracy are detailed in [42]. Compared with commonly used RLD-2 and MLE, IEM is much less sensitive to the background noise [46]. Moreover, the lifetime calculations are very simple and suitable for on-FPGA or on-chip implementations. On-FPGA real-time wide-field lifetime imaging has been successfully demonstrated on microfluidic mixing [14]. Figure 3(b) also shows an $F$-value curve for the 7-gate IEM with a total count of 1024 (IEM is a biased estimator and the quantization error which comes to effect at $\tau < 0.1T$ (for $M = 7$) is included in the $F$-value curve as in [42]. Without considering the impact of the background noise, the error performance of IEM-7 is similar to RLD-2. The effect of increasing $M$ in IEM is similar to increasing $R$ in GRLD; both cause their $F$-value curves to shift towards the smaller $\tau/T$ region. To further enhance the lifetime imaging speed on our SPAD prototypes, we proposed
another hardware implementation algorithm based on center-of-mass methods. The calculated lifetime defined in [43] is re-written as:

\[
\tau_{\text{CMM}} \sim \frac{\int_0^T tf(t)dt}{\int_0^T f(t)dt} = \tau \frac{T e^{-T/\tau}}{1 - e^{-T/\tau}} \tag{7}
\]

\[
\tau_{\text{CMM}} = \left( \sum_{i=1}^{N_c} \frac{\bar{D}_i}{N_c} + \frac{1}{2} \right) h, \quad \tau = \Omega \left( \frac{\tau_{\text{CMM}}}{T} \right). T = \Omega \left[ \frac{1}{M} \left( \sum_{i=1}^{N_c} \frac{\bar{D}_i}{N_c} + \frac{1}{2} \right) \right]. Mh, \tag{8}
\]

where \(\bar{D}_i\) is the 10-bit TDC output of the \(i\)th captured photon and \(\Omega(\cdot)\) is a 1D look-up table for calibrating the error term in Equation (7). Its precision and accuracy Equations were derived in [15], and its \(F\)-value curve (shown in Figure 3(b)) considering the accuracy term for 256-gate CMM with \(N_c = 1,024\) showing great photon effectiveness with a wide optimization range. We applied Equation (7) to achieve high-speed lifetime sensing [20,43] and imaging [15]. The lifetime calculations are mostly carried out on FPGAs. To demonstrate its performance, a widefield microscope was adapted to accommodate the SPAD array. A dilute solution of 10 \(\mu\)m fluorescent beads (G1000, Duke Scientific, Palo Alto, CA, USA) was used to simulate cells flowing through the system. The fluorescent emission was captured by the SPAD camera and the delay of every captured photon with respect to the laser pulses was calculated by the in-pixel TDC. The 10-bit CMM lifetime codes were generated from the FPGA through Equation (8), and the lifetimes were easily calculated with the characterization data of the TDC array. Figure 4(a) shows CMM codes at a video frame rate of 50 fps, and Figure 4(b) shows a single frame of the estimated lifetime video. The size of the bead is about 10 \(\mu\)m, and the speed of the beads is 200 \(\mu\)m/s. The experimental settings are optimized for the TDC sub-array with the resolution \(h = 78\) ps, Rows 1–16 [22], and the lifetime codes obtained from the other part (\(h = 160\) ps, Rows 17–32) are normalized. During one frame time the bead moves approximately 4 \(\mu\)m, 40% of its width, producing non-trivial motion blur. The average lifetime is around 2 ns, in a good agreement with the lifetime obtained by using the burst integrated fluorescence lifetime (BIFL) measurement with a Becker & Hickl card (SPC850). The microscope was based on a Nikon Eclipse Ti-E inverted microscope platform. Illumination was via a polarization maintaining fibre to the Nikon TIRF attachment using a 4 W supercontinuum laser (Fianium SC400) to provide picosecond laser pulses at 20 MHz. Excitation wavelengths are selected using bandpass filters prior to launching into the fibre. In this instance a 488 nm bandpass filter was used (470 ± 22 nm, Chroma Inc, Bellows Falls, VT, USA) unless stated otherwise. Fluorescence lifetime images were acquired with a Nikon 40\(\times\) Plan Fluor Objective (0.75 NA) using the 0.13 \(\mu\)m CMOS 32 \(\times\) 32 SPAD+TDC chip [22] arranged in the image plane of the microscope side port. The microfluidic system consisted primarily of a Mitos T-Junction Chip (100 \(\mu\)m channel) connected via 1/16” Teflon tubing to a Mitos P-pump (The Dolomite Centre, Royston, UK). The instrument response function (IRF) was measured and found to have a FWHM of 0.6 ns.
The analog version of Equation (7), denoted as analog mean-delay (AMD) hereafter, has been reported earlier using stand-alone components for lifetime sensing applications [47]. Recently, an AMD instrumentation considering the impact of IRFs and limited measurement window problems has been carried out almost at the same time with our digital CMM by Kim [48,49]. Their algorithms can work efficiently with low-cost single-channel PMTs, and it has been demonstrated in a confocal scanning microscope [50]. Padilla-Parra et al. also used Equation (7) to calculate the minimal fraction of interacting donor (MFD) in fluorescence resonance energy transfer (FRET) experiments [51,52]. The measurement window of the MFD, however, should be much larger than the largest lifetime component; otherwise the contrast capability would be lost, similar to the problems revealed in [15] and [49]. By automatically calibrating such problems, the MFD can be easily applied to our SPAD systems for FRET applications as well. Recently, a Bayesian FLIM analysis method was proposed, and it outperformed the commonly used MLE and LSM especially on noisy single-exponential data sets [53,54]. This algorithm is promising especially in single-cell cytometry. The Bayesian methods for multi-exponential decays are under development.

2.3. Error Performance of FLIM Algorithms on Data Collected by CMOS SPAD Arrays

To compare the error performance of different FLIM algorithms, widefield lifetime images of a uniform aqueous solution of Rhodamine B were taken in different acquisition time. The imaging system was set up on a Nikon TE2000U inverted microscope. The excitation source was a PicoQuant pulsed diode laser with a wavelength of 470 nm coupled through the epi-fluorescence port of the microscope using a Nikon B-2A filter cube. The laser pulse rate is 20 MHz and the maximum power reaching the back focal plane of the objective is about 90 µW. The sample was imaged onto the 32 × 32 SPAD camera directly attached to one of the camera ports using a 20× objective (Nikon Plan Apo, 20×, NA 0.45). The TDC resolution is set to be 78 ps with the on-chip phase-locked loop enabled [22]. Two measurement windows of \(Mh \sim 4.1\tau\) and \(Mh \sim 17\tau\) were chosen to calculate lifetimes. Figure 5(a,b) show the precision curves in terms of photon counts for different algorithms. The ideal curve is the theoretical precision of MLE (or CMM) without considering system non-idealties. CMM.
and MLE behave almost the same in both plots, and they are about 0.5 ~ 1.5 dB away from the ideal curve. The first window $Mh/\tau \sim 4.1$ is optimized for RLD-2, where RLD-2, IEM, and CGRLD ($S/R = 0.2/3.2$) work equally well. Figure 5(b) shows the reason why we need a fast FLIM algorithm other than RLD-2. The CGRLD with $S/R = 0.25/12.25$ still performs worse than the CGRLD with $S/R = 0.2/3.2$ making it inefficient in photon collecting due to low duty cycle operations (optimized window $> 17\tau$). The conclusions drawn from Figure 5(a,b) are in good agreement with the theoretical Figure 3(a,d).

Figure 5. Precision plots for (a) $Mh = 4.1\tau$ and (b) $Mh = 17\tau$.

2.4. Contrast Capability of Algorithms on Bi-Exponential Decays

Assume a bi-exponential decay $f(t) = A_1 \exp(-t/\tau_1) + A_2 \exp(-t/\tau_2)$ with $\tau_1$, $\tau_2$ being the lifetimes and $A_1$, $A_2$ being the pre-scalars. Although the above FLIM algorithms are for single-exponential data, it is worth comparing the average lifetimes produced for bi-decay data:

$$\tau_{ave} = (A_1\tau_1 + A_2\tau_2)/(A_1 + A_2).$$

(9)

The average lifetimes defined by IEM [42], RLD-2 [21], CMM [15], and CGRLD are:

$$\tau_{IEM} = \frac{h \cdot \sum_{j=0}^{M-1} (C_jN_j)}{N_0 - N_{M-1}},$$

$$N_j = \int_{jh}^{(j+1)h} f(t)dt = A_1\tau_1(x'_1 - x'^{j+1}_1) + A_2\tau_2(x'_2 - x'^{j+1}_2),$$

(10)

or $\tau_{IEM} \sim \frac{\int_0^{(M-1)h} f(t)dt}{f(0) - f([M-1]h)} = \left[\frac{A_1\tau_1(1-x'^{M-1}_1) + A_2\tau_2(1-x'^{M-1}_2)}{A_1 + A_2 - A_1x'^{M-1}_1 - A_2x'^{M-1}_2}\right]$ if $M \gg 1$ [42],

$$\tau_{RLD-2} = \frac{0.5T}{\ln \left(\frac{N_1}{N_2}\right)} = \frac{0.5T}{\ln \left(\int_0^{0.5T} f(t)dt\right)} = \frac{0.5T}{\ln \left(\frac{A_1\tau_1(1-e^{-0.5T/\tau_1}) + A_2\tau_2(1-e^{-0.5T/\tau_2})}{A_1\tau_1(e^{-0.5T/\tau_1} - e^{-0.5T/\tau_2}) + A_2\tau_2(e^{-0.5T/\tau_2} - e^{-T/\tau_2})}\right)},$$

(11)
\[\tau_{\text{CMM}} = T \cdot \Omega \left( \int_0^T t \cdot f(t) dt \right) = T \cdot \Omega \left( \int_0^T f(t) dt \right)\]

\[\tau_{\text{CGRLD}} = \frac{-h}{\ln(x)} \left( \sum_{i=0}^{R-S-1} x^i \right)^{\frac{1}{S}} = \frac{1}{N_S^T} \left( \frac{\int_0^T f(t) dt}{\sum_{i=0}^{R-S-1} x^i} \right)^{\frac{1}{S}} = \left[ \frac{A_1 \tau_{x_1} \left( x_1 - x_1^M \right) T + A_2 \tau_{x_2} \left( x_2 - x_2^M \right)}{A_1 \tau_{x_1} \left( x_1 - x_1^M \right) + A_2 \tau_{x_2} \left( x_2 - x_2^M \right)} \right]^{\frac{1}{S}},\]

where \(x_1 = \exp(-h/\tau_1), x_2 = \exp(-h/\tau_2), \) \(\Omega(\cdot)\) is the 1-D look up table described in [15], and \(h = T/M\) for IEM/CMM and \(h = T/(S + R)\) for CGRLD. From Equation (10), the average lifetime obtained by IEM will be almost the same with the commonly used Equation (9) when \(M \gg 1\). The conclusion will be the same for CGRLD for \(R \gg 1\). Figure 6 shows the average lifetimes versus \(A_1/(A_1 + A_2)\) obtained by Equations (9–13) at (a) \(\tau_1 = 0.33, \tau_2 = 3.3, T = 8\) ns and (b) \(\tau_1 = 2, \tau_2 = 4, T = 20\) ns. In Figure 6(b), IEM-7 and CGRLD (\(S/R = 0.2/3.2\)) have similar performances. In both cases, the RLD curves are farther away from Equation (9).

**Figure 6.** Average lifetime versus \(A_1/(A_1 + A_2)\) plots obtained by different methods for (a) \(\tau_1 = 0.33, \tau_2 = 3.3, T = 8\) ns and (b) \(\tau_1 = 2, \tau_2 = 4, T = 20\) ns.

To demonstrate how the above-mentioned single-exponential FLIM algorithms perform on bi-exponential data, we apply them to lifetime imaging data obtained by multi-photon FLIM. The experimental procedure is given in full elsewhere [55]. Briefly, early generation transplants of the P22 rat carcinosarcoma were grown in transparent window chambers, surgically implanted into the fascial layer of a dorsal skin flap of male BD9 rats [56,57]. For imaging of the blood vessel architecture, the molecular probe FITC conjugated Bovine serum albumin was injected intravenously as a blood pool contrast agent. After 100 minutes the FITC-BSA was observed to diffuse from the vasculature and the FLIM image taken. The intensity image in Figure 7(a) shows very little contrast between the vessels and the tumor bulk. Compared with the intensity image, the lifetime images Figure 7(b–f) obtained by 150-bin CMM, 51-bin IEM, Equation (9), RLD-2, and CGRLD show much better contrast.
The lifetimes in the vessels are smaller than those in the extra-vascular space. Figures 7(c,d) show very similar results except that there is a slightly larger variation in the IEM image shown in Figure 8(b). From Figures 6(a), 7(b) and 7(e), RLD and CMM generate larger lifetimes in the extra-vascular space due to, for example, the $\tau_2^2$ term in Equation (12). If $A_1$ is small throughout the image, the RLD and CMM produce less contrast than the other algorithms. They favor a larger $A_1$ distribution. The image data can also be analyzed with a bi-exponential fit algorithm throughout the image, and the lifetime components are $\tau_1 = 0.3 \pm 0.06$ ns and $\tau_2 = 3.4 \pm 0.4$ ns. The lifetime histograms for different algorithms are shown in Figure 8(a), and Table 1 shows the calculated lifetimes. Figure 8(b) shows that the lifetimes obtained by Equation (9) are highly correlated with those obtained by the 51-bin IEM. From Figure 8(b), the lifetime histogram showing a slightly larger variation in the IEM data reveals that the bi-exponential algorithms are crucial for examining most biological image data, but we are still benefitting from the high imaging speed and compactness single-exponential models can provide.

**Figure 7.** Images of FITC-albumin in a BD9 rat bearing a P22 tumour, 1 hour 40 minutes post administration. (a) Intensity image and lifetime images obtained by (b) 150-bin CMM; (c) 51-bin IEM; (d) Equation (9): $\tau_{\text{ave}} = (A_1 \tau_1 + A_2 \tau_2)/(A_1 + A_2)$; (e) 2-gate RLD; and (f) CGRLD with S/R = 0.2/3.2.
Table 1. Calculated lifetimes obtained by different algorithms.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Bi-Decay Model</th>
<th>Equation (9)</th>
<th>IEM M = 51</th>
<th>RLD</th>
<th>CMM</th>
<th>CGRLD S/R = 0.2/3.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau$ (ns)</td>
<td>$\tau_1 = 0.3 \pm 0.06$</td>
<td>$\tau_2 = 3.4 \pm 0.40$</td>
<td>$1.13 \pm 0.24$</td>
<td>$1.18 \pm 0.28$</td>
<td>$2.45 \pm 0.37$</td>
<td>$2.13 \pm 0.34$</td>
</tr>
</tbody>
</table>

Figure 8. (a) Lifetime histograms obtained by different FLIM algorithms and (b) $\tau_{\text{ave}}$ versus $\tau_{\text{IEM}}$ plot.

4. Conclusions

We have reviewed the latest developments on CMOS based FLIM systems and introduced several non-iterative high-speed time-domain algorithms suitable for massive sensor arrays. The theoretical error equations of the FLIM algorithms and their optimized operation conditions are discussed and compared. We have drawn from the theory several conclusions different from those in previously published literature and demonstrated the proposed gating schemes or FLIM algorithms on single-exponential data obtained from 0.13 $\mu$m CMOS SPAD arrays. The experimental results show good agreement with the theory. The contrast capabilities of the proposed algorithms on bi-exponential data are also discussed and demonstrated with an in vivo two photon fluorescence lifetime imaging data of FITC-albumin labeled vasculature of a P22 rat carcinosarcoma. With emerging CMOS imaging sensors, the proposed techniques capable of rapidly producing lifetime images with enough contrast show great potential in scientific and clinical applications.

Acknowledgments

The authors acknowledge funding support from the EPSRC (D. R. Matthews EP/C546105/1; V. Visitkul and D. Tyndall EPSRC Studentships), Royal Society Paul Instrument Fund (S. M. Ameer-Beg) and an endowment from Dimbleby Cancer Care (S. M. Ameer-Beg). We gratefully acknowledge Gillian Tozer and Iain Wilson for their support. The SPAD chip was supported by the European Community within the Sixth Framework Programme of the Information Science Technologies, Future and Emerging Technologies Open MEGAFRAME project (www.megaframe.eu).
References


Appendix

Similar to the analysis methods introduced in [45], we assume that the fluorescence decay function $f(t) = A \exp(-t/\tau)u(t)$, with $\tau$ being the lifetime and $A$ the prescaler, but here $h$ is the gate width of the "first" gate. Suppose $EN_c = \int_0^{T=R_h} A \exp(-t/\tau) dt = A \tau \left[ 1 - e^{-R_h/\tau} \right] = A \tau \left[ 1 - x^R \right]$, where $x = e^{-h/\tau}$, then the expectation values for $N_1$ and $N_2$ in both categories can be obtained as:

\[
\begin{align*}
EN_1 &= \int_0^h A \exp(-t/\tau) dt = A \tau \left( 1 - e^{-h/\tau} \right) = EN_c \left( 1 - x \right) / \left( 1 - x^R \right) \quad \text{and} \\
EN_2 &= \int_{0}^{R_h} A \exp(-t/\tau) dt = A \tau \left( e^{-R_h/\tau} - e^{-h/\tau} \right) = EN_c \left( x^S - x^R \right) / \left( 1 - x^R \right),
\end{align*}
\]

(A1)

We adopt the similar notation as in [45] for the recorded counts $N_j$, which are Poisson distributed with respective mean values $EN_j$ and standard deviations $\sigma N_j = \left( EN_j \right)^{1/2}$, $j = 1, 2$. From Equation (A1), $EN_1$ and $EN_2$ are related as $EN_c \left( x^S - x^R \right) - EN_2 \left( 1 - x \right) = 0$. We define:

\[
g \left( \hat{x} \right) = N_1 \left( \hat{x}^S - \hat{x}^R \right) - N_2 \left( 1 - \hat{x} \right).
\]

(A2)

Usually, the lifetime can be extracted from the recorded $N_1$ and $N_2$ in each pixel by solving $g \left( \hat{x} \right) = 0$ and $\hat{x} = -h/\ln \left( \hat{x} \right)$ and the prescaler $\hat{A} = A + \sigma A$ can be obtained by solving $\hat{A} = N_v / \left[ \hat{x} \left( 1 - e^{-h/\tau} \right) \right]$. There is often no experimental interest in the parameter $A$ [45], so we concentrate here on the derivation of $\sigma$. With the variations $\sigma N_j$ in the recorded counts $N_j$ ($j = 1, 2$), the calculated $\hat{x}$ will contain a variation $\sigma$ with respect to its mean value $x$ accordingly, $\hat{x} = x + \sigma_x$. To express $\sigma$ in terms of $\sigma N_j$ for the UGRLD we substitute $N_j = EN_j + \sigma N_j \left( j = 1, 2 \right)$ into Equation (A2) and $g(x + \sigma x) = 0$, and we can obtain:

\[
\begin{align*}
\left( EN_1 + \sigma N_1 \right) \left[ (x + \sigma_x)^S - (x + \sigma_x)^R \right] - \left( EN_2 + \sigma N_2 \right) \left( 1 - x - \sigma_x \right) &= 0, \\
\Rightarrow \sigma_x \left[ EN_2 + EN_1 \left( Sx^{S-1} - Rx^{R-1} \right) \right] &= \sigma N_1 \left( 1 - x \right) - \sigma N_2 \left( x^S - x^R \right) - \left( EN_2 + \sigma N_2 \right) \left( 1 - x - \sigma_x \right) = 0,
\end{align*}
\]

(A3)

where the higher order terms $\sigma_x^2$, $\sigma_x^3$, ..., are negligible, and the random variables $N_j$ and $N_2$ are independent for the UGRLD.

To derive the $F$-value for the CGRDL, we substitute $N_j = EN_j + \sigma N_j$ and $EN_2 = EN_2 + \sigma N_2$ into Equation (A2) with $EN_1 = EN_1 + \sigma N_1$ and $EN_2 = EN_2 + \sigma N_2$:

\[
\begin{align*}
\left( EN_1 + \sigma N_1 + \sigma N_2 \right) \left[ (x + \sigma_x)^S - (x + \sigma_x)^R \right] - \left( EN_2 + \sigma N_2 + \sigma N_2 \right) \left( 1 - x - \sigma_x \right) &= 0, \\
\Rightarrow \sigma_x \left[ EN_2 + EN_1 \left( Sx^{S-1} - Rx^{R-1} \right) \right] &= \sigma N_1 \left( 1 - x \right) - \sigma N_2 \left( x^S - x^R \right) - \left( EN_2 + \sigma N_2 + \sigma N_2 \right) \left( 1 - x - \sigma_x \right) = 0,
\end{align*}
\]

(A4)
where \( EN_{ov} = \int_{h}^{\tau} A \exp(-h/\tau)dt = EN_x\left(x^s - x\right)/(1-x^R) \) and the random variables \( N_1 \), \( N_{ov} \) and \( N_2 \) are mutually independent for the CGRLD. Differentiating \( x = e^{-h/\tau} \) and from Equations (A3) and (A4), we obtain:

\[
\begin{align*}
    \sigma_x &= \frac{\tau}{h} \sqrt{\left(1-x^s\right)^2 EN_x + \left(x^s - x^R\right)^2 EN_x + C \cdot EN_{ov}} \\
    C &= \begin{cases} 
        0, & \text{for UGRLD} \\
        -2(1-x)(x^s - x^R), & \text{for CGRLD}
    \end{cases}
\end{align*}
\]

\[
\begin{align*}
    \sigma_x \cdot \sqrt{EN_x} &= \tau \cdot \frac{\sqrt{1-x^R}}{h} \cdot \sqrt{\left(1-x\right)^2 \left(x^s - x^R\right)^2 + \left(x^s - x^R\right)^2 \left(1-x\right) + C \cdot (x^s - x)} \\
    \frac{\sigma_x}{\tau} \cdot \sqrt{EN_x} &\sim \frac{\sigma_x \cdot \sqrt{EN_x}}{\sigma_x} \left(1 + \frac{\sigma_x^2}{2EN_x}\right) \\
    \frac{\sigma_x^2}{\tau} &\geq \tau \cdot \frac{\sqrt{1-x^R}}{h} \cdot \frac{\sqrt{\left(1-x\right)^2 \left(x^s - x^R\right)^2 + \left(x^s - x^R\right)^2 \left(1-x\right) + C \cdot (x^s - x)}}{\left(x^s - x^R\right) + (1-x)(Sx^{R-1} - Rx^{R-1})}, \text{ when } \frac{\sigma_x}{2EN_x} \ll 1.
\end{align*}
\]