Supplementary Information

Title: pawFLIM: reducing bias and uncertainty to enable lower photon count in FLIM experiments.
Authors: Mauro Silberberg and Hernán E. Grecco

1. Phasor rotation: an interpretation

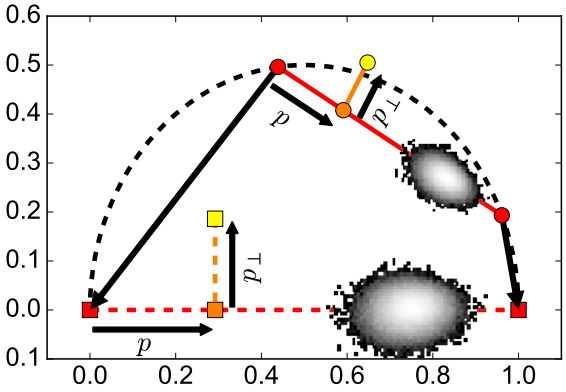

Figure 1. An alternative geometrical interpretation of the normalized scalar projection. Applying (1), the solid red biexponential line is mapped to the dashed red line, rotating and rescaling phasor distributions. After this transformation, p and $p_\perp$ of a noisy phasor (yellow) are simply the real and imaginary part of (1).

In the case of a biexponential decay, all phasors lie along a line joining the monoexponential components, R_1 and R_2 (Fig. 1, red circles). If these are known, we can solve (**??**) for p.

$$p = \frac{R - R_2}{R_1 - R_2} \tag{1}$$

However, in the presence of noise a phasor R won't necessary lie on this segment. Applying directly (1) would yield a complex p value. The usual estimator is the normalized scalar projection to the biexponential line, which is equivalent to a least squares minimization [**?**]. A noisy phasor is projected perpendicularly to the biexponential line (**Sup Fig. 1**, yellow and orange circles). Then, p is estimated as the distance between that point and R_2, normalized by the length of the segment defined by R_1 and R_2 (**Sup. Fig. 1**, solid red line).

But there is another way of interpreting this estimation. Equation (1) is an affine transformation (i.e., a combination of a translation, a rotation and a rescaling) that maps the biexponential segment defined by R_1 and R_2 to the $[0, 1]$ real segment (**Sup. Fig. 1**, red circles and squares). After this *rotation*, the normalized scalar projection is equivalent to taking the real part of (1), while the normalized perpendicular distance is its imaginary part.

This transformation is useful in comparing phasor distributions, since their standard deviation in the real direction is the measure of error in p, as they are rotated and rescaled appropriately.

2. Determination of lifetimes by global analysis

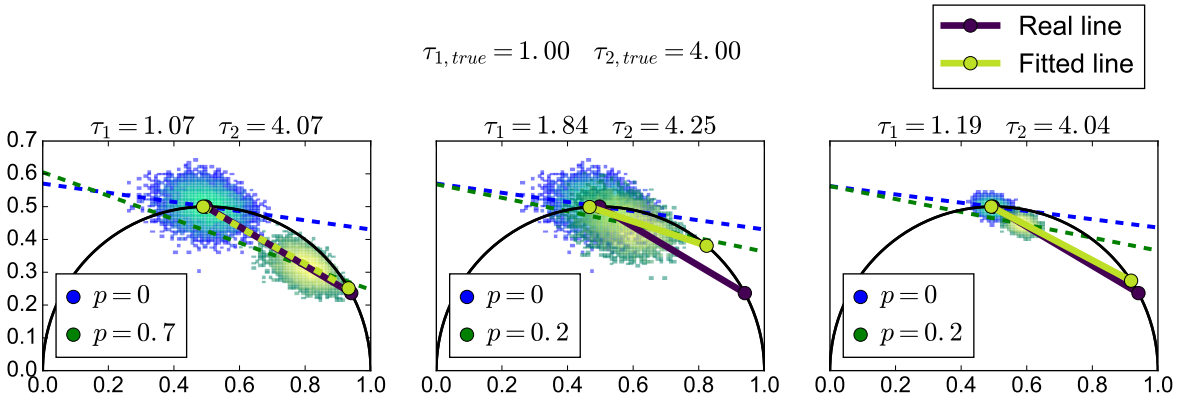

Figure 2. Linear fitting to determine the biexponential line. Each color corresponds to a cloud of a single photon fraction p, and in dotted line of the same color a linear fit to that cloud. **(A)** When there's sufficient variation in p, the line is accurately determined. **(B)** When there is not, the direction of each cloud eclipses the biexponential direction. **(C)** Increasing the number of detected photons, the biexponential direction prevails, recovering the biexponential line more accurately. Note that the direction of each phasor cloud does not change despite the increase in photon count.

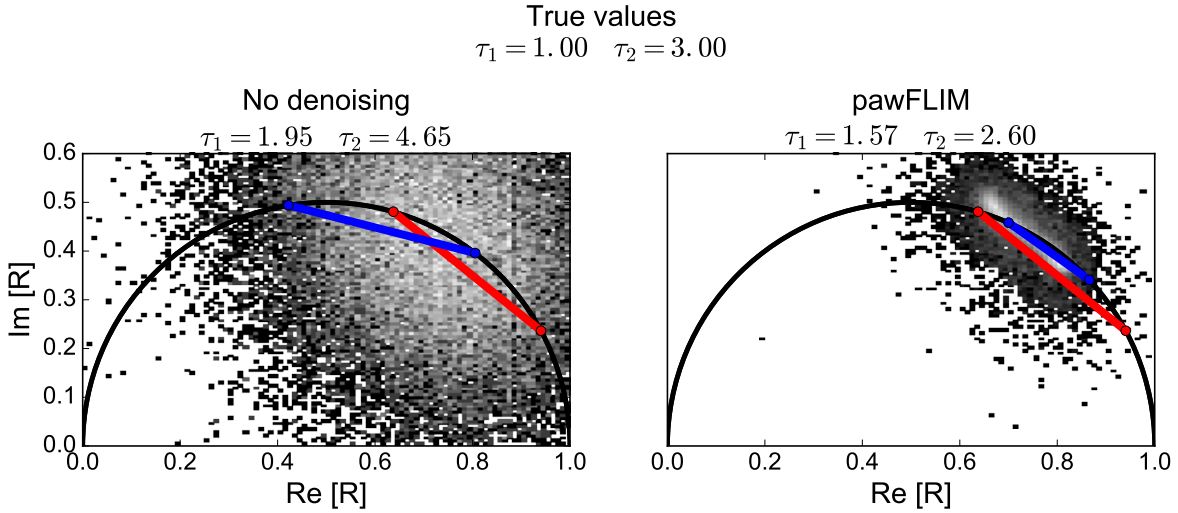

Figure 3. Linear fitting to determine the biexponential line. The denoising algorithm enhances the phasor distribution, in turn improving the global determination of lifetimes.

With a single phasor, we can't estimate all the necessary parameters that describe a biexponential decay. In that case, there are three parameters (two lifetimes and photon fraction) and two "measurements" (real and imaginary parts). This is more evident in the phasor plot representation: there are several lines that go through a single phasor. One would have to fit the decay in the time domain or use at least two harmonics. Both approaches imply non-linear minimizations.

Global analysis allows us to determine global lifetimes from a single harmonic by using multiple measurements [**?**]. Again, this is evident in the phasor plot, given that biexponential decays lie along a line joining the monoexponential phasors (**Sup. Fig. ??**). With at least two phasors of different photon fraction, a line is defined, and global lifetimes can be estimated by finding the intersection of the line with the monoexponential semicircle [**?**].

With more than two phasors, we can find the best line by a least squares minimization, which is a linear procedure. One might assume that the accuracy in determining the line can always be increased by increasing the number of phasors. As shown in **Sup. Fig. 2**, this is not the case.

It has been known that there is a requirement of a sufficient variation in p to estimate the global lifetimes [**?**], a fact which is more evident in the phasor plot representation.

Clouds corresponding to a single p fraction aren't circular, but are elongated along a direction which doesn't necessarily coincide with the biexponential line direction. When there's a low variation in p, these clouds blend together and the direction of a linear fit is dominated by the direction of each cloud. By increasing the number of photons, these clouds get smaller and closer to the true biexponential line and is the direction of this last one that prevails in the linear fit.

In general, reducing the size of the phasor distributions helps in the linear fit. Hence, in addition to p, we introduced the normalized perpendicular distance to the biexponential line, $p_\perp$, as a measure of interest in the determination of the global lifetimes τ_1 and τ_2. A decrease in variance both in p and in $p_\perp$ would aid the linear fit. Nevertheless, note that a bias in $p_\perp$ would displace the linear fit. On the contrary, a bias in p wouldn't hurt on this matter.

3. Uncertainty due to the IRF

Both excitation and detection distort the fluorescence decay, an effect which is summarized by the impulse response function (IRF) of the system. Since the excitation pulse $E(t)$ is broad, each molecule k gets excited at a different time $W_k \sim E(t)$. Later on, each molecule emits a photon at a time $X_k \sim D(t)$, where $D(t)$ is given either by (**??**) or (**??**). Finally, each photon is detected with a certain jitter, and the time of arrival of each photon will have an extra component $Y_k \sim G(t)$, given by the detection system. Hence, the detected time of arrival will be a sum of three random variables $Z_k = W_k + X_k + Y_k$. Since we're only interested in the time of emission of the molecule,

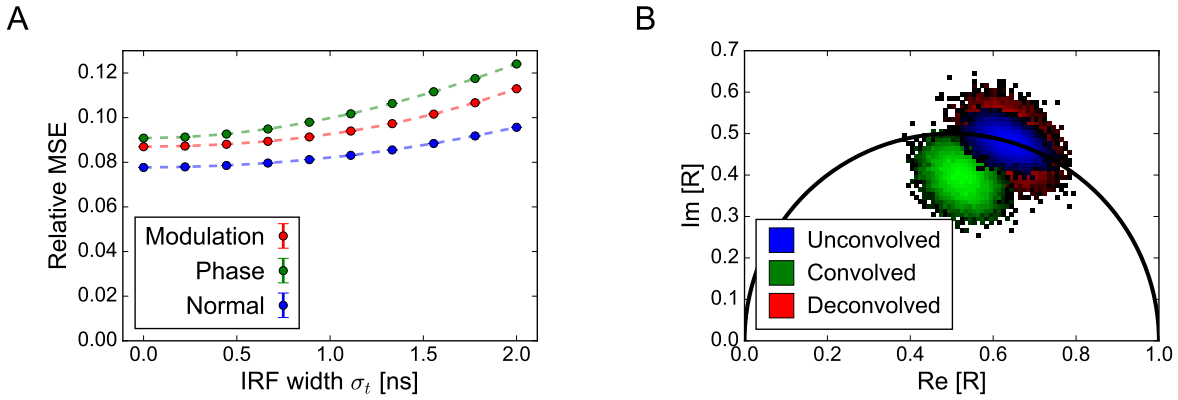

Figure 4. Uncertainty due to the IRF. (A) Mean squared error of the different lifetime estimators as a function of the IRF width σ_t, relative to the simulated lifetime. Monte Carlo simulation of TCSPC for $N = 200$ detected photons, true lifetime $\tau = 3$ ns, detection window $T = 25$ ns, and 7 million iterations. **(B)** Phasor distributions for $\sigma_t = 2.5$ ns. Note that the deconvolved distribution (red) it both bigger and has a different shape than the *unconvolved* one (blue).

the effects of the excitation and the detection are grouped in a single random variable whose distribution is given by $IRF(t)$.

Now, the distribution of a sum of independent random variables is given by the convolution of the individual PDF's. Then, as its is usually introduced, the *expected experimental* fluorescence F can be expressed as a convolution of the fluorescence decay D with the instrument response function IRF:

$$F(t\,|\,\tau) = D(t\,|\,\tau) * \mathrm{IRF}(t) = \int_{-\infty}^{\infty} D(u)\,\mathrm{IRF}(t-u)\,du \tag{2}$$

The IRF has to be characterized to perform a deconvolution and recover the fluorescence decay. But although we could measure the IRF as accurately as desired in an independent experiment, the IRF adds an uncertainty to the measured decay which deconvolution cannot remove. This seems evident in a photon by photon fashion, since each detection time has an independent and different contribution by the IRF.

While we could independently measure the IRF as accurately as desired (i.e. with as many photons as necessary), a (convolved) fluorescence decay will have a limited number of photons and hence only a sample of the IRF. Only by measuring both the IRF and the convolved fluorescence decay with an infinite amount of photons each, we may perform a deconvolution and recover the true fluorescence decay.

To show this, let's consider two random variables X and Y corresponding to the true fluorescence decay and the IRF. We measure the sum $Z = X + Y$.

In general, we will measure N photons arrival times and compute a phasor as:

$$R = \frac{1}{N}\sum_{k}^{N} e^{i\omega Z_k} = \frac{1}{N}\sum_{k}^{N} e^{i\omega(X_k+Y_k)} = \frac{1}{N}\sum_{k}^{N} e^{i\omega X_k} e^{i\omega Y_k} \tag{3}$$

which is another random variable.Note that we're not considering the binning process, that is, there are infinite continuous channels. In such case, we'd have to consider a multinomial distribution of photons, and covariances between bins. As such, we can calculate its expected value:

$$\mathrm{E}\left[R\right] = \mathrm{E}\left[\frac{1}{N}\sum_{k}^{N} e^{i\omega Z_k}\right] = \frac{1}{N}\sum_{k}^{N}\mathrm{E}\left[e^{i\omega Z_k}\right] = \mathrm{E}\left[e^{i\omega Z}\right] = \mathrm{E}\left[e^{i\omega X}e^{i\omega Y}\right] = \mathrm{E}\left[e^{i\omega X}\right]\ \mathrm{E}\left[e^{i\omega Y}\right] \tag{4}$$

last step due to independence between X and Y.

The first term is no other thing than the Fourier transform of the fluorescence decay:

$$\mathrm{E}\left[e^{i\omega X}\right] = \int e^{i\omega x} f_X(x)dx = \int e^{i\omega x}\frac{e^{-t/\tau}}{\tau}dx = \frac{1}{1-i\omega\tau} \tag{5}$$

here shown for a monoexponential decay. Analogously, for the second term we have:

$$\mathrm{E}\left[e^{i\omega Y}\right] = \mathcal{F}[IRF] = R_{IRF} \tag{6}$$

As is known, deconvolution is translated to a division in Fourier space:

$$R \to \frac{R}{R_{IRF}} \tag{7}$$

It can be seen that this procedure corrects the mean value of the phasor:

$$\mathrm{E}\left[\frac{R}{R_{IRF}}\right] = \frac{1}{R_{IRF}}\mathrm{E}\left[R\right] = \mathrm{E}\left[e^{i\omega X}\right] \tag{8}$$

where R_{IRF} is unaffected being a constant.

Subsequently, we can consider the covariance matrix of the phasor. In particular, we will focus on the variances of the real and imaginary parts separately. We'd like to compare the convariance matrices of the *unconvolved* or true fluorescence with the deconvolved fluorescence decay.

When there's no contribution of the IRF (ie $Y = 0$ always) we have:

$$\mathrm{Var}\left[\mathrm{Re}\left[R_{Y=0}\right]\right] = \mathrm{Var}\left[\frac{1}{N}\sum_{k}^{N}\cos\omega X_k\right] = \frac{1}{N^2}\sum_{k}^{N}\mathrm{Var}\left[\cos\omega X_k\right] = \frac{1}{N}\mathrm{Var}\left[\cos\omega X\right] \tag{9}$$

$$\mathrm{Var}\left[\mathrm{Im}\left[R_{Y=0}\right]\right] = \frac{1}{N}\mathrm{Var}\left[\sin\omega X\right] \tag{10}$$

More accurately, we should say that the IRF is a delta function centered at $t = 0$. Instead, if the IRF isn't negligible, we have for the deconvolved phasor:

$$\mathrm{Var}\left[\mathrm{Re}\left[\frac{R}{R_{IRF}}\right]\right] = \frac{1}{N}\mathrm{Var}\left[\frac{1}{\left|R_{IRF}\right|^2}\Big(\mathrm{Re}\left[R_{IRF}\right]\mathrm{Re}\left[R\right] - \mathrm{Im}\left[R_{IRF}\right]\mathrm{Im}\left[R\right]\Big)\right] \tag{11}$$

$$\mathrm{Var}\left[\mathrm{Im}\left[\frac{R}{R_{IRF}}\right]\right] = \frac{1}{N}\mathrm{Var}\left[\frac{1}{\left|R_{IRF}\right|^2}\Big(\mathrm{Re}\left[R_{IRF}\right]\mathrm{Im}\left[R\right] - \mathrm{Im}\left[R_{IRF}\right]\mathrm{Re}\left[R\right]\Big)\right] \tag{12}$$

We can show that these two covariance matrices are not equal (or more generally, similar) by showing that their traces differ. Hence, we define their difference as Δ:

$$\Delta = \mathrm{tr}\left(\mathrm{Cov}\left[\frac{R}{R_{IRF}}\right]\right) - \mathrm{tr}\left(\mathrm{Cov}\left[R_{Y=0}\right]\right) \tag{13}$$

Applying properties of the variance and expectation value we arrive at:

$$\Delta = \frac{1}{N}\frac{1-\left|R_{IRF}\right|^2}{\left|R_{IRF}\right|^2} \geq 0 \tag{14}$$

which would be 0 if $|R_{IRF}|^2 = 1$, which would only be the case with a delta-like function. Then, the matrices are not similar. Also, since the trace is the sum of the eigenvalues and $\Delta \geq 0$, we can hint that the deconvolved covariance matrix is bigger than the *unconvolved* one. To be certain, one would have to show that the determinant is bigger too.

As is shown by Monte Carlo simulations (**Sup. Fig. 4**), the width of the IRF produces an increase in the uncertainty in lifetime estimations, in disagreement with what was reported in [**?**].

4. Description of pawFLIM's wavelet algorithm

In pawFLIM, we adapted the translational invariant Haar wavelet basis to 2D complex data with different covariance matrices, and used a global significance level to decide whether to threshold coefficients.

In 1D, a Haar wavelet transform can be thought as sums (or averages) and differences of neighboring pixels x_1 and x_2:

$$a = \frac{x_1 + x_2}{\sqrt{2}} \tag{15}$$

$$d = \frac{x_1 - x_2}{\sqrt{2}} \tag{16}$$

which are called approximation and details coefficients in the wavelet literature. The $\sqrt{2}$ factor is a normalization factor. A denoising may be applied by thresholding detail coefficients (i.e., setting to zero), which is equivalent to hypothesis testing [**?**]. Then, x_1 and x_2 can be reconstructed as:

$$x_1 = \frac{a + d}{\sqrt{2}} \tag{17}$$

$$x_2 = \frac{a - d}{\sqrt{2}} \tag{18}$$

which, in case of having thresholded d, are simply the average $(x_1 + x_2)/2$.

In general, given an array of values $\{x_k\}$, a multiresolution analysis is performed by applying this scheme recursively with the approximation coefficients a, giving a set

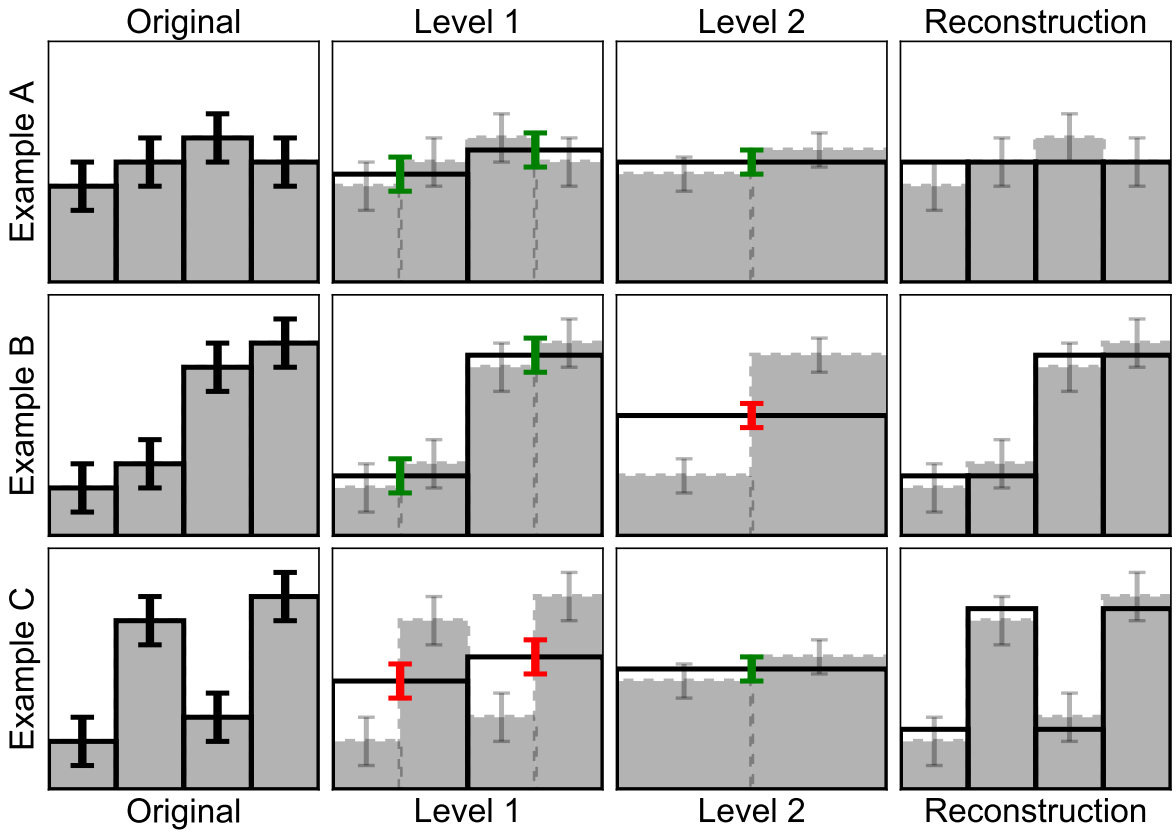

Figure 5. Examples of Haar wavelet thresholding to illustrate non-local averaging (case C). At each level, the average of two neighboring values is considered, and averaged (green) if they are similar or not (red).

of detail coefficients $d_{l,k}$ corresponding to different levels l:

$$a_{l+1,k} = \frac{a_{l,2k} + a_{l,2k+1}}{\sqrt{2}} \tag{19}$$

$$d_{l+1,k} = \frac{a_{l,2k} - a_{l,2k+1}}{\sqrt{2}} \tag{20}$$

where $a_{0,k} = x_k$.

To extend this to data with different variance, we considered taking a weighted average as the approximation coefficient at each level:

$$a = \frac{\dfrac{x_1}{\sigma_1^2} + \dfrac{x_2}{\sigma_2^2}}{\dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_2^2}} = \sigma_a^2 \left(\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2} \right) \tag{21}$$

$$d = x_1 - x_2 \tag{22}$$

For each d, a Z-score is calculated:

$$Z = \frac{d}{\sqrt{\sigma_1^2 + \sigma_2^2}} \tag{23}$$

and thresholded according to a preset significance level.

The extension to vector values is immediate:

$$\mathbf{a} = \mathbf{\Sigma_a} \left(\mathbf{\Sigma_1^{-1}} \mathbf{x_1} + \mathbf{\Sigma_2^{-1}} \mathbf{x_2} \right) \tag{24}$$

$$\mathbf{d} = \mathbf{x_1} - \mathbf{x_2} \tag{25}$$

where $\mathbf{\Sigma}$ are covariance matrices. To resolve whether to keep the coefficient, a χ^2 is calculated:

$$\chi^2_{(k)} = \mathbf{d}^T (\mathbf{\Sigma_1} + \mathbf{\Sigma_2})^{-1} \, \mathbf{d} \tag{26}$$

where k is the number of dimensions.

We have just described the extension of 1D Haar wavelets from scalar data to covariance-weighted vector data. The 2D or multidimensional Haar wavelet is described elsewhere and an extension of those to covariance weighted vector data is analogous.

We're currently using a global significance value to threshold, although there are different alternatives to be considered, such as controlling the false discovery rate [**?**, **?**].