2D-Fluorescence Lifetime Correlation Code: Mathematical Basis, Tutorial and Technical Notes

Premashis Manna

1 Introduction

This document describes the 2D fluorescence lifetime correlation (2D-FLC) code originally written by Toru Kondo (toru.kondo.c8@tohoku.ac.jp) while at the Schlau-Cohen Lab, MIT. This provides step-by-step instruction of how to use the code. It also contains some mathematical basis and technical notes on the technique. For detailed information on data collection and analysis, please visit [1] and the Supporting Information section of the article.

2 Mathematical background

Fluorescence decay ($I(t)$) of a sample measured by a TCSPC set-up contains information about its excited state lifetime. Often times, this decay is exponential in nature. However, the true lifetime signal is convoluted with the instrument response function that takes into account of the finite width of the excitation pulse, along with the temporal broadening due to the electronics involved, monochromator, detectors etc. The convoluted nature of the signal demands that mere subtraction of the IRF would not work but iterative reconvolution is needed to extract the lifetime information [2].

The time domain fluorescence lifetime decay data ($I(t)$) can be expressed in frequency- (k) space as the following way:

$$I(t) = \int_0^\infty a'(k)e^{-kt}dk \tag{1}$$

This is the so-called Laplace transform of the time-domain data. To convert this into lifetime or τ-space, in Eqn 1, we need to replace $k = 1/\tau$. This gives us,

$$
\begin{aligned}
I(t) &= \int_{\infty}^{0} a'(1/\tau) e^{-t/\tau} (-\frac{1}{\tau^2}) d\tau \\
&= \int_{0}^{\infty} \frac{a'(1/\tau)}{\tau^2} e^{-t/\tau} d\tau \\
&= \int_{0}^{\infty} a(\tau) e^{-t/\tau} d\tau
\end{aligned} \tag{2}
$$

where, $a'(1/\tau)/\tau^2 = a(\tau)$.

Inverse Laplace transform of Eqn 2 would provide the distribution of $a(\tau)$. In other words, it converts the fluorescence lifetime data from t-space to τ-space.

$$
\begin{aligned}
a(\tau) &= \int_{\infty}^{0} I(t) e^{t/\tau} dt \\
&\sim \sum_{i=1}^{\infty} I(t_i) e^{t_i/\tau} \\
&\sim \sum_{i=1}^{n} I(t_i) e^{t_i/\tau}
\end{aligned} \tag{3}
$$

In the first step, the integration is replaced with summation for the discretization of the data which is needed for the coding. In the second step, the upper limit of summation is restricted to a finite number, n.

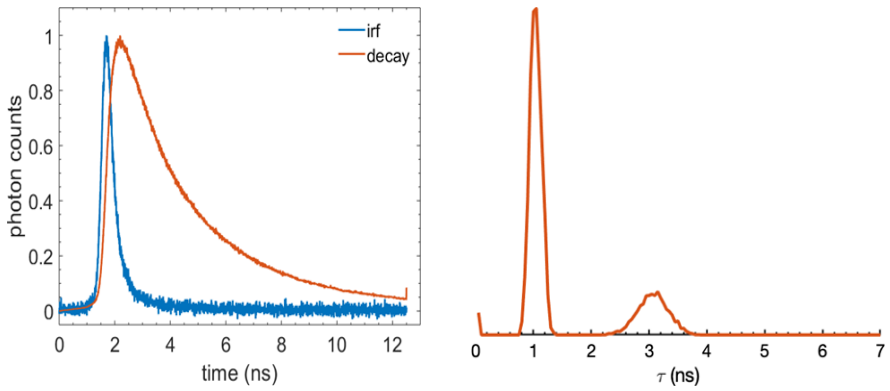
Now, inverse Laplace transform (ILT) is ill-posed which means many solutions exist that can adequately describe the decay behavior with a reasonable χ^2 value ($\chi^2 \sim 1$). Under this condition, noise can abruptly change the solution. Again, Fourier transform (FT) does not work to extract the lifetime distribution for the following reasons – a) due to abrupt nature of the rise of the fluorescence data b) truncation of the decay at long time scale. FT works best with periodic signal but not instantaneous signal decay. To overcome this problem of instability of the solution of ILT, we employ maximum entropy method while minimizing χ^2. In practice, instead of minimizing χ^2, we minimize Q which is given by:

$$
Q = \chi^2 - \frac{2S}{\eta} \tag{4}
$$

where, S is the Shanon-Jones entropy and expressed as:

$$
S = \sum_{i=1}^{n} a_i(\tau_i) - m_i - log\frac{\alpha_i(\tau_i)}{m_i} \tag{5}
$$

m_i in the above equation is the initial distribution of lifetime. In our code, we take it as a flat distribution of lifetime.



(a) (b)

Figure 1: (a) Experimentally determined instrument response function (irf) and a simulated lifetime decay trace having lifetime of 1 ns and 3 ns. (b) Lifetime distribution obtained by 1D-inverse Laplace transform of the time-domain data.

We minimize χ^2 in non-linear least square fits. But here, for MEM, we also want to maximize the Shanon-Jones entropy (S) to zoom in on some set of solutions which are uncorrelated to each other [3]. The regularization constant (η) balance between the optimization of χ^2 and S. In the code, we do it sequentially – first we make the η lower to give more importance to entropy maximization. Then slowly, we increase the value of η and this gives more priority to the fitting of data i.e. χ^2. This is to ensure that the solutions does not get trapped into some local minima rather than converging to a global minima. The negative sign in the Eqn 4 reflects that entropy is getting maximized while we minimize Q.

3 Tutorial

For running the 2D-FLC code, the user have to go through these following processes. Files needed to run these codes are fluorescence decay data (simulated or real) and instrument response function (IRF).

3.1 Generating Simulated Data

Use the code TK-MyMain-Simu-PhotonStream.m to generate simulated data. Before running the actual data, it is advisable to test the code with some simulated data. In the code, one can specify the number of lifetime components, their relative population and the corresponding inter-conversion rates.

3.2 1D Search

First, one have to run 1D-ILT code to convert the time-domain decay data to tau-domain data as shown in Figure 1. As the iterative reconvolution is performed on the data set, we have observed that the resulting lifetime distribution ($a(\tau)$) is very much sensitive towards the position or shift of IRF. Therefore, the IRF position is systematically varied to minimize the χ^2 (Figure 2). Once we have a full scan of IRF positions and the corresponding lifetime distribution, four of five distributions having low χ^2 values are averaged to obtain a mean lifetime distribution. This mean distribution is then used in the next step of the processes. We call this procedure as **1D Search**. Here is the step-by-step protocol for running **1D Search** part of the code. More documentation can be found in the associated matlab files.

1. Preparation
   - remove baseline of IRF, and normalize it before analysis.
2. Use TK-MyMain-Create2DFDC-cor-SeparateData-BootStrap-v02.m and set the following:
   - set "BootStrap-PhotonNum-factor"
   - set "BootStrap-GroupNum-factor"
   - set "DivideNum"
   - set "DataFolderPath"
   - set "FileName-DataList"
   - set "tt1-Step"
3. Use TK-MyMain-Fit-1DMEM-02.m and
   - set "y0"
4. Use TK-MyMain-Search-RiseIRF-1DMEM.m and
   - set "LoadFileName-IRF"

3.3 2D Search

After running the **1D Search**, the lifetime distribution obtained is used as input for **2D Search**. In this part of the code, first, 2d fluorescence lifetime correlation map ($M(\Delta T, t', t")$)is constructed as described in [4, 1]. Then inverse Laplace transform is applied in 2D lifetime decay data to extract $\tilde{M}(\Delta T, \tau', \tau")$ (Figure 3). Here also different IRF positions are scanned to obtain a set of optimal IRF position that gives minimal χ^2 for the solution.

Purpose of running the **2D Search** is to obtain the optimal IRF position. Therefore, here, ΔT is just set to one value. Also, the regulator constant also set to some low number, usually 100.

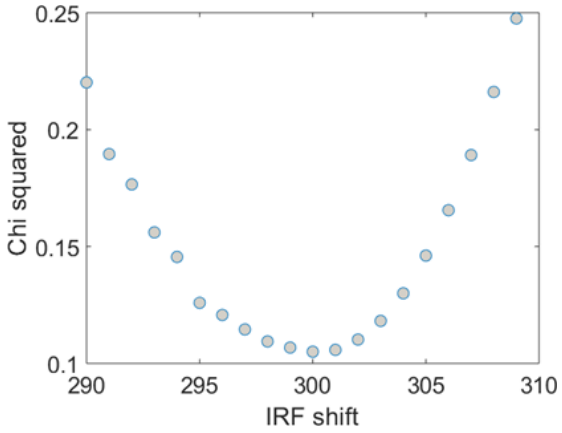
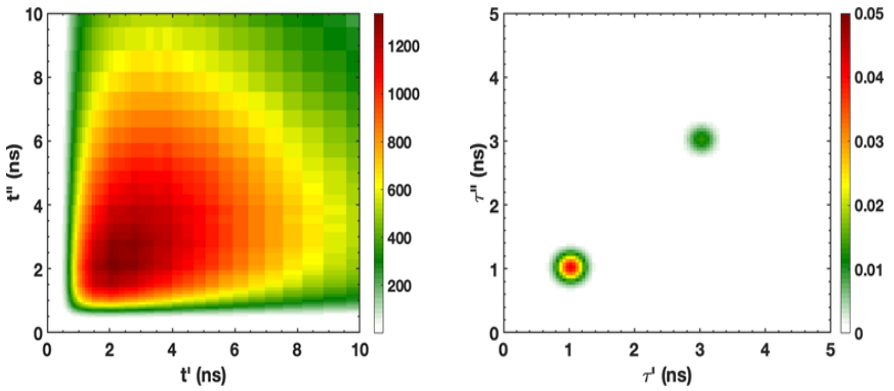


Figure 2: Position of IRF is systematically varied to minimize the χ^2.

1. Use TK-MyMain-Create2DFDC-cor-SeparateData-BootStrap-v02.m and
   - set "BootStrap-PhotonNum-factor"
   - set "BootStrap-GroupNum-factor"
   - set "DivideNum"
   - set "DataFolderPath"
   - set "FileName-DataList"
   - set "tt1-Step"
2. Use TK-MyMain-Fit-2DMEM-04.m and
   - set "y0" and "Fix1orNot0orAbs2-y0"
3. Use TK-MyMain-GFit-2DMEM.m and
   - set "G-TrialNumFor-RegulatorConst" (set around 100)
4. Use TK-MyMain-Search-RiseIRF-2DMEM.m and set
   - set "LoadFileName-IRF", "Hozzon-estimates", and "RisePoint-IRF-start"
     (Set the "Hozzon-estimates", *i.e.* the number of lifetime states and their initial values as obtained from **1D Search**).

3.4 Average 2D

Now once we have identified optimal IRF positions for this data set, we like to perform the final run to obtain 2D fluorescence lifetime correlation. Use the 5 good IRF positions (associated with lowest χ^2) that we obtained earlier in **2D Search**. This time, increase the trial number for regulator constant to ~ 300.



(a) (b)

Figure 3: (a) 2D fluorescence lifetime map ($M(\Delta T, t', t")$) constructed as a function of ΔT.(b) This map is inverse Laplace transformed to obtain $M(\Delta T, \tau', \tau")$.

1. TK-MyMain-Run-Ave2DMEM.m
   - set "LoadFileName-IRF"
   - set "dT" and "ddT"
   - set "Tau-Select"
   - If "Tau-Select" = 2, set "Hozzon-estimates"
   - If "Tau-Select" = 2, set "LoadFileName-Stored-Mat-A"
2. Use TK-MyMain-Create2DFDC-cor-SeparateData-BootStrap-v02.m and
   - set "BootStrap-PhotonNum-factor"
   - set "BootStrap-GroupNum-factor"
   - set "DivideNum"
   - set "DataFolderPath" and "DataListFromExternalFile"
   - If ""DataListFromExternalFile"" = 1, data listed in "FileName-DataList" are loaded.
   - set "tt1-Step"
3. Use TK-MyMain-GFit-2DMEM.m and
   - set "G-TrialNumFor-RegulatorConst" (set around 300)
4. Use TK-MyMain-Fit-2DMEM-04.m and
   - set "TrialNumFor-RegulatorConst" and "y0"
5. Finally, run TK-MyMain-Run-Ave2DMEM.m

- set
  Var-RisePoint-Imax = 5 ; % number of run to average
  Var-RisePoint-Icenter = 303 ; % center of the rise point
  Tstart = 11 ;
  Tend = 311 ;
  SaveFileName = 'TimeSplit4-LHCII-pH7-Remove-i005-cor0mi0-TTT1' ;
  Hozzon-estimates = [0,... % estimates from the 1D search
  1, 1, 0.2,...
  1, 3, 0.2] ;

Results of 2D-FLC analysis (after Global fit for all ΔTs):
Gresult-EstQ-Global = fitting indicator = [RegulatorConst, EstQ, Kai2, EntropyS]
Gresult-Mat-2DFDC-lin-t = time t in linear
Gresult-Mat-2DFDC-log-t = time t in log
Gresult-Mat-A-Global = Matrix A (lifetime distribution) = [distribution along Tau, state number]
Gresult-Mat-G-dTfun = Matrix G (correlation function) = [state num, state num, ΔT index]
Gresult-Mat-M-2DFLC-dTfun = Matrix M (2D-FLC) = [Tau index, Tau index, ΔT index]
Gresult-Mat-M-Model-lin-dTfun = Matrix M (2D-FD) = [t index, t index, ΔT index], where t is in linear
Gresult-Mat-M-Model-log-dTfun = Matrix M (2D-FD) = [t index, t index, ΔT index], where t is in log
Tau = lifetime = [Tau index]
dT = ΔTs = [ΔT index]
MeasurementTime-s = measurement time in second for each molecule

Input Data (produced from TCSPC photon stream data)
Mat-2DFDC-lin = Matrix M (2D-FD) = [t index, t index, ΔT index], where t is in linear
Mat-2DFDC-log = Matrix M (2D-FD) = [t index, t index, ΔT index], where t is in log.

3.5 Correlation Fitting

This part of the code fits the lifetime auto- and cross-correlation as obtained from the previous section.

1. Use TK-MyMain-Analyze-03-NotRatio.m and
   ⇒ set parameters

(A) Ana-estimates-Select = 1 or 0 Initial fitting parameters are set to be input values (0) or values stored in "Ana-estimates".

(B) TrialNumber = trial number for the fitting

(C) FitNormalizedCor-y1n0 = 1 or 0 or nm

The correlation fitting is performed for normalized curves, i.e. fitting weight is equal for all correlation curves (1), or for absolute curves, i.e. not normalized (0). If set to be (nm), only correlation between state n and m is weighted by a factor given by "FitWeightFactor".

(D) ComponentNumber = a number of dynamic components

(E) StateAssign = assignment of lifetime state, defining the order of state for rate matrix (e.g. [1, 2, 2] means that 1st low/column in rate matrix is related to state 1 while 2nd and 3rd are related to state 2).

(F) FitA = a scale factor of correlation, which should be 1.

(G) FitE = optical extinction coefficient, which should be 1 in the present study = [component num index, lifetime state num index].

(H) FitQ = fluorescence quantum yield, which should be 1 in the present study = [component num index, lifetime state num index].

(I) FitRateM = dynamic rate matrix = [lifetime state num index, lifetime state num index, component num index].

(J) Fity0 = just base line, which should be 0 in the present study = [lifetime state num index, lifetime state num index].

(K) Parameters labeled by "Free0orFix1orAbs2——-" = 0 or 1 or 2 or $n > 2$.

The fitting parameters are set as a free parameter (0), fixed to be the input value (1), fixed to be an absolute/positive value (2).

If set to be (n), the parameter is set as a global value and linked with other parameters set to be the same number n. e.g. If Free0orFix1orAbs2-Q = [3, 1 ; 2, 3]; intensity of component 1 at state 2 (set to be 1) is fixed, intensity of component 2 at state 1 (set to be 2) is absolute, and two intensities of component 1 at state 1 and component 2 at state 2 (both set to be 3) are linked with each other and fitted as a one global value.

2. Start run with TK-MyMain-CorFit-NotRatio.m

   Output ⇒ "Ana-estimates", "Ana-estimates-y0", "Ana-EquPop-V-Substate"

   "Ana-estimates" contain...

   Low 1: "StateAssign"

   Low 2: "Ana-estimates-A" = result for "FitA"

   Low 3: "Ana-estimates-E" = result for "FitE"

   Low 4: "Ana-estimates-Q" = result for "FitQ"

   After Low 5: "Ana-estimates-RateM" = result for "FitRateM"

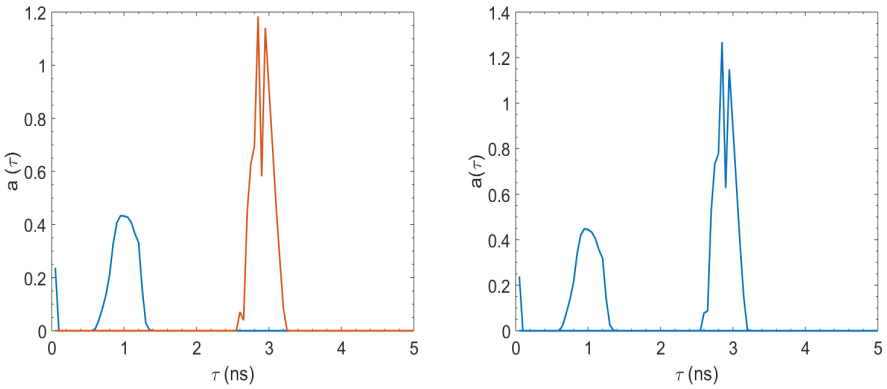
"Ana-estimates-y0" = result for "Fity0"

"Ana-EquPop-V-Substate" = population of each state = [lifetime state index, component index]

3. Display figure with TK-MyMain-Fig.m
   - set "DisplayAll0orMap1orCor2"
4. Display correlation figure with TK-MyMain-Fig-Cor.m

4 Technical Notes

1. Increasing the regulator constant trial number (G-TrialNumFor-RegulatorCons) from 300 to 400 did not change the lifetime distribution significantly. Although it increased the computation time from ∼7000 s to ∼12000 s. Therefore, ∼ 300 might be a good number for it.



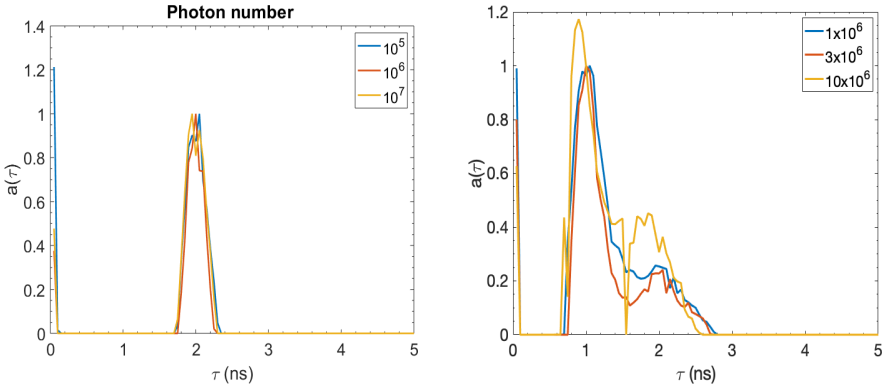
(a) (b)

Figure 4: Lifetime distribution of the simulated data having equal 1 ns and 3 ns population. (Left) with regularization constant trial number 300 and (right) 400.

2. Maximum entropy method (MEM) does not need any specific model to fit the data. The initial distribution is considered to be flat here (m_i, Eqn 5). Discretization of the data is preferred in log(τ) space which makes the initial choice of flat distribution physically meaningful [5].
3. Each resolved lifetime component is associated with a width which in principle could be related to the heterogeneity of the system. However, the width is also dependent on the photon number which makes the interpretation complicated. Lifetime distribution width cannot quantify the real heterogeneity of the lifetimes unless one quantifies inherent noise in the decay to the width. Noise in the data increases the width of the lifetime distribution and therefore peak resolution decreases [6].

4. A good fit criterion of $\chi^2 \sim 1$ could be obtained for many different distribution of $\alpha(\tau)$. The optimum distribution is the one which fits the data adequately ($\chi^2 \sim 1$) and maximizes the value of Shanon-Janes entropy function S. Given multiple distribution reasonably satisfy the distribution, MEM optimization ensures the distribution with minimum number of peaks (less structured) and maximum width for each peak. This is to ensure that that data is not over-interpreted.

5. Complete decay and number of iterations is important for proper construction of lifetime distribution. Width of the distribution increases in case of lower number of photons.



(a) (b)

Figure 5: The width of lifetime distribution as a function of photon number for a system with a single (a) 2 ns lifetime and (b) having two populations with 1 ns and 2 ns lifetimes.

6. According to Swaminathan et al. [5] signal-to-noise ratio controls the width of lifetime distribution. Figure 5a shows the lifetime distribution of single-exponential decays with different photon number. With the increase of photon number from 10^5 to 10^6, the width of the distribution decreases little bit and the background peak gets diminished. But from 10^6 to 10^7, the width of the peak did not change appreciably. Actually, it broadened a little bit.

   For a system with multiple lifetime components, with the increase of photon number, the peaks tend to resolve well (Figure 5b). However, the longer lifetime component is always broadened probably due to the truncation of the data at 12.5 ns which is the time-window for our 80 MHz sample excitation.

References

[1] Toru Kondo, Jesse B Gordon, Alberta Pinnola, Luca Dall'Osto, Roberto Bassi, and Gabriela S Schlau-Cohen. Microsecond and millisecond dynamics in the photosynthetic protein lhcsr1 observed by single-molecule correlation spectroscopy. *Proceedings of the National Academy of Sciences*, 116(23):11247–11252, 2019.

[2] Darren A Smith, Grant McKenzie, Anita C Jones, and Trevor A Smith. Analysis of time-correlated single photon counting data: a comparative evaluation of deterministic and probabilistic approaches. *Methods and applications in fluorescence*, 5(4):042001, 2017.

[3] AK Livesey and JC Brochon. Analyzing the distribution of decay constants in pulse-fluorimetry using the maximum entropy method. *Biophysical journal*, 52(5):693–706, 1987.

[4] Kunihiko Ishii and Tahei Tahara. Two-dimensional fluorescence lifetime correlation spectroscopy. 2. application. *The Journal of Physical Chemistry B*, 117(39):11423–11432, 2013.

[5] R Swaminathan and N Periasamy. Analysis of fluorescence decay by the maximum entropy method: influence of noise and analysis parameters on the width of the distribution of lifetimes. In *Proceedings of the Indian Academy of Sciences-Chemical Sciences*, volume 108, page 39. Springer, 1996.

[6] Jeremy M Shaver and Linda B McGown. Maximum entropy method for frequency domain fluorescence lifetime analysis. 1. effects of frequency range and random noise. *Analytical chemistry*, 68(1):9–17, 1996.