

Published in final edited form as:

Comput Phys Commun. 2013 April 1; 184(4): 1322–1332. doi:10.1016/j.cpc.2012.12.022.

Vectorized data acquisition and fast triple-correlation integrals for Fluorescence Triple Correlation Spectroscopy

William K Ridgeway^{a,b,c}, David P Millar^a, and James R Williamson^{a,b,c,*}

^aDept. of Molecular Biology, The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla CA 92037, USA

^bDept. of Chemistry, The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla CA 92037, USA

^cThe Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla CA 92037, USA

Abstract

Fluorescence Correlation Spectroscopy (FCS) is widely used to quantitate reaction rates and concentrations of molecules *in vitro* and *in vivo*. We recently reported Fluorescence Triple Correlation Spectroscopy (F3CS), which correlates three signals together instead of two. F3CS can analyze the stoichiometries of complex mixtures and detect irreversible processes by identifying time-reversal asymmetries. Here we report the computational developments that were required for the realization of F3CS and present the results as the Triple Correlation Toolbox suite of programs. Triple Correlation Toolbox is a complete data analysis pipeline capable of acquiring, correlating and fitting large data sets. Each segment of the pipeline handles error estimates for accurate error-weighted global fitting. Data acquisition was accelerated with a combination of off-the-shelf counter-timer chips and vectorized operations on 128-bit registers. This allows desktop computers with inexpensive data acquisition cards to acquire hours of multiple-channel data with sub-microsecond time resolution. Off-line correlation integrals were implemented as a two delay time multiple-tau scheme that scales efficiently with multiple processors and provides an unprecedented view of linked dynamics. Global fitting routines are provided to fit FCS and F3CS data to models containing up to ten species. Triple Correlation Toolbox is a complete package that enables F3CS to be performed on existing microscopes.

Keywords

Fluorescence Triple Correlation Spectroscopy; F3CS; FCS; Multiple-tau; Fluorescence; Data Acquisition

1. Overview

Fluorescence Correlation Spectroscopy (FCS) analyzes chemical kinetics by observing fluctuations that spontaneously occur in a small sample at equilibrium [1]. FCS is a powerful complement to single-molecule experiments, tolerating less-than ideal samples and detecting

© 2012 Elsevier B.V. All rights reserved

*Corresponding author. jrwill@scripps.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

dynamics on multiple time scales. FCS has diverse applications in biophysics, molecular biology and live cell imaging [2], and FCS theory is being actively developed to create new spectroscopic methods [3].

FCS uses raw fluorescence intensity data $i(t)$ to calculate an experimental correlation function $G(\tau)$ by asking whether fluctuations at time t persist at time $t + \tau$:

$$\begin{aligned} G(\tau) &= \frac{\langle \delta i(t) \delta i(t+\tau) \rangle}{\langle i(t) \rangle \langle i(t) \rangle} \\ \langle i(t) \rangle &= \frac{1}{T} \sum_{d=0}^{T-1} i(t_d) \quad (1) \\ \delta i(t) &= i(t) - \langle i(t) \rangle \end{aligned}$$

The number of samples T can exceed 10^{10} , providing excellent statistics. Hardware-based [4, 5] and software-based [6, 7, 8, 9] correlation systems routinely calculate correlations using a logarithmic series of τ values that span more than five orders of magnitude. This gives FCS a dynamic range few other techniques enjoy, but precludes the use of Fourier-based numerical techniques for calculating the integral. The amplitudes of FCS curves contain a wealth of information about particle concentrations and photon yields, and if multiple molecules are labelled with distinct fluorophores detected in different channels $i_\alpha(t)$ and $i_\beta(t)$, the Cross-Correlation integral $G_{\alpha \times \beta}(\tau)$ [10] reveals the concentrations of particles with both labels:

$$G_{\alpha \times \beta}(\tau) = \frac{\langle \delta i_\alpha(t) \delta i_\beta(t+\tau) \rangle}{\langle i_\alpha(t) \rangle \langle i_\beta(t) \rangle} \quad (2)$$

Cross-correlation identifies linked binding and is particularly valuable for analyzing biological interactions in large molecules that are not amenable to FRET. Fluorescence Triple Correlation Spectroscopy (F3CS) was recently developed by extending the cross-correlation principle to three signals [11, 12] using a more complicated triple-correlation integral,

$$G_{\alpha \times \beta \times \gamma}(\tau_1, \tau_2) = \frac{\langle \delta i_\alpha(t) \delta i_\beta(t+\tau_1) \delta i_\gamma(t+\tau_2) \rangle}{\langle i_\alpha(t) \rangle \langle i_\beta(t) \rangle \langle i_\gamma(t) \rangle} \quad (3)$$

The third signal allows F3CS to resolve binary and ternary particles that are indistinguishable to Cross-Correlation FCS and the second delay time provides a means to analyze linked kinetics and detect time-reversal asymmetries indicative of irreversible processes.

F3CS requires fast triple correlation integrals. Rather than designing custom hardware to do this, we took a software approach that minimized adoption costs for us and for others. This approach required a data acquisition program capable of recording multiple channels at MHz sampling rates, fast multithreaded routines to calculate correlation integrals and a means to fit large amounts of data in a global manner to resolve populations of molecules that differ by composition and / or diffusion time. Triple Correlation Toolbox addresses these specific needs, enables F3CS to be performed on existing microscopes, and allows biochemists and physical chemists to quickly collect and analyze data.

1.1. Triple Correlation Toolbox is a suite of programs for collecting, integrating and fitting data

An outline of the triple-correlation experiment and data analysis pipeline is shown in Figure 1 and the major programs are detailed in Table 1.1.

2. Data acquisition accomplished by mixing TTL circuitry with 128-bit SSE operations

F3CS requires the ability to record photon arrival times with $\sim 1\mu\text{s}$ resolution across three independent channels, and we sought to perform this task using a commonly available multifunction data acquisition card (PCIe-6251, National Instruments) such as those found in other laboratory-based high-performance data acquisition schemes [8, 13]. However, established photon detection schemes were limited to $\sim 200\text{ kHz}$ sample rates on this hardware because they made poor use of the limited bandwidth across the card, which forced data on the card to be transferred to RAM at a rate faster than the CPU could handle. One scheme for example, used 32-bit counters [6] to count photon arrivals during a short time period. This required 4 bytes per channel per timepoint, even though most counts were 0 or 1. Another scheme recorded only the time between successive photons [7], which worked well for weak signals but was rapidly overwhelmed by concentrated samples and high photon yields. To circumvent these problems, we designed a digital waveform acquisition scheme that only used 1 bit per channel per timepoint and took advantage of the deep hardware buffers commonly associated with digital I/O. The scheme can record 8 independent channels at 20 MHz on our hardware.

The waveform acquisition scheme used a fast but primitive memory chip to record random photon arrivals, and the state of this memory was read by the computer at regular intervals (Fig. 2A). The scheme starts when a photon is detected by any one of three Avalanche Photo Diode (APD) photon detectors. The APD temporarily ignores further photons and emits a brief TTL electrical pulse (Timing diagram, Fig. 2B). The TTL pulse is fed into the clock pin of the primitive memory, an off-the-shelf counter-timer chip (CD74AC163E, 100 MHz clock rate, Texas Instruments.), and the pulse toggles the level of the lowest counter output from low to high or from high to low. The output remains at this new level until the next photon is read. The waveform detection circuit reads the output as 0 or 1 and then moves these single-bit data to RAM. Photon arrivals can then be inferred by finding instances when data at time i differ from data at time $i + 1$. The scheme requires that the length of the waveform time bin be equal to or shorter than the dead time of the APDs, otherwise two photons per waveform time bin would appear to be zero photons. Here, waveform bins were 50 ns wide and the APD dead time was 50–60 ns.

Sequential datapoints were compared using single bit-wise XOR operations performed using Streaming Single Instruction Multiple Data Extensions (SSE) [14, 15] to parallelize this operation across 128-bit registers such that the CPU could keep pace with the data acquisition process. The resulting photon records were binned to achieve the desired time resolution (typically 800 ns) and the final values for the three channels of interest were written to disk. In short, high-speed data acquisition required the speed of dedicated hardware, the speed of parallel SSE operations, and the efficient representation of photon data during the crucial card-to-RAM transfer stage.

The TTL / SSE detection scheme did not introduce electronic artifacts into the data, as determined by cross-correlating photon arrivals from an incoherent light source (room lights). While crosstalk would appear as non-zero amplitude, correlations were zero within error (Fig. 3) demonstrating that the acquisition technique can be used for its intended purpose. Suppression of crosstalk was achieved by careful signal path routing, correct

termination of APD pulses, independent digital ground lines on all channels, bypassing the +5 V rail with fast capacitors (Polypropylene 100 V 220 pF 10%, Wima), and grounding all unused channels of the waveform circuit. The $G(0, 0)$ term (not shown in the log-log-linear plot in Fig. 3A) did have a finite amplitude, but this is of limited consequence as $G(0, 0)$ is not interpreted anyway due to shot noise [16, 17]. For triple auto-correlations, the terms $G(0, 0)$, $G(0, \tau_2)$ and $G(\tau_1, 0)$ similarly contain contributions from the statistics of photon detection and detector afterpulsing that preclude their analysis [18, 11].

3. Data acquisition software

Data acquisition and primary data inspection tasks were split amongst four programs (Fig. 4A). Microscope alignment and tuning requires real-time readout using the TTL / SSE scheme, and was accomplished with the GUI-based program F3CS_AlignTool. F3CS_DAQ is a command-line program that records photon arrivals to disk, in a binary format. If required, the data are reversed in time by F3CS_Reverser or resampled and converted to more user-friendly text files using F3CS_TimeTrace.

3.1. F3CS_DAQ provides robust data acquisition with laser control

The main data acquisition program, F3CS_DAQ, controls laser excitation, streams large hundreds of GB of data to the disk during long (10 hr) experiments, and minimizes two sources of data disruption: optical contamination from debris and card mis-reads described above. Debris were responsible for ~ 90% of data disruption during typical usage. Overruns were minimized by implementing data transfer and file I/O processes in separate threads, using circular buffers to minimize the effects of slow file I/O, and using vectorized code as described above. Code was originally hand-tuned in assembly for 2007-era Athlon 64 processors ($\times 2$ 4200+ 2.2 GHz), and the hand-tuned code runs on later AMD processors, but not Intel. More recent versions of gcc (>4.4.3) with auto-vectorization routines [19] now obviate the need for hand-coding by vectorizing the key loops that act on `data[i]` and `data[i + 1]` simultaneously. The gcc auto-vectorized code is only two- to three-fold slower than hand-coded assembly. To provide portability to Intel CPUs, both assembly and vectorizable C code are included, and both are sufficiently fast that overruns should not adversely affect data quality, especially on computers more modern than ours.

Data become contaminated when debris (detergent aggregates, undissolved Trolox, protein aggregates) is laser-trapped in the focal volume of the microscope. Trapping is exacerbated with the high laser power required for multi-photon excitation and elicits fluorescence spikes that can harm the APDs. To break the trap, F3CS_DAQ shuts the laser for one second when a user-defined count-rate is exceeded (typically $2\times$ the present count rate, higher with dilute concentrations) and the shutter repeatedly closes for five seconds until fluorescence falls below this limit. To clear non-fluorescent debris, which suppresses fluorescence and is more difficult to detect, shutters close for five seconds after every five hundred seconds of data acquisition.

3.2. F3CS_Reverser reverses the time order of data in raw photon count files

As discussed below, correlation integrals can be calculated from reversed and non-reversed photon count data. F3CS_Reverser reverses raw photon count data from .w3c files and outputs the reversed data as .r.w3c files that are read by correlation integral software. When F3CS_Reverser works with a series of consecutively-recorded .w3c files, data from the last .w3c file end up in the first .r.w3c file.

3.3. F3CS_TimeTrace converts raw photon count files into text files with a given time resolution

This is useful for visual inspection of the data (Fig. 4B) and basic photometric tasks, e.g. measuring FRET in bulk solution. The output is not used by the analysis pipeline.

3.4. F3CS_AlignTool: A graphical user interface for real-time display of microscope data

F3CS_AlignTool assists with microscope tuning and sample verification prior to data acquisition (Fig. 5). There are three modes: Alignment for displaying intensity traces and running averages for three channels, Burst for very rapid sampling of time-traces, and FCS for real-time double-correlation FCS analysis of the sample.

The first mode, Alignment, plots fluorescence from three channels over time and displays the photon count rate in MHz. Channels α , β , γ are blue or green, yellow and red respectively. This example shows the final alignment of the yellow channel. The first 50 reads (~ 10 s, adjustable) are averaged for the top intensity figure, while the bottom intensity figure is instantaneous.

The Bursts mode displays 50 ns – 512 μ s time-resolution data. Time is plotted along the horizontal axis, and intensity is inverted on the vertical axis, with longer coloured bars indicating more photons. The vertical scale can be adjusted on the right, in units of photons. The time scale is such that each pixel represents the time resolution selected on the far right; this example employs 512 μ s bins to watch bursts when individual fluorescently-labeled RNA molecules diffuse through the focal volume.

The FCS mode provides a rapid but imprecise FCS measurement. Data are the rolling average of 6×1 second observations, used for determining whether the system is setup correctly before data are collected using F3CS_DAQ. As correlation and data acquisition are both computationally intensive they are performed alternately, and so it takes approximately 15 s to display the first curve and curves are updated every 3 s. As acquisition times are relatively short, data are relatively imprecise. For tuning the setup, laser power in particular, a few qualitative estimates of data are provided, but are not rigorous and should not be used to analyze data. For example, N reflects the approximate number of molecules per focal volume from small- τ data which are frequently contaminated by afterpulsing, so the N estimate is often inaccurate for autocorrelation data. η reflects the number of photons per molecule as a benchmark of system performance (units of kHz/molecule), and S/N estimates signal to noise by calculating the ratio of $G(0)$ amplitude to the standard deviation of $G(\tau)$ values over each of the three autocorrelations. Three correlation modes select which channels to correlate and display. Mode 0,1 displays $G_{\alpha \times \alpha}(\tau)$, $G_{\alpha \times \beta}(|\tau|)$, $G_{\beta \times \beta}(\tau)$ as green yellow and red respectively. Mode 1,2 displays $G_{\beta \times \beta}(\tau)$, $G_{\beta \times \gamma}(|\tau|)$ and $G_{\gamma \times \gamma}(\tau)$, Mode 2,0 displays $G_{\gamma \times \gamma}(\tau)$, $G_{\gamma \times \alpha}(|\tau|)$, and $G_{\alpha \times \alpha}(\tau)$.

4. Correlation integral routines

Data recorded by F3CS_DAQ can be correlated in a variety of ways (Fig. 6A, Table 2). Depending on the desired sign of the delay times, raw photon-count data from F3CS_DAQ are either correlated directly after acquisition or following a time-reversal step. The resulting correlation data are then read by F3CS_Outlier2 and F3CS_Outlier3 in order to eliminate outliers, average data and estimate experimental errors.

4.1. F3CS_2FCS calculates double correlation (FCS) integrals

Six double-correlation integrals are calculated using a software version of the symmetric multiple-tau scheme [4, 5] that was originally designed to be performed quickly in hardware using small integers:

$$\begin{aligned}
 G_{\alpha \times \beta}^{\text{sm}\tau}(\tau) &= \frac{A_2}{2^j} \sum_{h=0}^{T/2^j-1} \left[\left(\sum_{k=0}^{2^j-1} \delta i_{\alpha}((2^j h + k)t_Q; 0) \right) \left(\sum_{k=0}^{2^j-1} \delta i_{\beta}((2^j h + k)t_Q - \tau; \tau) \right) \right] \\
 A_2^{-1} &= T \overline{i_{\alpha}(0)} \overline{i_{\beta}(\tau)} \\
 \delta i_x(t; \tau) &= i_x(t) - \overline{i_x(\tau)} \\
 \overline{i_x(\tau)} &= \frac{1}{T} \sum_{d=0}^{T-1} i_x(d t_Q - \tau)
 \end{aligned} \tag{4}$$

Here, T is the total number of data points used to calculate the correlation, t_Q is the time resolution of the raw data (typically 800 ns), and each τ has a j value that depends on the way the integral was set up (below).

A_2 fulfills the Symmetric Normalization criterion [4] by calculating the average value of i_x as $\overline{i_x(\tau)}$ instead of $\overline{i_x(0)}$. The resulting correlation function is less skewed by instrumental drift such as laser power oscillations and focus creep. The same lagged averages $\overline{i_x(\tau)}$ are used to calculate $\delta i_x(t; \tau)$ to ensure the sum of all $\delta i_x(t; \tau)$ converges to 0. This requires a unique calculation of δi_x for every unique τ value considered.

Evaluating Eqn. 4 at every timepoint requires data structures containing running counts of zero-delay data $\{i_{\alpha}(t), i_{\beta}(t), i_{\gamma}(t)\}$ and finite-delay data $\{i_{\alpha}(t - \tau), i_{\beta}(t - \tau), i_{\gamma}(t - \tau)\}$. The use of fractal data structures each containing lower and lower time resolution data [11] allowed the delayed data to be moved and stored in a memory-efficient way. Negative delay times are used so that raw data are accessed only once, are used to calculate fast-time-resolution correlations, and then are summed with other values in gradually larger delay-time bins as new data are read in and t increases. The three zero-delay and three finite-delay data structures are multiplied in all nine combinations to generate three auto and six cross correlations. To improve data quality, the final cross correlations are calculated by averaging pairs of correlations, which is equivalent to averaging positive and negative delay times under the assumption that the system is in equilibrium and the two correlations are equal [20, 21],

$$\begin{aligned}
 G_{\alpha \times \beta}(|\tau|) &= \frac{G_{\alpha \times \beta}(-\tau) + G_{\beta \times \alpha}(-\tau)}{2} \\
 &= \frac{G_{\alpha \times \beta}(-\tau) + G_{\alpha \times \beta}(\tau)}{2}
 \end{aligned} \tag{5}$$

Returning to Eqn. 4, the value j in the sum limits varies as τ increases. The relationship between j and τ is determined by n , an integer used to generate a series of τ values that span a large range in a pseudo-logarithmic fashion. For convenience, n is always a power of two, and larger values of n result in more points spanning the range (Fig. 7A), which generates correlation curves with finer time-resolution and more data points for fitting (Fig. 7B). For double correlations, calculation times scale approximately as $n^{3/4}$ (Fig. 7C), and $n = 32$ is routinely chosen as a compromise between time-resolution and computational cost. To reduce computation time, data sets are broken into overlapping fragments and processed by individual threads, which results in very efficient scaling of performance with additional cores.

4.2. F3CS_AxBxG and two similar programs calculate triple-correlation (F3CS) integrals

F3CS integrals use the same data structures as FCS integrals, but on a more complex schedule [11], and each τ_1 or τ_2 has a j or ω value that again depends on the values of n selected,

$$G_{\alpha \times \beta \times \gamma}^{\text{smr}}(\tau_1, \tau_2) = \frac{\mathbb{A}_3}{2^{j+\omega}} \sum_{h=0}^{T/2^{\omega}-1} \left[\sum_{k=0}^{2^{\omega-j}-1} \left[\left(\sum_{l=0}^{2^j-1} \delta i_{\alpha}((2^{\omega}h+2^j k+l)t_{\omega};0) \right) \right. \right. \\ \left. \left. \times \left(\sum_{l=0}^{2^j-1} \delta i_{\beta}((2^{\omega}h+2^j k+l)t_{\omega} - \tau_1; \tau_1) \right) \right] \right] \\ \times \left(\sum_{k=0}^{2^{\omega}-1} \delta i_{\gamma}((2^{\omega}h+k)t_{\omega} - \tau_2; \tau_2) \right) \Bigg]; \quad \omega \geq j; \quad \tau_1 \leq \tau_2 \\ \mathbb{A}_3^{-1} = T i_{\alpha}(0) i_{\beta}(\tau_1) i_{\gamma}(\tau_2) \quad (6)$$

When calculating triple-cross-correlations, each channel $\{i_{\alpha}, i_{\beta}, i_{\gamma}\}$ contributes to the integrand. Six unique correlations are calculated ($G_{\alpha \times \beta \times \gamma}(\tau_2, \tau_1)$, $G_{\alpha \times \gamma \times \beta}(\tau_2, \tau_1)$, $G_{\beta \times \gamma \times \alpha}(\tau_2, \tau_1)$, $G_{\beta \times \alpha \times \gamma}(\tau_2, \tau_1)$, $G_{\gamma \times \alpha \times \beta}(\tau_2, \tau_1)$, $G_{\gamma \times \beta \times \alpha}(\tau_2, \tau_1)$), each over the triangular domain $(\tau_1, \tau_2); \tau_1 \leq \tau_2$. For convenience, data are stored, displayed and fit as three pairs of correlations ($G_{\alpha \times \beta \times \gamma}(\tau_2, \tau_1)$, $G_{\beta \times \gamma \times \alpha}(\tau_2, \tau_1)$, $G_{\gamma \times \alpha \times \beta}(\tau_2, \tau_1)$) over the full (square) positive domain by using the following type of mapping:

$$G_{\alpha \times \gamma \times \beta}(\tau_1, \tau_2) \rightarrow G_{\alpha \times \beta \times \gamma}(\tau_2, \tau_1) \quad \tau_1 > \tau_2 \quad (7)$$

As triple correlations generate many more data points than double correlations (e.g. $n = 32 \rightarrow 170,000$ datapoints) and take longer to compute (Fig. 7B,C), keeping n within the range 2 – 8 creates smaller data sets that are easily handled by fitting programs but still have an adequate number of photons per data point.

4.3. F3CS_Outlier2 cleans up double-correlation data contaminated by aggregates in solution

When debris contaminates correlation data (Fig 6B), it is usually over a fast enough time scale (< 1 s) that the contamination is restricted to 1–2 of the overlapping datasets from which individual FCS curves are calculated. Individual contaminated curves (Fig 6C) are often distinct enough that they can be identified and rejected [22] before they are averaged together to create the final data (Fig 6D). F3CS_Outlier2 identifies and eliminates contaminated curves using an outlier identification strategy similar to a recently-published algorithm [23], except that the selection stage used here is based on a general statistical test.

Briefly, the algorithm reduces each calculated curve to a sum, GS , by summing $G(\tau)$ over a range of delay values ($10 \mu\text{s} \leq \tau \leq 100 \text{ ms}$). For each of the six auto- and cross-correlations, the average \overline{GS} and standard deviation σ_{GS} of the GS distribution is computed, and for every sub-curve, the weighted deviation from the mean is calculated as $d = |GS - \overline{GS}|/\sigma_{GS}$. If the largest value of d exceeds a user-defined threshold χ , the associated sub-curve is removed from the dataset, along with the other five auto- and cross-correlations derived from the same data, and the process repeats until $d < \chi$ for all d . By using χ values in the range of 2.8 – 3.2, anomalies are reliably removed while preserving data and the success of the outlier rejection algorithm is apparent when looking at final data (Fig 6D).

After outlier rejection, F3CS_Outlier2 performs two additional functions: binning correlation curves calculated from data collected over a selected time range (typically 13 – 500 s) and estimating experimental errors for each point in each correlation curve as the

standard error of the mean. In practice, the progress of the outlier rejection program can be evaluated by comparing these experimental error estimates with the point-to-point fluctuations in the correlation curve, which should be of similar magnitude in the absence of outliers.

4.4. F3CS_Outlier3 cleans up triple-correlation data contaminated by aggregates in solution

The triple correlation version of outlier-rejection program (output, Fig. 6E) performs data rejection based only on double correlation data embedded in the .bin2 files, so that the same portions of a data set are rejected in both double and triple-correlation workflows.

5. Data fitting routines

At the final stage of the data analysis pipeline, two fitting programs allow different levels of analysis to be performed (Fig. 8A). The simplest program, F3CS_LocalFit, fits individual $G(\tau)$ curves to basic FCS fit equations [24] that describe apparent particle number and hydrodynamic properties of the sample under consideration. The second fitting program, F3CS GlobalFit, globally analyzes a full set of double and triple correlation data by considering the influence of every molecular species on every correlation curve. This latter approach requires a more sophisticated treatment of cross-talk, but can return actual, not apparent, particle numbers of complex mixtures.

In both programs, user control over the fitting process is provided by input files that specify initial guesses for fit parameters and determine whether to hold each parameter constant or let it float in the fit. A full list of user-specifiable fit parameters is given in the user manual. The fit processes use a Levenberg-Marquardt algorithm from the Gnu Scientific Library [25]. The algorithms are fast enough and robust enough to handle the large amount of data generated by triple correlation functions, which for a set of six double and three triple correlations is approximately 14,000 data points. Once the fit converges, calculated fits and weighted residuals are written to tab-delimited .dat files that can be plotted with many packages, notably Gnuplot. Several Gnuplot scripts are automatically generated with which to rapidly review data and assess fit qualities. Once fits are acceptable, fit parameters are written to a .dat file for further analysis with external packages.

5.1. F3CS_LocalFit fits FCS curves to apparent particle numbers

The fit functions for auto-correlations $G_{\alpha \times \alpha}(\tau)$ and cross-correlations $G_{\alpha \times \beta}(\tau)$ are as follows:

$$G_{\alpha \times \alpha}(\tau) = \frac{1}{N} \left[\theta G_j(\tau) + (1 - \theta) G_k(\tau) \right] T(\tau) + \frac{A(\tau)}{\langle i_{\alpha}(t) \rangle} + G(\infty) \quad (8)$$

$$G_{\alpha \times \beta}(\tau) = \frac{1}{N} \left[\theta G_j(\tau) + (1 - \theta) G_k(\tau) \right] T(\tau) + G(\infty) \quad (9)$$

Here, N is the total number of molecules per focal volume, $G_j(\tau)$ is the correlation function of species j , $T(\tau)$ corrects for triplet state excitation and $A(\tau)$ corrects for the opto-electronic artifact afterpulsing in auto-correlation data only. Each FCS curve is fit to a mixture of two species j and k , with the fraction of species j represented as θ . Two species fits are useful for determining the apparent extent of binding. $G(\infty)$ terms allow for instrument drift at long delay times, as even slight amounts of instrument drift can be detected with great precision at long lag times [11] and can dominate the fitting process if not allowed to float.

Afterpulsing was corrected for as described [26], using calibration data collected once per day by observing a random light source and fit to the polynomial

$$A(\tau) = k_2\tau^{-2} + k_3\tau^{-3} + k_4\tau^{-4} \quad \tau < 50 \mu s \quad (10)$$

For each species with a unique diffusion time D_j , the per-species correlation function is

$$G_j(\tau) = f_\gamma M_j(\tau; 1) + (1 - f_\gamma) M_j(\tau; \gamma) \quad (11)$$

$$M_j(\tau; \gamma) = \frac{1}{1 + \tau/\gamma\tau_{D,j}} \sqrt{\frac{1}{1 + \tau/\omega^2\gamma\tau_{D,j}}} \quad (12)$$

$$\tau_{D,j} = \frac{r_0^2}{4D_j}; \quad \omega = \frac{z_0}{r_0}$$

The decay function $M_j(\tau; 1)$ accounts for the shape of the decay curve, under the assumption that the microscope focal volume is a gaussian ellipsoid with horizontal and vertical $1/e^2$ radii r_0 and z_0 . The molecular species j diffuses with characteristic diffusion time D_j . Using two decay functions $M_j(\tau; 1)$ and $M_j(\tau; \gamma)$, in a ratio determined by f_γ , is a common strategy to improve fits [27] (a similar approach did not yield better fits to triple-correlation data). Triplet state transitions in FCS data are described [27] by the fraction of molecules in the triplet state $T_{0,j}$ and the average triplet-state lifetime $\tau_{f,j}$

$$T(\tau) = 1 + \frac{T_{0,j}\theta e^{-\tau/\tau_{f,j}} + T_{0,k}(1 - \theta)e^{-\tau/\tau_{f,k}}}{1 - T_{0,j}\theta - T_{0,k}(1 - \theta)} \quad (13)$$

F3CS_LocalFit treats each fit parameter in one of three ways: by holding it fixed, fitting it to a specific curve, or fitting it globally to all curves. For example, a binding reaction could be probed using one $G\beta \times \gamma(\tau)$ curve per time point (Fig. 8B). Global variables would include the microscope parameters ω , γ and f_γ and intrinsic molecular parameters τ_D for both unbound and bound species. Local parameters would include N and θ values that are specific to individual reactions. Plotting θ vs. time would yield the desired kinetic binding curve.

5.2. F3CS_GlobalFit fits double and triple correlation curves to a single model

A large amount of information about a mixture of chemical or biochemical species in a sample can be recovered by global fitting of FCS [28, 29, 30] and F3CS curves calculated from the same fluorescence data. The present aim of fitting is to recover the vector of species concentrations $\vec{N} = (N_1, N_2, \dots, N_M)$ and by so doing, understand the populations of various molecules in the sample (Fig. 8C–D). While in principle each species contributes to each FCS and F3CS curve, identification relies on labeling strategies to bias certain species to contribute disproportionately to certain curves. The fitting routine accounts for this with weight factors $Y_{\alpha,j}$, which correspond to the brightness of species j at detector α , and experimental data are weighted as follows:

$$G_{\alpha \times \alpha}(\tau) = \frac{\sum_j Y_{\alpha,j}^2 V_{\alpha,\alpha} N_j M_j(\tau; 1) T_j(\tau)}{\langle i_\alpha(t) \rangle \langle i_\alpha(t+\tau) \rangle} + \frac{A(\tau)}{\langle i_\alpha(t) \rangle} + G_{\alpha \times \alpha}(\infty) \quad (14)$$

$$G_{\alpha \times \beta}(\tau) = \frac{\sum_j^M Y_{\alpha,j} Y_{\beta,j} V_{\alpha,\beta} N_j M_j(\tau;1)}{\langle i_{\alpha}(t) \rangle \langle i_{\beta}(t+\tau) \rangle} + G_{\alpha \times \beta}(\infty) \quad (15)$$

$$T_j(\tau) = 1 + \frac{T_{0,j} e^{-\tau/\tau_{f,j}}}{1 - T_{0,j}}$$

$$\langle i_{\alpha}(t) \rangle = b_{\alpha} + \sum_j^M Y_{\alpha,j} N_j \quad (16)$$

$$G_{\alpha \times \beta \times \gamma}(\tau_1, \tau_2) = \left(\frac{\gamma_3}{\gamma_2^2} \right) \frac{\sum_j^M Y_{\alpha,j} Y_{\beta,j} Y_{\gamma,j} N_j M_j(\tau_1, \tau_2)}{\langle i_{\alpha}(t) \rangle \langle i_{\beta}(t+\tau_1) \rangle \langle i_{\gamma}(t+\tau_2) \rangle} + G_{\alpha \times \beta \times \gamma}(\infty, \infty)$$

$$M_j(\tau_1, \tau_2) = \frac{1}{1 + 4\tau_1(\tau_2 - \tau_1)/3\tau_{D,j}^2 + 4\tau_2/3\tau_{D,j}} \quad (17)$$

$$\times \sqrt{\frac{1}{1 + 4\tau_1(\tau_2 - \tau_1)/3\omega^4\tau_{D,j}^2 + 4\tau_2/3\omega^2\tau_{D,j}}}$$

Compared with local fits in the previous section, the auto- and cross-correlation terms appear simpler because multiple species i and j are treated via the sum over all M species considered. The additional term $V_{\alpha,\beta}$ is an observation volume correction factor. With wide-field detection, all values are 1 within error, while confocal detection volumes across multiple colors can vary on the order of 10% depending on how well optics are corrected. The correction factor is obtained by observing single fluorescent species that fluoresces in two different channels. Other varieties of triple-correlation data e.g. $G_{\alpha \times \alpha \times \alpha}(\tau_1, \tau_2)$, $G_{\alpha \times \alpha \times \beta}(\tau_1, \tau_2)$ present unique information that could be included along the lines of high order correlation analysis [30], but were left out of the fitting to keep the fitting matrices manageably small. The ratio γ_3/γ_2^2 describes the relative sizes of double and triple correlation focal volumes (~ 1.65) and is calibrated experimentally as described [11].

The apparent brightness matrix $Y_{\alpha,j}$ is a product of matrices describing molecular properties and instrument factors,

$$Y = C \Xi Q \text{Occ} \quad (18)$$

Here, C is a diagonal matrix of correction terms that account for drift of one detector vs another, Ξ is a matrix of bleedthrough values that quantifies the effects of detection optics on how well each fluorophore is detected by each APD, Q is a diagonal matrix with the brightness values of each fluorophore, and Occ is a labeling occupancy matrix that contains information about experimental design and which dye is present on which species.

F3CS_GlobalFit jointly fits auto-correlation FCS (Eq 14), cross-correlation FCS (Eq 15), F3CS (Eq 17) and intensity values (Eq 16) by weighting each piece of data by its error estimate, and the amplitudes of weighted fit residuals do not show major differences between different types of data. As three-dimensional plots are relatively difficult to visualize, a separate version (F3CS_GlobalFit_3D_Plot) uses the library PLplot to generate .ps files containing shaded plots of data, fits, and residuals (e.g. Fig 6E).

6. Time reversal asymmetry detection for analysis of irreversible processes

Irreversible processes such as non-equilibrium steady state reaction networks [20] can be detected by a finite-valued difference curve [11]

$$\Delta G_{\alpha \times \beta \times \gamma}(\tau_1, \tau_2) = G_{\alpha \times \beta \times \gamma}(\tau_1, \tau_2) - G_{\alpha \times \beta \times \gamma}(-\tau_1, -\tau_2) \quad (19)$$

6.1. F3CS_Difference

The analysis (Fig. 9A) first requires calculating triple-correlation integrals using both raw data (Fig. 9B) and time-reversed data (Fig. 9C) generated with F3CS_Reverser. F3CS_Difference then reads both sets of correlation integrals and subtracts one from the other (Fig. 9D). If the forward and backwards correlations are identical, then the difference will be zero within error, and not much can be concluded either way. If the correlations differ and the difference curves contain non-zero features, as they do in Fig. 9D, the system is irreversible. The data in Fig. 9 are from a sample under intense laser irradiation which causes irreversible photobleaching of the fluorescent dye. Forward and backward F3CS data are subtracted after the data have been purged of outliers using the same outlier-rejection algorithms discussed earlier.

7. Concluding remarks

As Triple Correlation Toolbox is a complete package, any research group with a three-color single-molecule microscope and an inexpensive hardware card can now use F3CS and FCS together to measure populations in samples of unprecedented complexity. Many programs are generally useful for FCS alone and the data acquisition program provides an inexpensive but information rich detection scheme for any microscope, in addition to performing FCS or F3CS.

For program usage, installation and a tutorial, please see the accompanying user manual, F3CS_Manual.pdf.

Acknowledgments

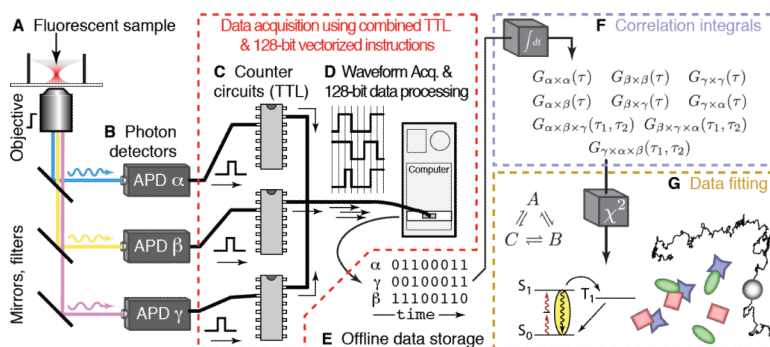
The authors thank MT Sykes for his insightful comments regarding most areas of this project and WS Young for comments and computing infrastructure.

9. Role of the funding source The National Institutes of Health (R37-GM-53757 to J.R.W., GM-044060 to D.P.M.) are gratefully acknowledged for funding. This material is based upon work supported under a National Science Foundation Graduate Research Fellowship (W.K.R.). Funding agencies had no involvement with study design; collection, analysis and interpretation of data; the writing of the report; or the decision to submit the article for publication.

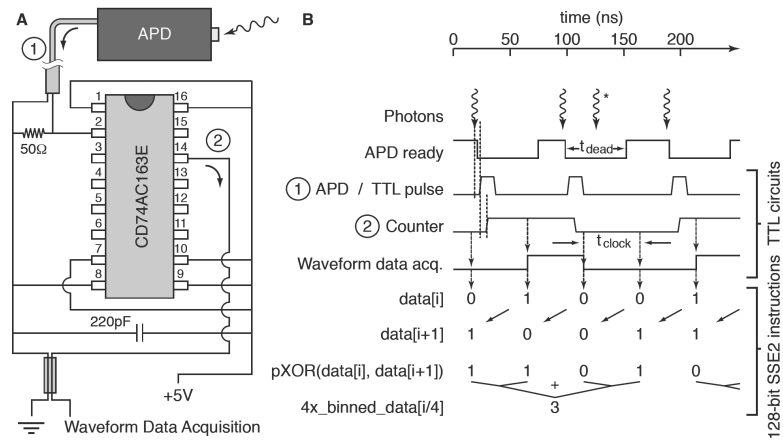
References

- [1]. Magde D, Elson E, Webb W. Phys. Rev. Lett. 1972; 29:705.
- [2]. Haustein E, Schwille P. Annu. Rev. Biophys. Biomol. Struct. 2007; 36:151. [PubMed: 17477838]
- [3]. Digman MA, Gratton E. Annu. Rev. Phys. Chem. 2011; 62:645. [PubMed: 21219151]
- [4]. Schatzel K, Drewel M, Stimac S. J. Mod. Opt. 1988; 35:711.
- [5]. Schatzel K. Quantum Opt. 1990; 2:287.
- [6]. Magatti D, Ferri F. Rev. Sci. Instrum. 2003; 74:1135.
- [7]. Eid JS, Muller JD, Gratton E. Rev. Sci. Instrum. 2000; 71:361.
- [8]. Yang LL, et al. J Microsc. 2009; 234:302. [PubMed: 19493109]
- [9]. Laurence TA, Fore S, Huser T. Opt. Lett. 2006; 31:829. [PubMed: 16544638]
- [10]. Schwille P, Meyer-Almes FJ, Rigler R. Biophys. J. 1997; 72:1878. [PubMed: 9083691]
- [11]. Ridgeway WK, Millar DP, Williamson JR. J. Phys. Chem. B. 2012; 116:1908. [PubMed: 22229664]
- [12]. Wobma HM, et al. Phys. Chem. Chem. Phys. 2012; 14:3290. [PubMed: 22249466]

- [13]. Czerwinski F, Oddershede LB. *Comput. Phys. Commun.* 2011; 182:485.
- [14]. Oberman S, Favor G, Weber F. *Micro, IEEE.* 1999; 19:37.
- [15]. Raman SK, Pentkovski V, Keshava J. *Micro, IEEE.* 2000; 20:47.
- [16]. Qian H, Elson EL. *Biophys. J.* 1990; 57:375. [PubMed: 2317556]
- [17]. Qian H. *Biophys. Chem.* 1990; 38:49. [PubMed: 2085652]
- [18]. Palmer AG, Thompson NL. *Rev. Sci. Instrum.* 1989; 60:624.
- [19]. Naishlos, D. Proceedings of the 2004 GCC Developers Summit. 2004. Autovectorization in GCC; p. 105-118. CiteSeer
- [20]. Qian H, Elson E. *Proc. Natl. Acad. Sci. U. S. A.* 2004; 101:2828. [PubMed: 14970342]
- [21]. Steinberg IZ. *Biophys. J.* 1986; 50:171. [PubMed: 3730501]
- [22]. Bacia K, Schwille P. *Nat. Protoc.* 2007; 2:2842. [PubMed: 18007619]
- [23]. Ries J, et al. *Opt. Express.* 2010; 18:11073. [PubMed: 20588964]
- [24]. Davis LM, Williams PE, Ball DA, Swift KM, Matayoshi ED. *Curr. Pharm. Biotechnol.* 2003; 4:451. [PubMed: 14683437]
- [25]. Galassi, M., et al. *GNU Scientific Library Reference Manual.* 2 edition. Network Theory Ltd.; 2003.
- [26]. Zhao M, et al. *Appl. Optics.* 2003; 42:4031.
- [27]. Widengren J, Mets U, Rigler R. *J. Phys. Chem.* 1995; 99:13368.
- [28]. Eggeling C, Kask P, Winkler D, Jager S. *Biophys. J.* 2005; 89:605. [PubMed: 15849243]
- [29]. Hwang LC, Gosch M, Lasser T, Wohland T. *Biophys. J.* 2006; 91:715. [PubMed: 16632502]
- [30]. Palmer AG, Thompson NL. *Biophys. J.* 1987; 52:257. [PubMed: 3663831]

**Figure 1.**

The F3CS experiment and data analysis. Arrows represent the flow of information in the form of photons, electrons and data. (A) A fluorescent sample is detected by microscope optics with single-molecule sensitivity. (B) Three independent Avalanche Photo Diodes (APDs) convert photons into electrical signals (TTL) that are fed into (C) short term memory in the form of off-the-shelf counter-timer chips. (D) A data acquisition card reads the counter-timer chips every 50 ns and (E) photon counts from all three channels are written to disk. (F) Photon counts are later analyzed by a family of correlation integrals. (G) The resulting correlation curves are fit to global models of phenomena such as reversible and irreversible linked chemical reactions, triplet-state kinetics, complex interactions of bio-molecules (colored shapes) and diffusion time measurements.

**Figure 2.**

A data acquisition scheme that minimizes data transfer bandwidth in order to achieve high read speeds. (A) Diagram of the counter-timer circuit used to couple APD photon-detection pulses to waveform data acquisition (PCIe-6251, National Instruments). (B) Timing diagram showing the mixture of TTL-level signals and 128-bit SSE2 instructions used to convert photon arrivals into single-bit records, $\text{data}[i]$, which must be transferred from data acquisition hardware to RAM. The photon marked with an asterisk will not be detected due to APD dead time.

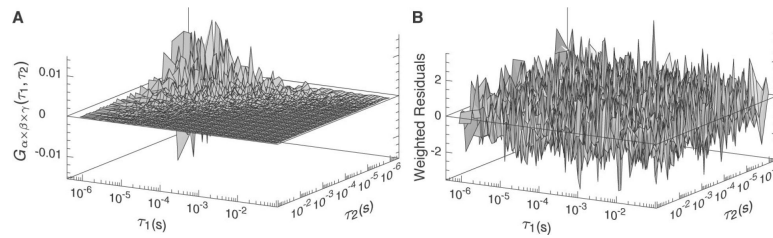


Figure 3.

Precision of a triple correlation of a random signal (room lights). (A) F3CS data are zero-valued, as expected, but are not obtained with equal precision throughout the dataset. Data were estimated with the least precision at short delay times and precision increased as delay times lengthened and more data were binned to create each point. (B) Weighted residuals after fitting the data to a single constant ($G(\tau_1, \tau_2) = C$). The trivial fit had a reduced chi-squared of approximately one ($\chi^2 = 1.04$), which verified that the data analysis pipeline correctly estimated and propagated experimental errors. The flat residuals show that data do not contain distortions that could potentially arise from cross talk between electronic components.

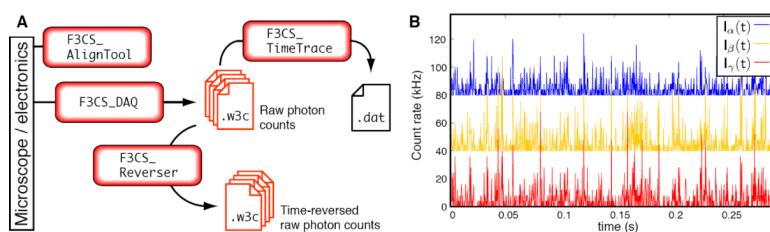


Figure 4. Software for data acquisition. (A) Workflow describing the relationship between several programs (rounded rectangles) and data files. (B) Sample time-trace data binned to 1 kHz; spikes result from molecules diffusing through the focal volume.

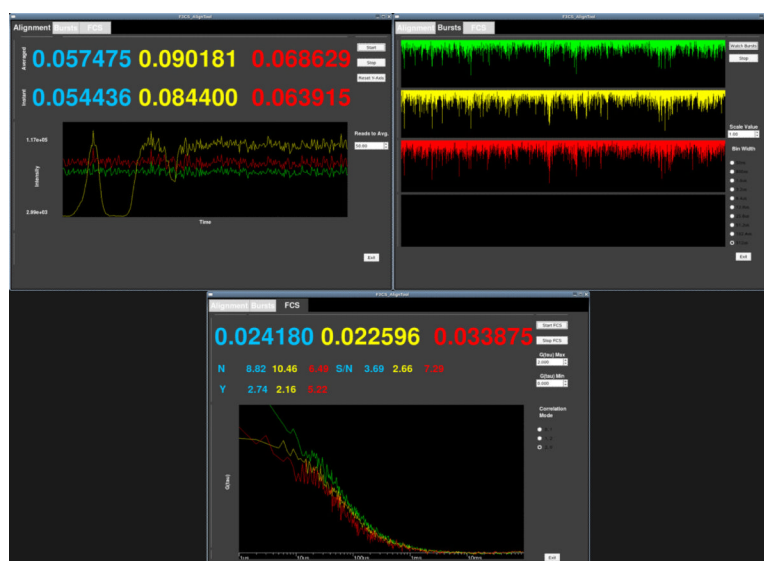
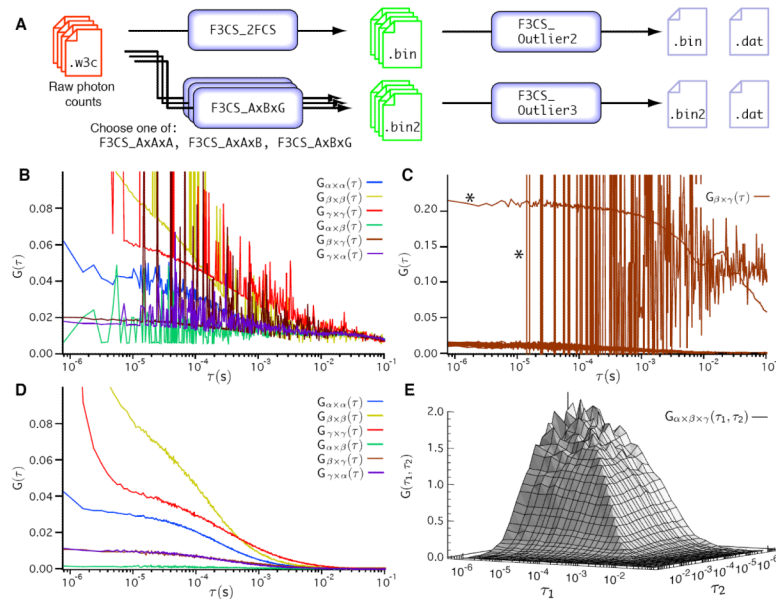


Figure 5.

F3CS_AlignTool: A Graphical User Interface tool for realtime display of fluorescence intensity and FCS data. The aesthetics look unusual because they were tuned for a dimmed monitor in a dark laser lab where this program is usually used to assist the user when setting up experiments, and because WKR has a very limited grasp of style. Program appearance also varies by operating system. Three main modes are reflected in three tabs: *Alignment* displays continuous, real-time intensity traces for aligning the microscope. The yellow channel APD is being aligned in this example. Green time-traces are used instead of blue throughout to increase low-light visibility. *Bursts* displays very fast intensity traces that allow the time-structure of the signals to be examined. In this example, individual molecules are observed. *FCS* computes quick auto- and cross-correlations to assess signal / noise and screen for extraneous fluctuations from laser Q-switching, table vibrations, et cetera.

**Figure 6.**

Software for calculating double and triple correlation integrals, and outlier rejection. (A) Photon count data are processed independently by two programs, one performing a set of six double-correlation FCS integrals, the other performing a set of three F3CS integrals. Three versions of the latter program each compute different sets of triple correlations; of these, F3CS_AxBxG is used most frequently. (B) The output of the correlation programs can not be used directly, as debris diffusing through the focal volume contaminates about 1% of FCS data and produces the jagged lines and raised amplitudes seen here. (C) As contaminants are usually restricted to a small portion of the dataset, correlations are calculated by breaking each data set into ~ 1 s fragments and correlating each fragment independently. Here, only two of the dataset fragments are contaminated (marked with *), and these curves are quite distinct from the rest. (D) The same six double correlations after outlier rejection programs eliminated the contaminated curves from the average. (E) Triple-correlation data cleaned using the same algorithm.

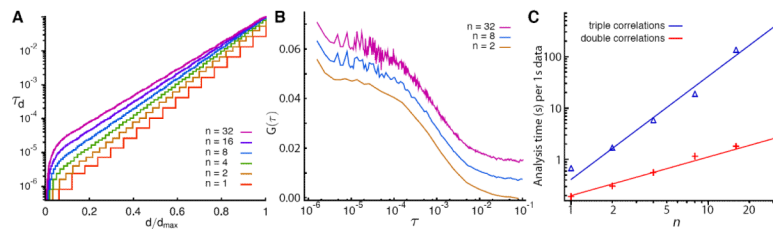
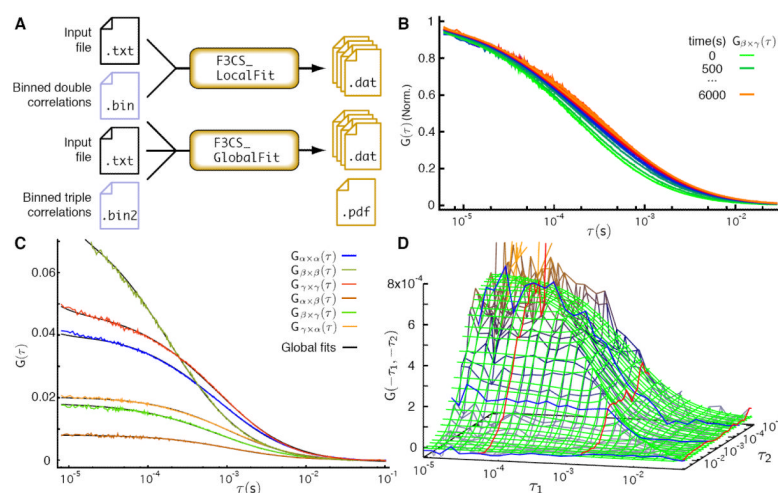
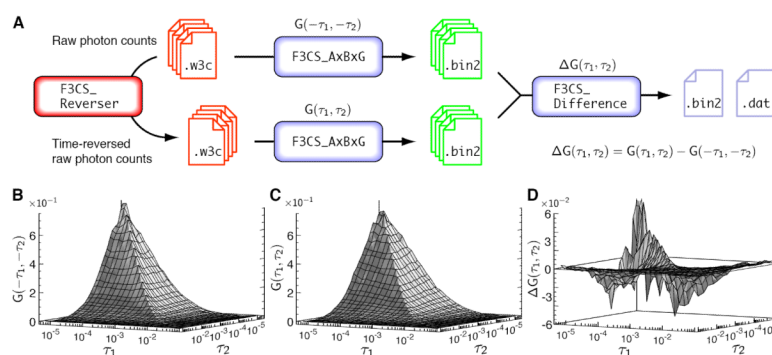


Figure 7.

The range of τ_D values used in correlation integral analysis spans several orders of magnitude, and how fine the coverage is within this series is determined by the integer n . (A) Delay times τ_D vs. d (delay time index, normalized) for several values of n show that larger n gives finer-scale coverage at longer τ_D times. (B) FCS curves calculated w/ different n display different degrees of time resolution and numbers of datapoints (curves shifted for clarity). (C) Computation time increases as $n^{3/4}$ for double correlations and as n^2 for triple correlations. Timings performed on a single core of an AMD Phenom II X6 1055T.

**Figure 8.**

Software for fitting double and triple correlation curves to models. (A) The workflow requires both input files and data to run two independent fitting programs: the light-weight program F3CS_LocalFit for fitting double-correlation curves independently of each other, and F3CS_GlobalFit for fitting double-correlation, triple-correlation and intensity data globally. (B) Sample fits using F3CS_LocalFit to fit multiple time points of a binding reaction using common instrument and molecular parameters but individual fraction bound (θ) parameters. (C) Double-correlation and (D) Triple-correlation data fit to the same global model.

**Figure 9.**

Software for time-reversal asymmetry detection. (A) The workflow largely follows the established flow for triple-correlation integrals, but performs the integrals twice – once on raw data and once on time-reversed data. For each timepoint, the correlation integrals are paired up and subtracted by F3CS_Difference. Data written to .dat files can be visualized using gnuplot. (B) Forward correlation (C) Reverse correlation and (D) Difference data illustrating irreversible photobleaching [11]

Table 1

Programs used to generate and analyze triple-correlation data.

Major programs in the Triple Correlation Toolbox suite					
Name	Language	Libraries	Run Time	Cores	File Size
Data Acquisition					
F3CS_Daq	C, SSE2 Assembly	NIDAQmx pthread	Real-time	2	96 MB per 13.4 s Data
F3CS_AlignTool	C	NIDAQmx gtk+-2.x	Real-time	1	–
Auto- and Cross-Correlation Integrals (high time resolution)					
F3CS_2FCS	C	pthread	26 s^{\dagger} per 13.4 s Data	$1 - N^{\ddagger}$	1 MB per 13.4 s Data
Triple Correlation Integrals					
F3CS_AxBxG	C	pthread	36 s^{\dagger} per 13.4 s Data	$1 - N^{\ddagger}$	19 MB per 13.4 s Data
Outlier rejection					
F3CS_Outlier2	C	–	1 – 20 s	1	140 kB per curve
F3CS_Outlier3	C	–	1 – 40s	1	300 kB per curve
Fit triple correlations globally with auto- and cross-correlations and intensity					
F3CS_GlobalFit	C	GSL	10 – 100 s	1	–
Fit individual auto- and cross-correlations (lightweight fitting program)					
F3CS_LocalFit	C	GSL	~ 10 s	1	–

[†]Timing for multithreaded programs is recorded for 2 cores (AMD Phenom II X6 1055T); correlation programs scale efficiently with new cores. Time increases if the correlation curve is calculated at more closely-spaced τ values.

[‡]The maximum number of cores is limited by the size of the data sets. The NIDAQmx library handles card I/O (National Instruments, Austin TX). GSL (Gnu Scientific Library) provides Levenberg – Marquardt non-linear least-squares optimization routines.

Table 2

Correlation integrals evaluated by each program.

Program	Correlation Integrals		
F3CS_2FCS	$G_{a \times a}(\tau)$	$G_{\beta \times \beta}(\tau)$	$G_{\gamma \times \gamma}(\tau)$
	$G_{a \times \beta}(\tau)$	$G_{\beta \times \gamma}(\tau)$	$G_{\gamma \times a}(\tau)$
F3CS_AxAxA	$G_{a \times a \times a}(-\tau_1, -\tau_2)$	$G_{\beta \times \beta \times \beta}(-\tau_1, -\tau_2)$	$G_{\gamma \times \gamma \times \gamma}(-\tau_1, -\tau_2)$
F3CS_AxAxA [†]	$G_{a \times a \times a}(\tau_1, \tau_2)$	$G_{\beta \times \beta \times \beta}(\tau_1, \tau_2)$	$G_{\gamma \times \gamma \times \gamma}(\tau_1, \tau_2)$
F3CS_AxAxB	$G_{a \times a \times \beta}(-\tau_1, -\tau_2)$	$G_{\beta \times \beta \times \gamma}(-\tau_1, -\tau_2)$	$G_{\gamma \times \gamma \times a}(-\tau_1, -\tau_2)$
F3CS_AxAxB [†]	$G_{a \times a \times \beta}(\tau_1, \tau_2)$	$G_{a \times a \times \beta}(\tau_1, \tau_2)$	$G_{\gamma \times \gamma \times a}(\tau_1, \tau_2)$
F3CS_AxBxG	$G_{a \times \beta \times \gamma}(-\tau_1, -\tau_2)$	$G_{\beta \times \gamma \times a}(-\tau_1, -\tau_2)$	$G_{\gamma \times a \times \beta}(-\tau_1, -\tau_2)$
F3CS_AxBxG [†]	$G_{a \times \beta \times \gamma}(\tau_1, \tau_2)$	$G_{\beta \times \gamma \times a}(\tau_1, \tau_2)$	$G_{\gamma \times a \times \beta}(\tau_1, \tau_2)$

[†] Calculated with time-reversed data