



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

A fast, flexible algorithm for calculating correlations in Fluorescence Correlation Spectroscopy

T. Laurence, S. Fore, T. Huser

October 19, 2005

Optics Letters

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

A fast, flexible algorithm for calculating correlations in Fluorescence Correlation Spectroscopy

Ted A. Laurence*, Samantha Fore, Thomas Huser

*Corresponding author: T. Laurence

T. Laurence, S. Fore and T. Huser are with Lawrence Livermore National Laboratory, Livermore, CA 94550 USA and the NSF Center for Biophotonics Science and Technology, University of California Davis, 4800 Second Avenue, Suite 2600, Sacramento, CA 95817 (phone: 925-422-1788; fax: 925-422-5137; e-mail: laurence2@llnl.gov).

Abstract

A new algorithm is introduced for computing correlations of photon arrival time data acquired in single-molecule fluorescence spectroscopy and fluorescence correlation spectroscopy (FCS). The correlation is first rewritten as a counting operation on photon pairs. For each photon, the contribution to the correlation function for each subsequent photon is calculated for arbitrary bin spacings of the correlation time lag. By retaining the bin positions in the photon sequence after each photon, the correlation can be performed efficiently. Example correlations for simulations of FCS experiments are shown, with comparable execution speed to the commonly used multiple-tau correlation technique. Also, wide bin spacings are possible that allow for real-time software calculation of correlations even for high count rates (~ 350 kHz). The flexibility and

broad applicability of the algorithm is demonstrated using results from single molecule photon antibunching experiments.

Copyright

OCIS codes

Introduction

Single-molecule fluorescence spectroscopy (SMFS) and fluorescence fluctuation spectroscopy (FFS) are powerful tools in the analysis of macromolecular dynamics and interactions. The many different SMFS and FFS methods introduced in recent years analyze the same basic data streams – sequences of photon arrival times from one or more detection channels. It has become standard practice¹⁻⁴ to first record the arrival times of all detected photons, with any accompanying information (lifetime, channel, etc.), and then perform any analysis on the resulting sequence.

One of the most common data analysis operations performed on these photon streams is the temporal correlation function used in fluorescence correlation spectroscopy (FCS)⁵. Calculated in the most straightforward manner, this can be an extraordinarily time-consuming process involving time lags that span several orders of magnitude on the time axis. However, with appropriate algorithms, this operation can be performed quickly. The multiple-tau correlation technique, originally implemented for digital hardware correlators, efficiently calculates temporal correlations in real time by limiting the number of correlation time lags calculated. Although only certain correlation time lags are calculated, the entire data set is used by successively reducing the time resolution of the data stream⁶. This algorithm has been

implemented in software, with appropriate modifications to account for the fact that, with sufficiently high time resolution, most of the recorded time bins contain no photon events⁷⁻⁹.

Here, we demonstrate a new correlation algorithm which is simpler in nature than the multiple-tau correlation technique. In contrast to the multiple-tau correlation technique, this algorithm allows for bins to be spaced in any way desired. The algorithm concept can be easily adapted to extensions of the temporal correlation function, including the PAID histogram³ and higher-order correlations^{10, 11}.

Theory

Correlation function rewritten in terms of photon timing data

If written in terms of two fluorescence or light scattering intensities, $I_A(t)$ and $I_B(t)$, the temporal correlation function is defined as $C_{AB}(\tau) = \langle I_A(t)I_B(t+\tau) \rangle / \langle I_A(t) \rangle \langle I_B(t+\tau) \rangle$. Here, time t and time lag τ are continuous variables.

The data recorded for a typical SMFS/FCS experiment are series of photon events i recorded with a discrete time stamp t_{Ai} with time resolution Δt for each detection channel A . Assuming stationarity, the ensemble averages in the expression for $C_{AB}(\tau)$ are converted to averages over all time. Averaging over a finite, experimental time T gives the correlogram $\hat{C}_{AB}(\tau)$. Within this finite time, there are N_A and N_B photons detected in the respective channels.

Using discrete time stamps simplifies the expression for the correlogram considerably. In terms of t , $I_A(t)$ is the number of photons i such that $t=t_{Ai}$; or $I_A(t) = n(\{i \ni t_{Ai} = t\}) / \Delta t$. $\{i \ni t_{Ai} = t\}$ is the set of all photons i such that $t_{Ai}=t$, and the operator n counts the number of

elements in the set. In this notation, using discrete time lag τ , we can express the correlogram as,

$$\hat{C}_{AB}(\tau) = \frac{n(\{i \ni t_{Ai} = t_{Bj} - \tau\})T/\Delta t}{n(\{i \ni t_{Ai} \leq T + \tau\})n(\{j \ni t_{Bj} \geq \tau\})(T/\Delta t)^2} \quad (1)$$

The restrictions on the average intensities used in the denominator are there for symmetric normalization¹².

Efficient correlation calculation by tracking bin positions

The above expression reduces the computation of the correlogram to a counting task. To compute $\hat{C}_{AB}(\tau)$ at the full resolution Δt , there are two obvious possibilities. First, one can iterate through each pair of photons, calculating the time lag $\tau = t_{Bj} - t_{Ai}$ and adding 1 to the correlogram $\hat{C}_{AB}(\tau)$. This is prohibitively slow, requiring $N_A N_B$ operations ($O(n^2)$ algorithm). An alternative is to use fast fourier transforms (FFT) to calculate the correlogram. This is a poor choice for FCS measurements, which require calculations of $\hat{C}_{AB}(\tau)$ over a very large dynamic range. A typical FCS experiment in our laboratory has a time resolution Δt of 12.5 ns (limited only by the time-resolution of the counter-timer data acquisition card; PCI-6602, National Instruments), and lasts for 5 minutes. We generally calculate time lags up to $\tau = 1$ s. To use the FFT at full resolution, this would require two arrays with 2.4×10^{10} elements each (most elements are zero).

The full time resolution is not needed or wanted in FCS measurements for the entire range of time lags, and a compromise is used to speed up the calculation. In the commonly used multiple-tau algorithm⁶, only the first 16 time lags are calculated at the full resolution. Subsequently, the

time resolution is halved by adding the number of photons counted in adjacent bins, and 8 more time lags are calculated. This process is repeated until the full desired dynamic range is reached. Originally developed for hardware correlators, this algorithm has recently been extended to work with the photon timing data described above⁷⁻⁹.

As we now describe, it is not necessary to use the rigid bin spacings of the multiple-tau algorithm or deal with the complications of halving the time resolution in order to calculate the correlogram. We specify M bins with desired time lag τ_k : say $[\tau_1^{\min}, \tau_1^{\max}]$, $[\tau_2^{\min}, \tau_2^{\max}]$, ..., $[\tau_M^{\min}, \tau_M^{\max}]$. There is no requirement on the spacing between these bins. The choice depends only on the time scales relevant to the experiment.

The algorithm is implemented as follows:

- 1) Initialize a correlogram Y_k with M bins to 0.
- 2) Starting with t_{A1} , find $t_{B_k^{\min}}$ such that $t_{B_k^{\min}-1} < t_{A1} + \tau_k^{\min} \leq t_{B_k^{\min}}$ and $t_{B_k^{\max}}$ such that $t_{B_k^{\max}-1} < t_{A1} + \tau_k^{\max} \leq t_{B_k^{\max}}$ for all $k = 1 \dots M$ (a linear search is adequate).
- 3) For every k , add $j_k^{\max} - j_k^{\min}$ to Y_k .
- 4) Repeat steps 2 and 3 for $t_{A2} \dots t_{AN_A}$. The important “trick” lies in this step. The bin limits j_k^{\min} and j_k^{\max} are only slightly adjusted when going from t_{A1} to t_{A2} in step 2 (figure 1). Hence, it is possible to use the previously found j_k^{\min} and j_k^{\max} as starting points to find new values that satisfy the conditions in step 2 for t_{A2} . This reduces the computational load dramatically.
- 5) Normalize according to equation (1).

The correlogram is initialized in step 1. In step 2, the boundaries of the M time lag bins in channel B are mapped out for the first photon in channel A, allowing for the contribution of photon pairs including the first photon to be simply calculated in step 3. Steps 2 and 3 are then repeated for the remaining photons in channel A. However, it is very important to notice that the boundaries of the M time lag bins are in general shifted only by a small amount when going from the first to second photons in channel A (figure 1). If channels A and B have the same count rates, then each time lag bin boundary in channel B will on average shift only one photon for every photon shifted in channel A. The number of operations scale as $N_A M$ rather than $N_A N_B$ operations, dramatically reducing the computational cost. The computational cost is similar to the multiple-tau algorithm, except now without restrictions on the time lag bins chosen.

Results and Discussion

Method compares favorably with multiple-tau algorithm

We first compare the performance of our new algorithm with the multiple-tau algorithm using simulated FCS data. A homogeneous fluorescent species is simulated to undergo three-dimensional Brownian diffusion through a confocal detection volume³. These are 10 simulations, where each molecule's average count rate (brightness) in the confocal detection volume is 35.4 kHz and the diffusion time of the molecule through this detection volume is 300 μ s. The average number of molecules is set at 0.1, 1.0, and 10.0 in three separate simulations, leading to total average count rates of 3.54 kHz, 35.4 kHz, and 354 kHz, respectively. A plot of calculated correlations for the simulation with an average of 0.1 molecules per detection volume and corresponding curve fits are shown in figure 2.

Table 1 compares the calculation times using our new algorithm and the multiple tau algorithm (performed on a 2 GHz Pentium M processor). In the second and third columns, the bin spacings are those dictated by the multiple tau algorithm, with time lags between 1 μ s and 2 s. At low count rates, there is very little difference in performance. However, at high count rates (354 kHz), the multiple tau algorithm is about 40% faster. When the reciprocal of the count rate is comparable to the minimum time bin (1 μ s), the successive reductions in the time resolution produces savings in the computation time.

Number of Photons (10 s simulations)	Time for Multiple Tau (s)	Time for current algorithm (s)	Time for large bin spacing (s)
35291	0.21	0.21	0.07
362483	1.9	2.1	0.44
3559930	16	23	4.0

For the fourth column, a wide spacing between bins is used (2 per decade between 1 μ s and 1s) using our new algorithm. Such a wide spacing may be used for calculating correlations in real time during the experiment. For example, for the highest count rate in the bottom row, neither algorithm can compute the full correlation within the simulation time of 10s. However, with wider bin spacings (fourth column), our software correlation routine would still be able to keep up with a count rate of 354 kHz. Widely spaced bins are problematic for the multiple-tau algorithm due to triangular averaging^{8, 9, 13}. Triangular averaging results from the finite size of the time bins used during the calculation of the correlation. The process used by the multiple-tau algorithm of adding the number of photons counted in adjacent bins to reduce the time resolution leads to larger triangular averaging. With the standard bin spacings used, the triangular averaging is small (<0.1%)^{8, 9, 13}. However, for the wider spacing of 2 bins per decade in column

4, the triangular averaging would be much more severe. In our new algorithm, the full time resolution of the photon arrival times is preserved through the entire calculation. The only triangular averaging is on the time scale of the photon arrival time (12.5 ns in our case).

Example correlations – photon antibunching

The current algorithm is most useful in cases where specific bin spacings are needed. For example, we are currently performing measurements that attempt to quantify the number of independent fluorescent quantum emitters using fluorescence antibunching; the experiments and experimental setup are described in reference¹⁴.

Briefly, we measure fluorescence from single oligonucleotides labeled with a red fluorophore (Atto 655, Atto-tec GmbH) immobilized on a coverslip. We collect data from single hairpin samples by scanning images (figure 3a) using a 2D piezo-scanner (MadCity Labs, Nano-Bio200, Madison, WI). For excitation of Atto 655, we use a picosecond pulsed laser (LDH 635-B, PicoQuant GmbH) driven by a 5 MHz external clock source generated by a counter-timer board (PCI-6602, National Instruments). The excitation laser is focused by a high numerical aperture microscope objective, and emitted fluorescence is collected by the same objective. The emission is then focused onto a pinhole (150 μm), and a 50/50 beamsplitter is used to form a Hanbury-Brown and Twiss interferometer with two avalanche photodiode detectors (SPCM-AQR-14, PerkinElmer). It is necessary to use two independent detectors for photon antibunching, because there is no ideal detector with negligible dead time^{15, 16}.

The APDs produce a TTL pulse for every photon detected which are timed using an 80 MHz clock. The same 80 MHz clock times both channels, and drives the 5 MHz clock for triggering the laser diode. Hence, the laser and photon timing are synchronized, facilitating later data analysis.

The random arrival times of photon events can be overcome by using pulsed excitation sources^{14, 17, 18}. The most efficient way to then obtain photon antibunching histograms is to calculate the cross-correlation function between the two channels of the Hanbury-Brown and Twiss interferometer. Comparing the number of correlated to anti-correlated events by measuring peak heights in the arrival time histogram can then, for example, be used to reveal the average number of quantum emitters in the observation volume^{14, 18, 19}. Figure 3b shows the correlation of all photon events obtained during the image scan (figure 3a) calculated with logarithmic spacings similar to those used for the FCS simulation in figure 2 (the spacing between bins is too large to observe the pulsed nature of the excitation). Note, that there are two significant time scales in the decay curve. The faster fluctuations at the $\sim 100 \mu\text{s}$ time scale as indicated by the slight shoulder in the plot are due to triplet-state-induced blinking of the fluorophores²⁰. The major, longer term decay at the $\sim 30 \text{ ms}$ time scale is due to the movement of the scanner²¹. Figure 3c shows the fine detail of the correlation function around two time points in the FCS curve in figure 3b (as indicated by the dash-dotted and full line in the figure). Comparison between the short time scale correlation and the long time scale correlations allows the quantification of the triplet state effects on the antibunching curve. Such effects will play a critical role in using photon antibunching to count the number of fluorophores for each individual molecule. The ability to make such comparisons is made possible using the flexible bin spacings.

Conclusion

We have presented a novel correlation algorithm that makes use of flexible bin spacings to quickly or accurately calculate correlation functions depending on the needs of the experiment. We have demonstrated the performance of this algorithm by calculating the correlation function

for simulated fluorescence correlation spectroscopy data. We have also demonstrated the large dynamic range over which this algorithm permits calculations by simultaneously calculating correlations on the nanosecond to millisecond time scales for a photon antibunching experiment. The correlation algorithm developed here allows researchers to adjust time lag bins to the resolution that is important to them without sacrificing performance. Also, this algorithm can be extended to other calculations, including the PAID histogram³ and higher-order correlations^{10,11}.

Acknowledgments

S. Fore acknowledges funding from the student employee graduate fellowship program at Lawrence Livermore National Laboratory. The Center for Biophotonics, an NSF Science and Technology Center, is managed by the University of California, Davis, under Cooperative Agreement No. PHY 0120999. Work at LLNL was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory, under contract No. W-7405-Eng-48.

References

1. J. S. Eid, J. D. Muller, and E. Gratton, "Data acquisition card for fluctuation correlation spectroscopy allowing full access to the detected photon sequence," *Rev Sci Instr* **V71**, 361-368 (2000).
2. J. R. Fries, L. Brand, C. Eggeling, M. Kollner, and C. A. M. Seidel, "Quantitative identification of different single molecules by selective time-resolved confocal fluorescence spectroscopy," *J Phys Chem a* **102**, 6601-6613 (1998).
3. T. A. Laurence, A. N. Kapanidis, X. X. Kong, D. S. Chemla, and S. Weiss, "Photon arrival-time interval distribution (PAID): A novel tool for analyzing molecular interactions," *J Phys Chem B* **108**, 3051-3067 (2004).
4. S. A. Soper, L. M. Davis, and E. B. Shera, "Detection and identification of single molecules in solution," *J. Opt. Soc. Am. B, Opt. Phys. (USA)* **9**, 1761-1769 (1992).
5. D. Magde, E. Elson, and W. W. Webb, "Thermodynamic fluctuations in a reacting system: measurement by fluorescence correlation spectroscopy," *Phys Rev Lett* **29**, 705-708 (1972).
6. K. Schatzel, "New concepts in correlator design," *Inst. Phys. Conf. Ser. No. 77: session 4 No. 77: session 4*, 175-185 (1985).

7. T. Wohland, R. Rigler, and H. Vogel, "The standard deviation in fluorescence correlation spectroscopy," *Biophysical Journal* **80**, 2987-2999 (2001).
8. D. Magatti, and F. Ferri, "Fast multi-tau real-time software correlator for dynamic light scattering," *Applied Optics* **40**, 4011-4021 (2001).
9. D. Magatti, and F. Ferri, "25 ns software correlator for photon and fluorescence correlation spectroscopy," *Rev Sci Instr* **74**, 1135-1144 (2003).
10. L. Edman, and R. Rigler, "Memory landscapes of single-enzyme molecules," *Proc Natl Acad Sci U S A* **97**, 8266-8271 (2000).
11. A. G. Palmer, and N. L. Thompson, "Molecular Aggregation Characterized by High-Order Autocorrelation in Fluorescence Correlation Spectroscopy," *Biophysical Journal* **52**, 257-270 (1987).
12. K. Schatzel, M. Drewel, and S. Stimac, "Photon correlation measurements at large lag times: improving statistical accuracy," *J. Mod. Opt. (UK)* **35**, 711-718 (1988).
13. K. a. R. P. Schatzel, "Noise on Multiple-Tau Photon Correlation Data," *SPIE vol. 1430 Photon Correlation Spectroscopy: Multicomponent Systems* **1430**, 109-115 (1991).
14. S. Fore, T. A. Laurence, Y. Yeh, R. Balhorn, C. W. Hollars, M. Cosman, and T. Huser, "Distribution analysis of the photon correlation spectroscopy of discrete numbers of dye-molecules conjugated to DNA," *IEEE Journal of Selected Topics in Quantum Electronics* **in press** (2005).
15. W. P. Ambrose, P. M. Goodwin, J. Enderlein, D. J. Semin, J. C. Martin, and R. A. Keller, "Fluorescence photon antibunching from single molecules on a surface," *Chem Phys Lett* **269**, 365-370 (1997).
16. T. Basche, W. E. Moerner, M. Orrit, and H. Talon, "Photon Antibunching in the Fluorescence of a Single Dye Molecule Trapped in a Solid," *Phys Rev Lett* **69**, 1516-1519 (1992).
17. B. Lounis, and W. E. Moerner, "Single photons on demand from a single molecule at room temperature," *Nature* **407**, 491-493 (2000).
18. K. D. Weston, M. Dyck, P. Tinnefeld, C. Muller, D. P. Herten, and M. Sauer, "Measuring the number of independent emitters in single-molecule fluorescence images and trajectories using coincident photons," *Analytical Chemistry* **74**, 5342-5349 (2002).
19. C. W. Hollars, S. M. Lane, and T. Huser, "Controlled non-classical photon emission from single conjugated polymer molecules," *Chem Phys Lett* **370**, 393-398 (2003).
20. J. Widengren, U. Mets, and R. Rigler, "Fluorescence Correlation Spectroscopy of Triplet States in Solution - a Theoretical and Experimental Study," *J Phys Chem* **99**, 13368-13379 (1995).
21. Y. Xiao, V. Buschmann, and K. D. Weston, "Scanning fluorescence correlation spectroscopy: A tool for probing microsecond dynamics of surface-bound fluorescent species," *Analytical Chemistry* **77**, 36-46 (2005).

Figure Captions

Figure 1: Calculating the cross-correlation of photon sequences in two channels (A and B) using our new algorithm. The maximum and minimum limits for M bins in time lag τ

$([\tau_1^{\min}, \tau_1^{\max}], [\tau_2^{\min}, \tau_2^{\max}], \dots, [\tau_M^{\min}, \tau_M^{\max}])$ are added to the arrival time of each photon in channel

A. Part **(a)** shows these limits (black arrow for τ_1^{\min} , gray dotted lines for other limits) for

photon 1 from channel A, and part **(b)** shows these limits for photon 2. In this example, $M=4$,

$\tau_1^{\min} = 0$, $\tau_2^{\min} = \tau_1^{\max}$, $\tau_3^{\min} = \tau_2^{\max}$, and $\tau_4^{\min} = \tau_3^{\max}$. To the left of each limit in (a), the index j of

the photon in channel B such that $t_{Bj} > t_{A1} + \tau_{\text{curr}}$, where, in turn, $\tau_{\text{curr}} = \tau_1^{\min}$, $\tau_{\text{curr}} = \tau_1^{\max}$, etc.

Similar indices j are shown in (b). The contribution to the correlogram Y_k for each time lag bin

k is calculated by subtracting the indices maximum and minimum indices, $j_k^{\max} - j_k^{\min}$. The

critical step in making this correlation calculation efficient is when switching between photons in

channel A. Note that in going from photon 1 in channel A to photon 2 (comparing (a) and (b)),

there are only small adjustments made in the values of j_k^{\min} and j_k^{\max} . Rather than searching

through the entire photon stream in channel B to find the new time lag bin limits for photon 2,

we keep all of the values found for photon 1, and make minor adjustments in their values so that

the relationships $t_{Bj} > t_{A2} + \tau_{\text{curr}}$ are satisfied for $\tau_{\text{curr}} = \tau_1^{\min}$, $\tau_{\text{curr}} = \tau_1^{\max}$, etc. This reduces the

computational load dramatically, required on order $N_A M$ operations.

Figure 2: Correlations and fits for simulated data of fluorescent molecules diffusing through a

Gaussian detection volume. The average number of molecules N in the detection volume is 0.1,

and the diffusion time τ_D through the detection volume is 300 μs . For the simulation detection

volume, the correlation curve satisfies, $C(\tau) = 1 + 1/\left[N(1 + \tau/\tau_D)\sqrt{1 + \tau/(25\tau_D)} \right]$. The

correlation calculated using our new algorithm with the quasi-logarithmic bin spacings from the

multiple-tau correlation algorithm (8 bins per octave, 160 total bins) is shown in black. The

dotted light gray line is a fit with $N = 0.099 \pm 0.002$ and $294 \pm 3 \mu\text{s}$. The dark gray line is the

correlation with wide bin spacings (2 bins per decade, 12 total bins). A fit (not shown) recovered the values $N = 0.10 \pm 0.01$ and $300 \pm 100 \mu\text{s}$. The wider spacing sacrifices some accuracy, but dramatically increases speed of computation.

Figure 3: Photon antibunching experiments are performed on fluorescently labeled DNA oligomers attached to a glass surface. **(a)** Image of surface-immobilized DNA oligomers. Image size $20 \mu\text{m} \times 20 \mu\text{m}$ **(b)** Long-time scale correlation function over the entire image (similar bin spacings to figure 2). **(c)** Short-time scale correlation functions with linearly spaced bins (25 ns bins). The non-normalized correlation for the region with time lags between $-1.25 \mu\text{s}$ and $1.25 \mu\text{s}$ is shown (central plot in figure 3c, dotted line in figure 3b), along with the regions within $1.25 \mu\text{s}$ of the $\pm 2 \text{ ms}$ time lags (side plot in figure 3c, black line in figure 3b). The spikes in figure 3c correspond to the timing of the pulses from the excitation laser. The solid gray line corresponds to the average peak height in the region between $-1.25 \mu\text{s}$ and $1.25 \mu\text{s}$, and the dotted gray line corresponds to the average spike height in the regions within $1.25 \mu\text{s}$ of $\pm 2 \text{ ms}$ time lag regions. The peak at $-0.2 \mu\text{s}$ with very low photon pair counts is due to photon antibunching (black arrow). It is not at 0 time lag due to a cable delay and a software adjusted digital delay.

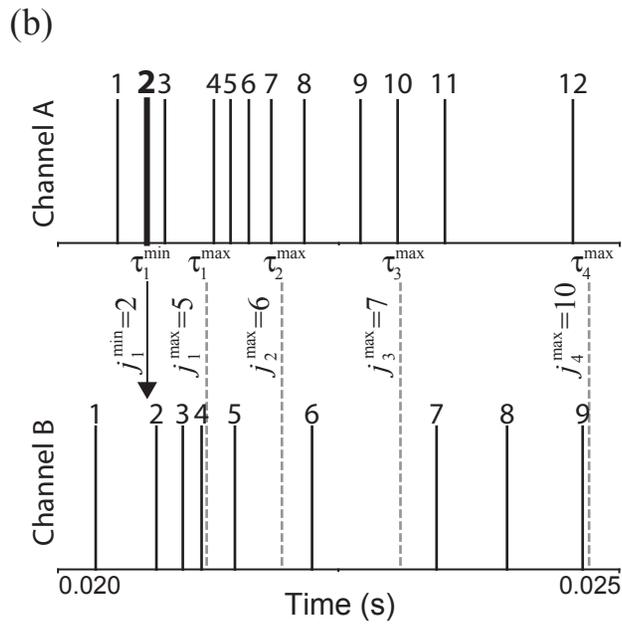
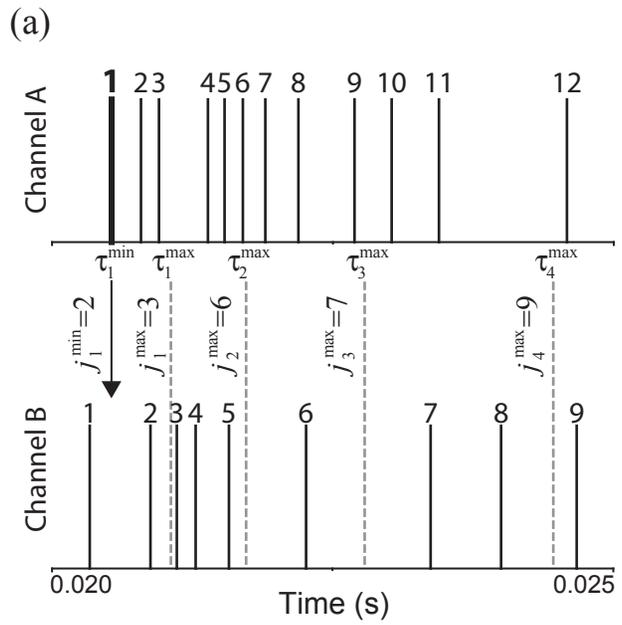


Figure 1, Laurence *et al.*

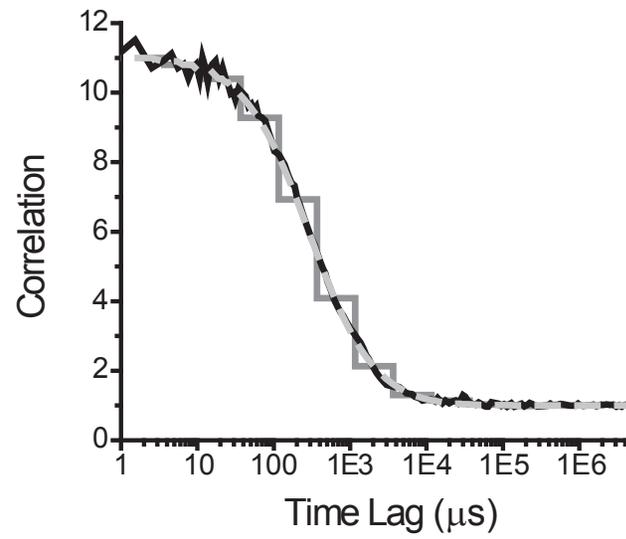


Figure 2, Laurence *et al.*

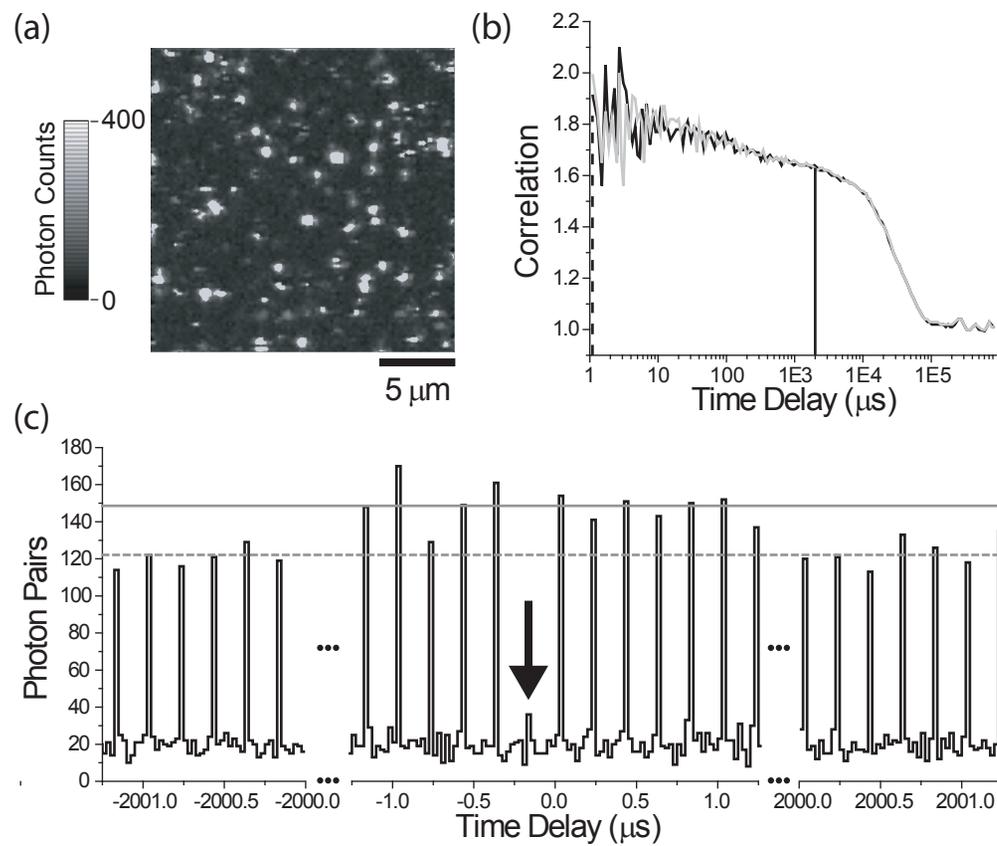


Figure 3, Laurence *et al.*