

Fig. S1: Study of the histogram of the distribution of values $\{i/N; 0 \leq i \leq N; 1 \leq N \leq 500\}$ represented in Fig. 2. A. Logarithm of the posterior probability (LPP) of bin number N (as defined in ref. ³⁶). The maximum is obtained for $N = 1$. B: Histogram of the distribution with 1, 10 and 100 bins. The distribution is uniform at this binning level. C: Histogram of the distribution with 1,000 bins: the discrete nature of the distribution (as represented in Fig. 2) starts to appear at such a binning resolution.

Fig. S2: Simplified Jablonski diagram for a FRET pair. Each molecule is defined by a ground state S_0 , an excited state S_1 and a triplet state T_1 , and the corresponding transition rates k_e , k_r , k_{ISC} and k_{ph} . In the absence of alternating excitation, only the donor molecule D is excited by a laser, with rate k_e . Each molecule can bleach when it is in its excited state, with a rate k_{bl} , and the donor can non-radiatively transfer its energy to the acceptor molecule with a rate k_{FRET} .

Fig. S3: Influence of dithering and averaging of the LPP on the optimal number of bin. A: the Log Posterior Probability (LPP) calculated on the original data (black curve) does not exhibit any maximum. Adding a small amount of noise (dithering) of amplitude 0.01 to the data suffices to yield a LPP with a maximum, although the curve remains noisy (not shown). Averaging the LPP obtained for 10,000 different dithered data sets result in a smooth LPP curve (red curve) with a clearly defined maximum. B: Detail of the LPP curves around their maximum. Each curve was obtained as the average over $R = 1$ to 10,000 LPP's of randomly dithered data sets. The position of the maximum ($M = 37$) does not vary for $R > 100$. C-D: PRH for the original and one dithered data set, for $M = 100$ (gray box histogram) and the optimal bin number $M = 37$ (black histogram). In both cases, the PRH build with the optimal bin number has lost its spikes and voids. In addition, the figure indicates that dithering results in a significant smoothing of the PRH constructed with 100 bins. E: An alternative way to define the maximum of the LPP (defined only for integer numbers, and still "noisy" even after significant averaging, see black curve) consists in fitting the LPP with a sum of exponentials (red curve: 4 exponentials). The location of the maximum of this smooth curve ($M = 33$) remains close to that of the averaged LPP ($M = 37$) in all studied cases (see Fig. S4 & S5 for further examples).

Fig. S4: Same as Fig. S3 for Sample 2 (PR ~ 0.7).

Fig. S5: Same as Fig. S3 for the combination Sample 1 + 2.

Fig. S6: Effect of the dithering amplitude on the optimal bin number. 3 cases are studied successively: Sample 1 (A-B: PR ~ 0.35), Sample 2 (C-D: PR ~ 0.7) and the combination of both (E-F: Sample 1 + 2). Without dithering, the Log Posterior Probability (LPP: gray curve in A, C, E) is very noisy and the definition of a maximum is impossible. While increasing the dithering amplitude, the averaged LPP over 1,000 trials becomes smoother and eventually exhibits a clear maximum. The value of M at this maximum (B, D, F) rapidly decreases when the dithering amplitude increases, to finally reach an asymptotic value.

Fig. S7: Effect of dithering on the optimal bin number. A-I: The dithering amplitude applied to the original data set (A) was increased in 0.01 steps from 0.01 to 0.1. For each amplitude, one dithered data set is histogrammed with 100 bins (red curve), and the optimal number of bins for this dithering amplitude. As the dithering amplitude increases, the spikes and voids of the 100 bin histograms are progressively smoothed out. The optimally binned histograms are devoid of them in all cases.

Fig. S8: Effect of dithering on the width of the optimally binned histogram. A: Superimposition of optimal histograms of the dithered data set on the original data set histogrammed with 100 bins (gray box histogram). The dithering amplitude was varied from 0.01 to 0.1, resulting in a slight widening of the histogram. B: Variation of the standard deviation (\sim width) of the histograms as a function of dithering amplitude.

Fig. S9: Comparison between $M = 100$ bin histograms and optimal binning histograms. A: Sample 1, B: Sample 2, C: Sample 1+2. The optimal bin number of the sum of both data sets is almost identical to the maximum of that of each separate data set as expected. The spikes and voids visible with $M = 100$ are eliminated, while the presence of two separate peaks is clearly visible in the combined data set.

Fig. S10: Background count distributions. A: Joint distribution of sizes and durations of 121,541 simulated bursts. B: Burst size distribution (BSD) without constraint on their duration (black histogram) and burst duration distribution (BDD) without constraint on their size (red histogram). These histograms show that the most representative bursts have duration of 1 to 2.5 ms and burst sizes between 50 and 100 counts. C: Comparison between the experimental distribution of background counts in bursts of size S and duration τ (gray box histograms) and best fit Poisson distributions (black histograms). The agreement is quite good between the two whenever the population of bursts is large enough, justifying the use of Poisson distributions in Eq. (36). Discrepancies existing for distributions with very few bursts would have a negligible effect on the final result due to their small weight in the sum of the right-hand-side of Eq. (36).

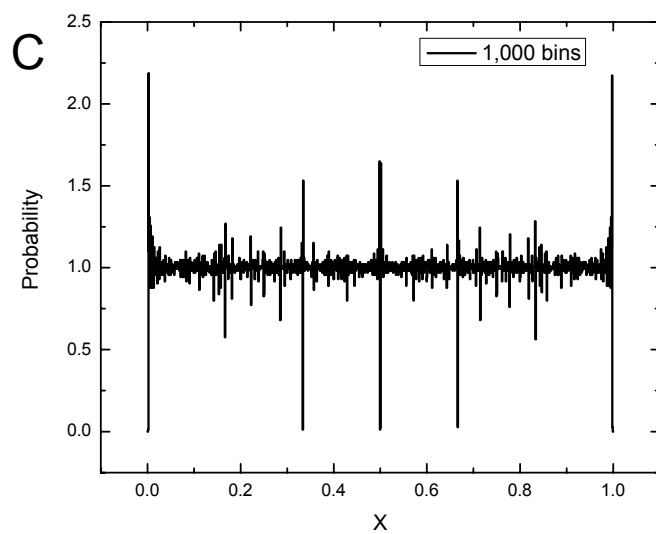
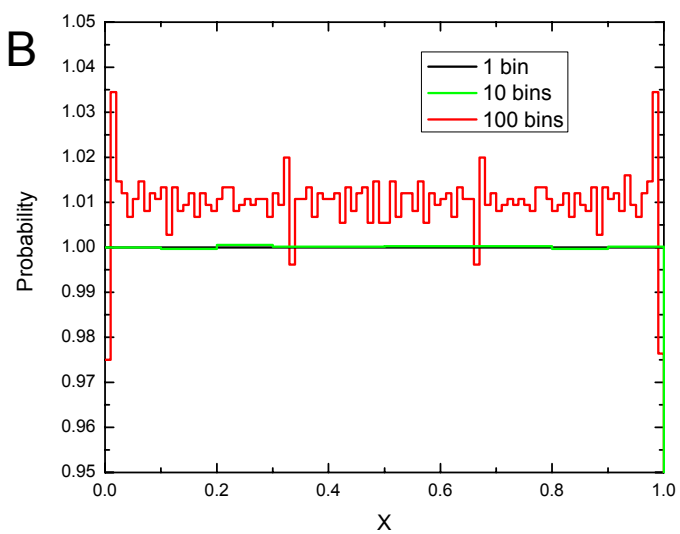
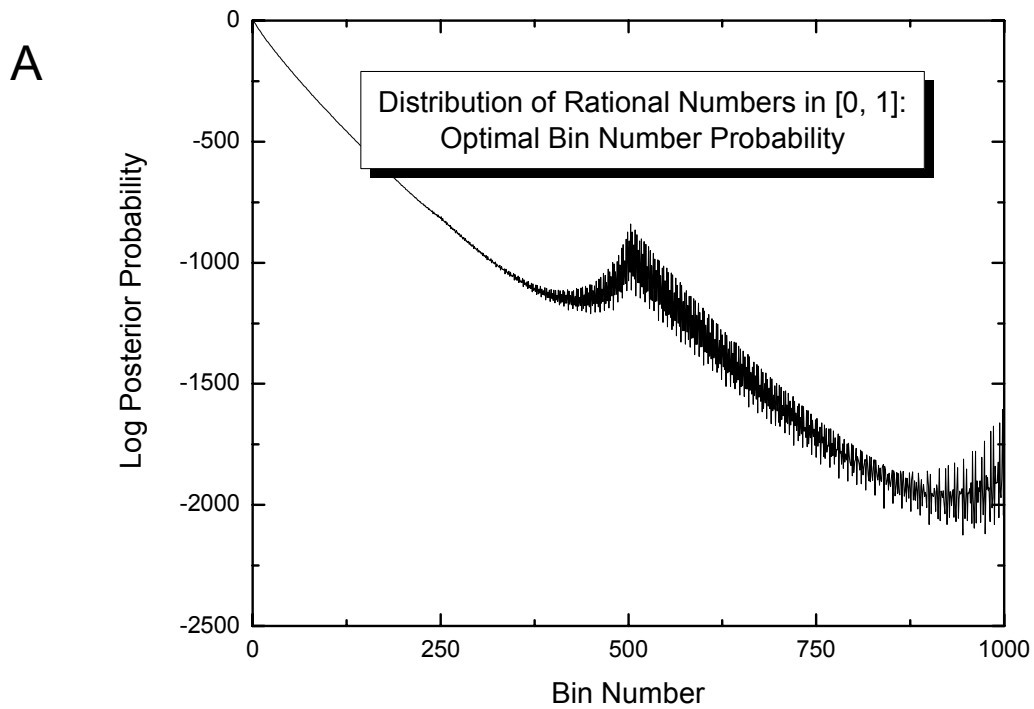


Fig. S1

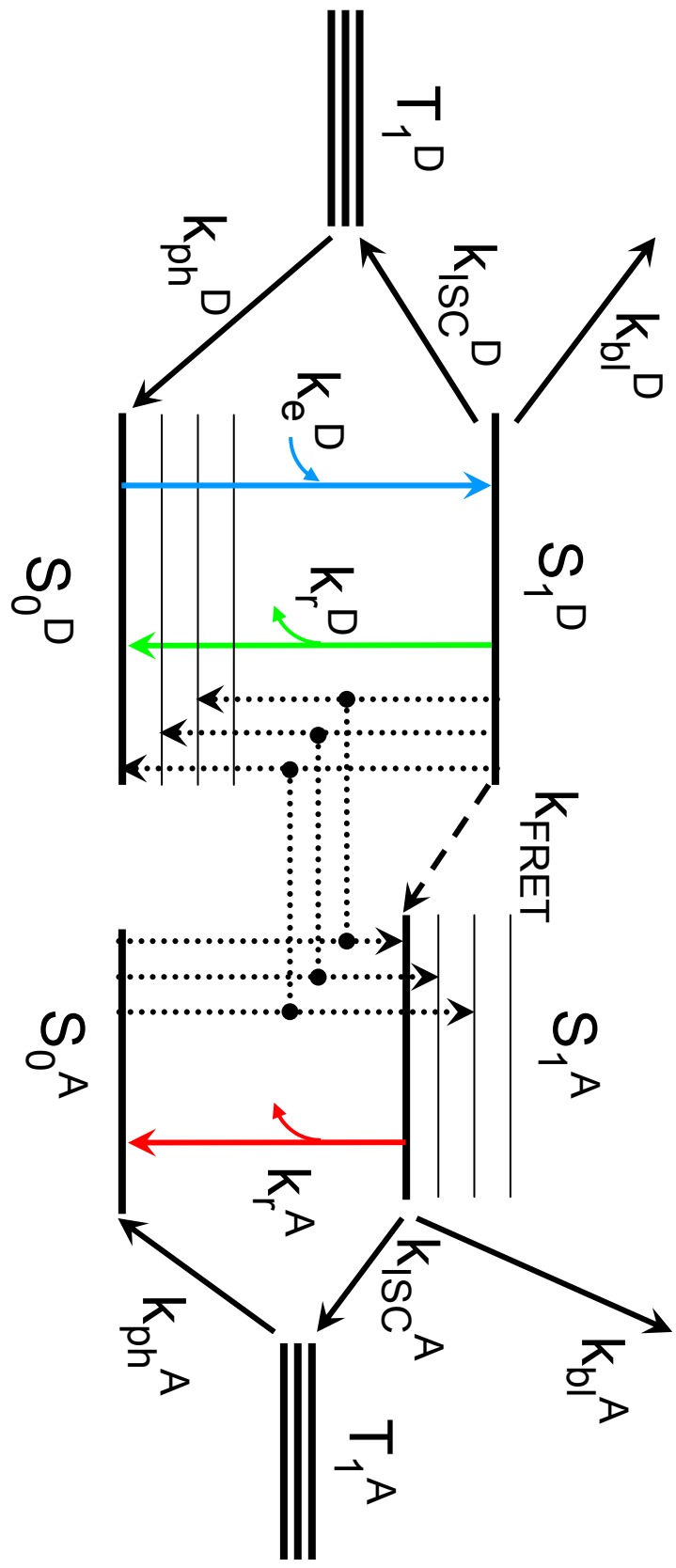


Fig. S2

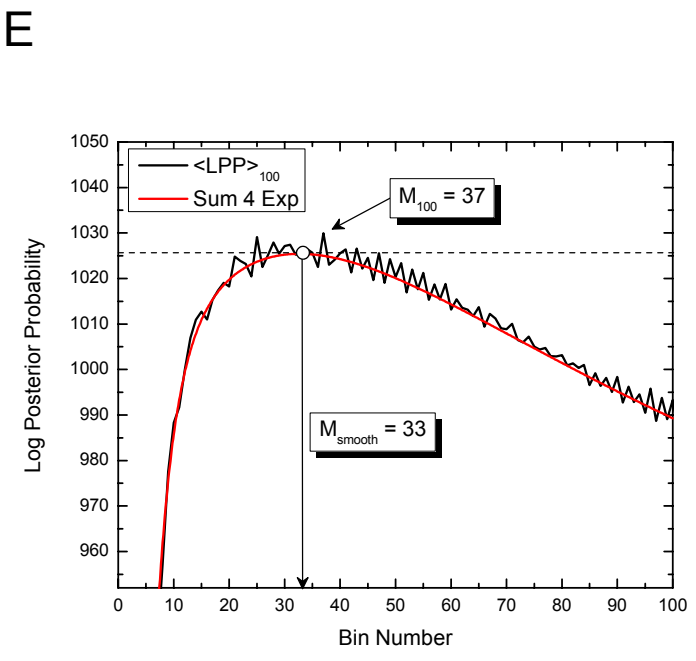
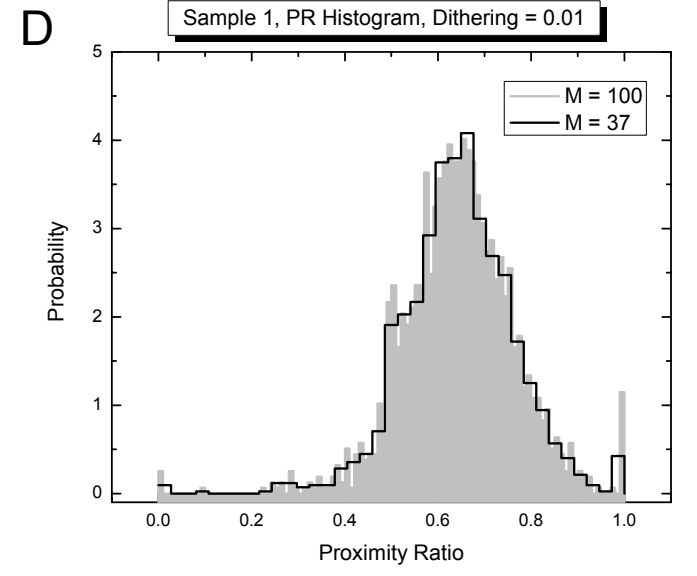
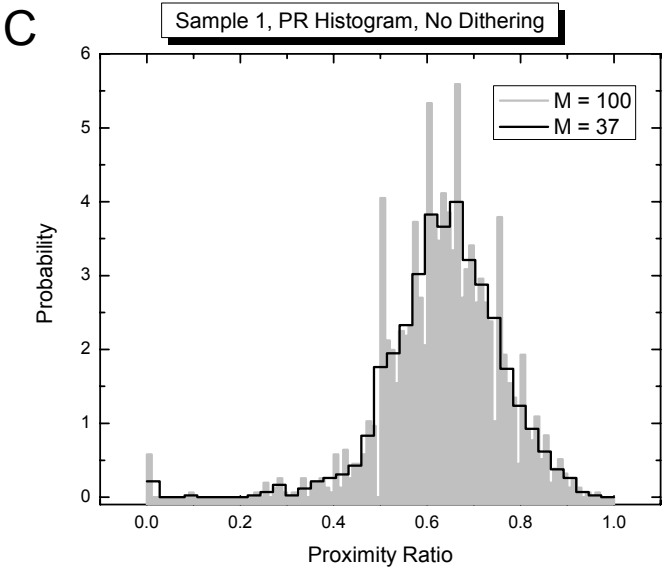
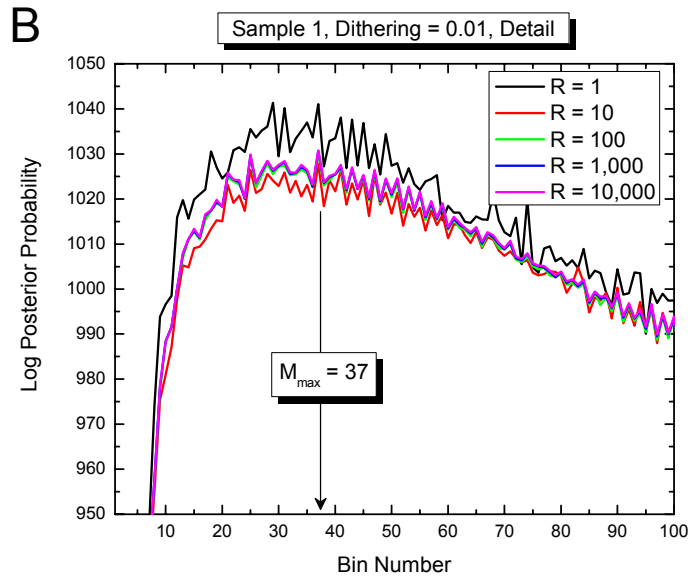
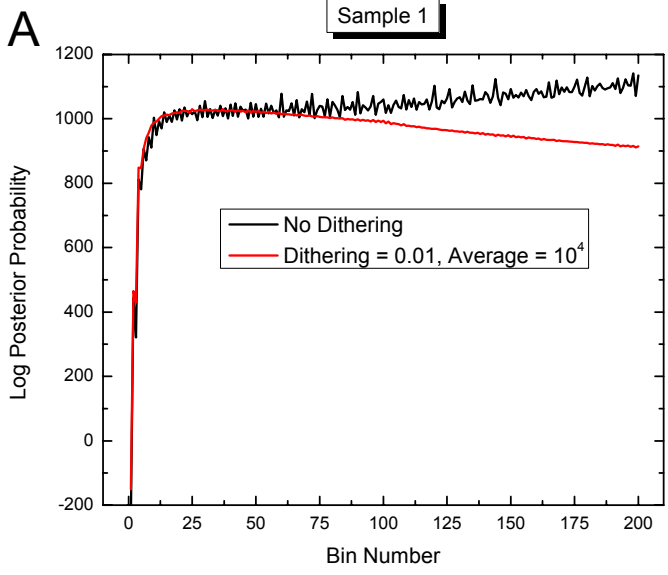


Fig. S3

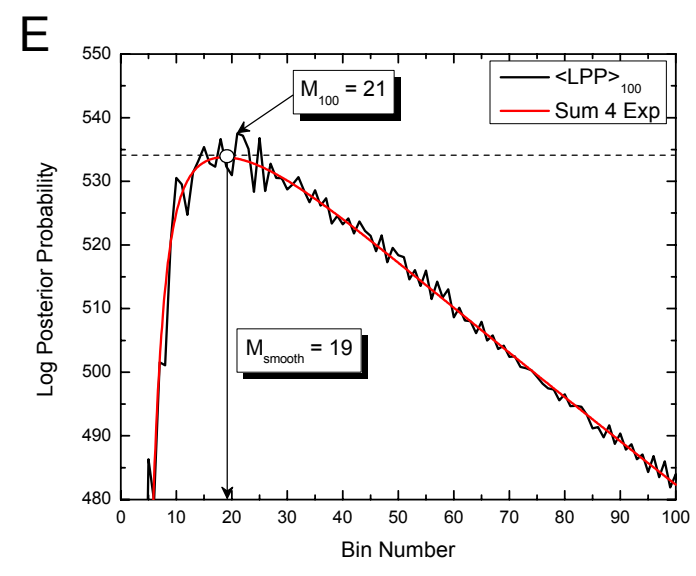
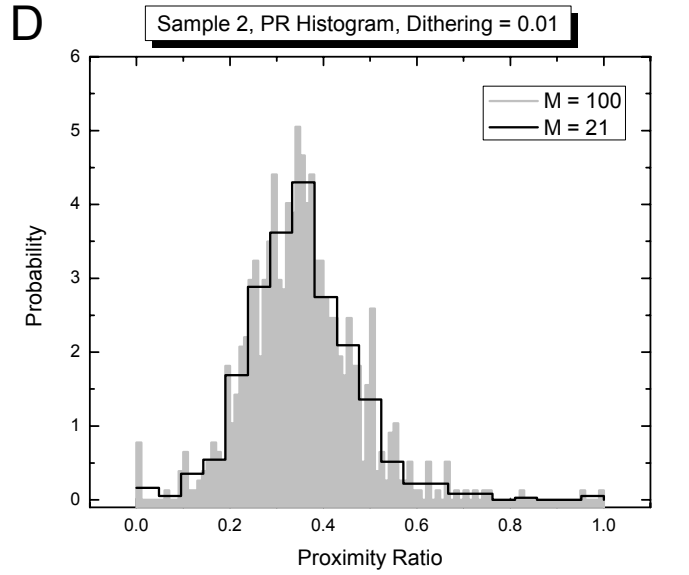
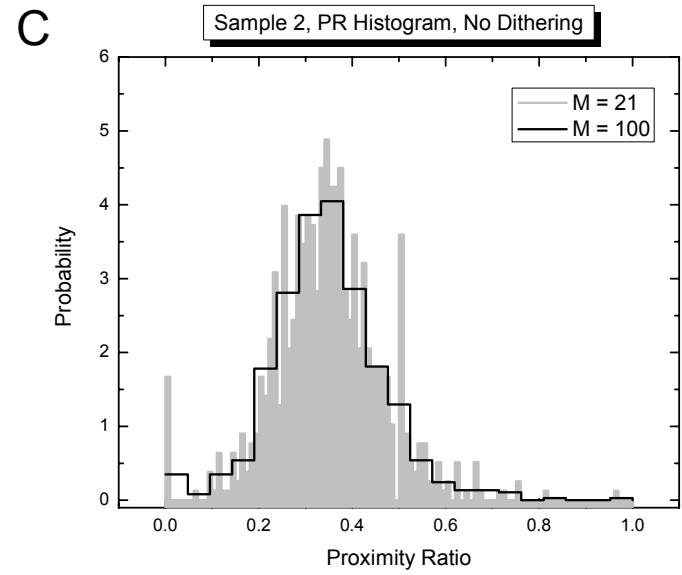
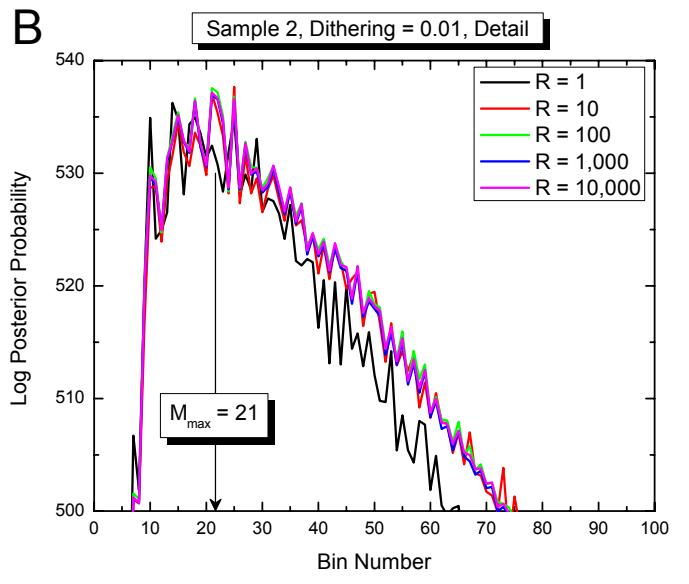
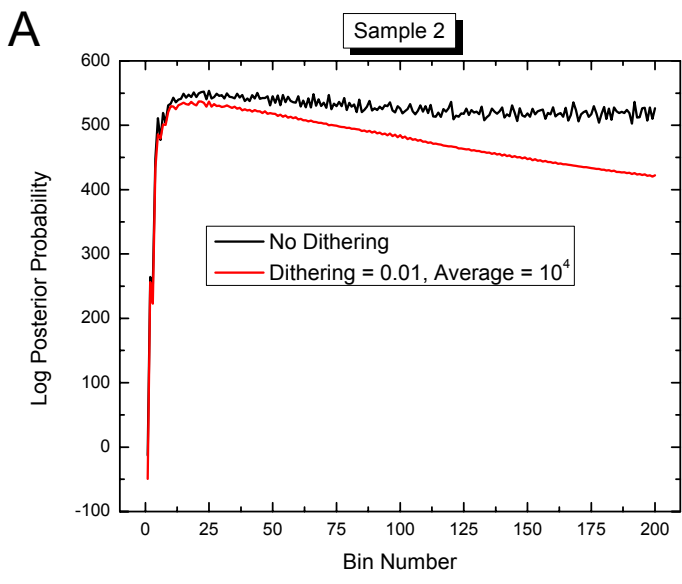


Fig. S4

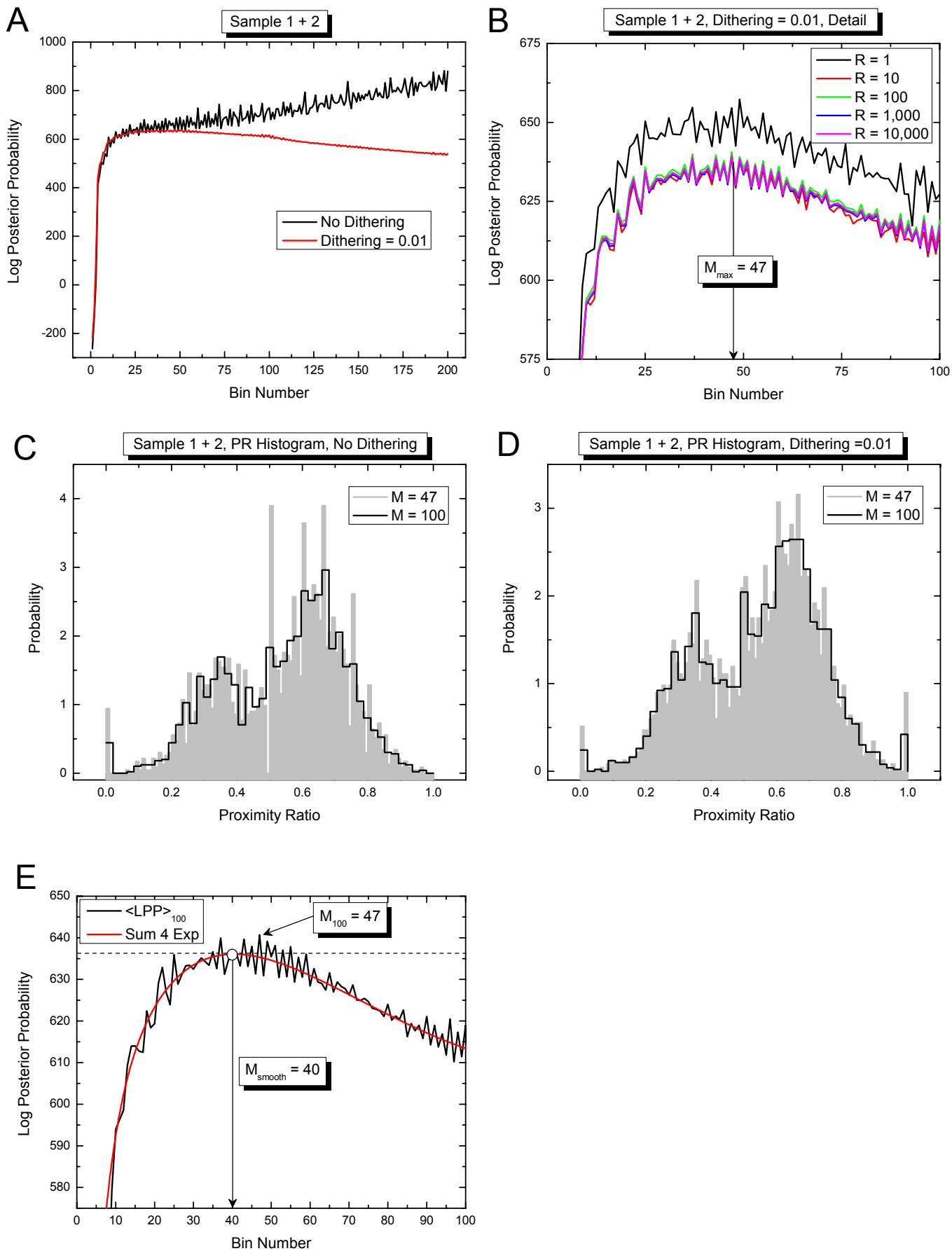


Fig. S5

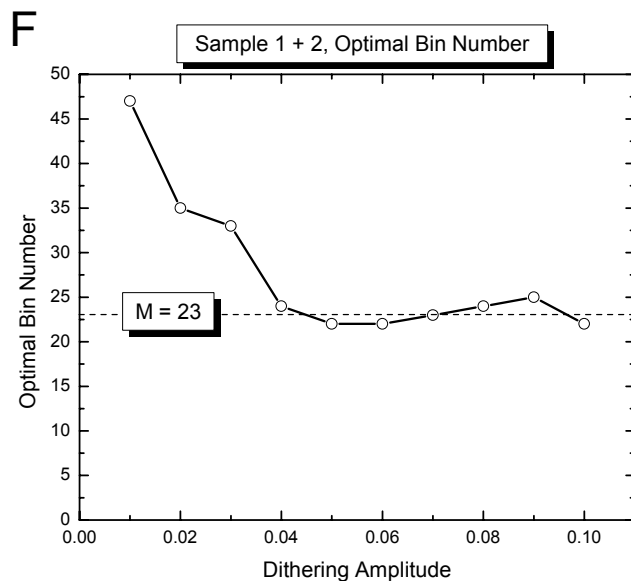
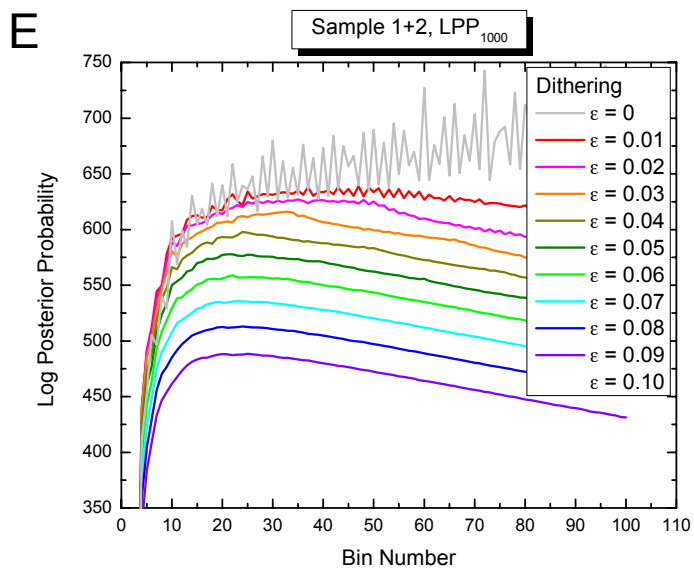
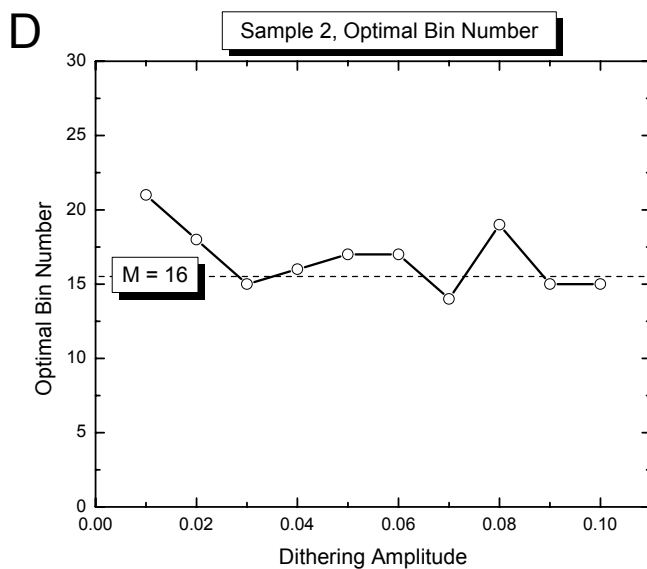
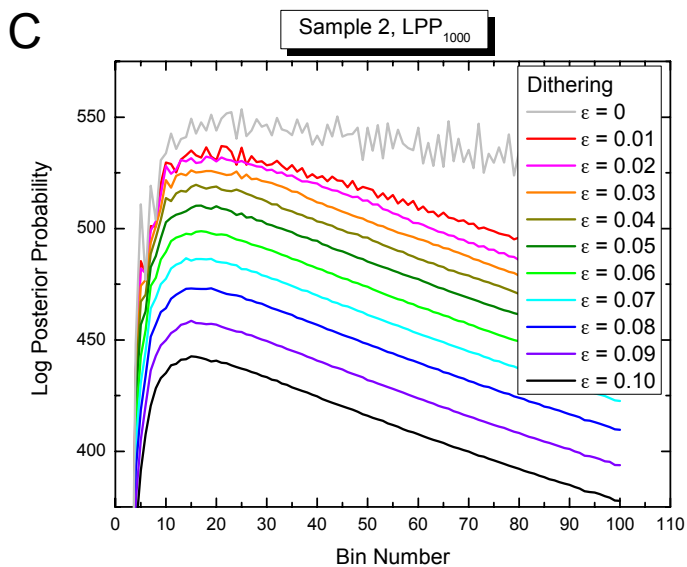
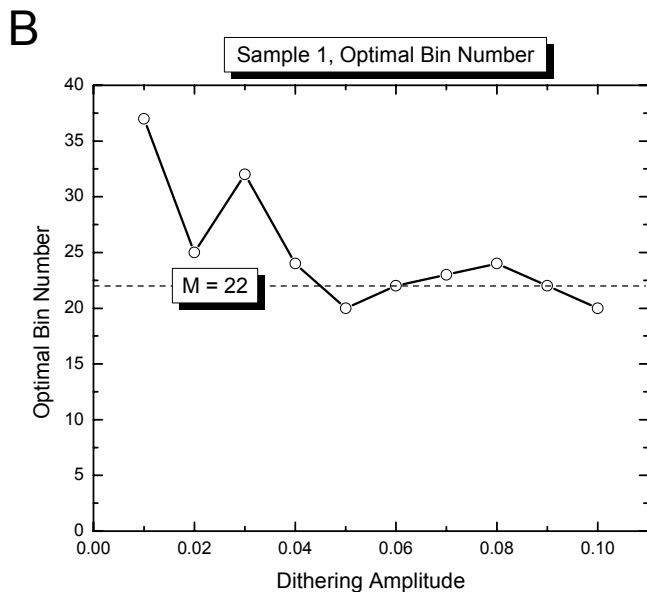
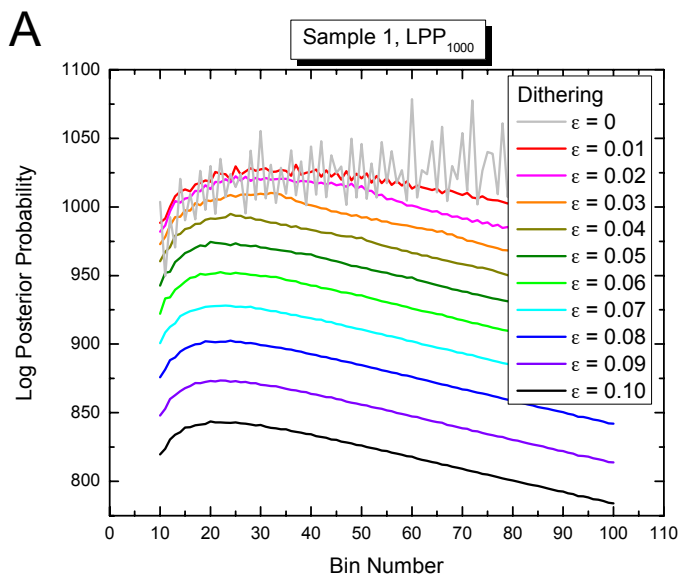


Fig. S6

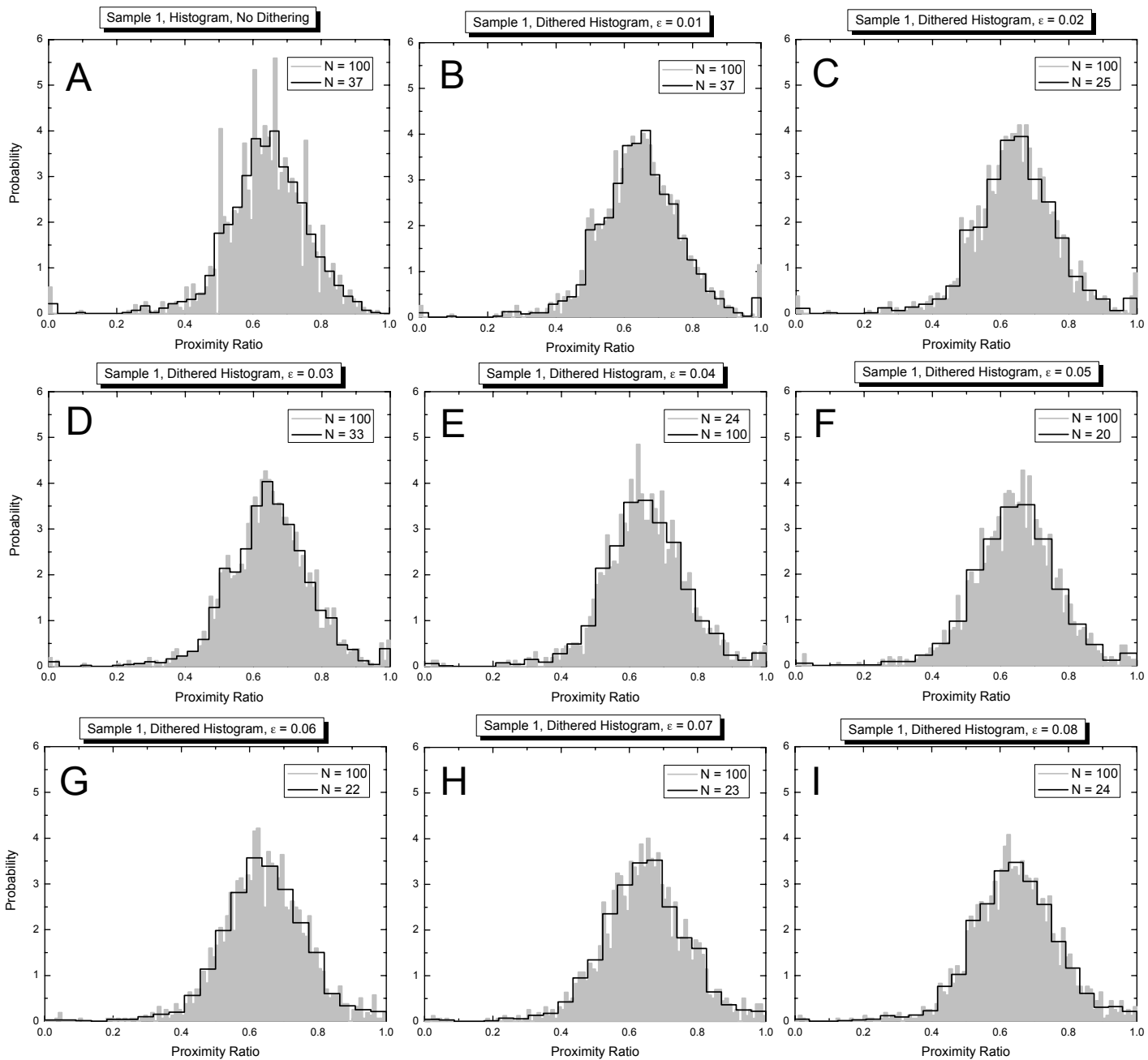
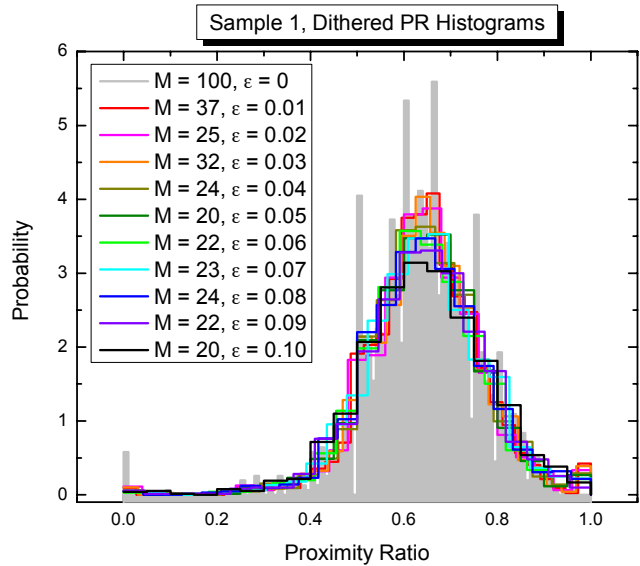
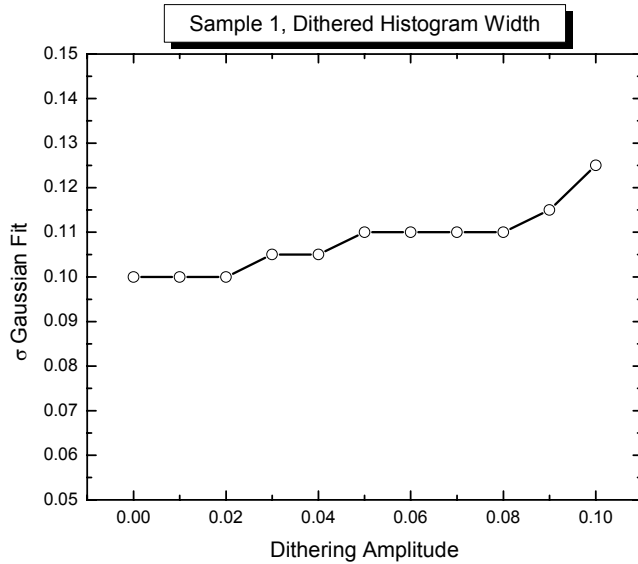


Fig. S7

A**B**

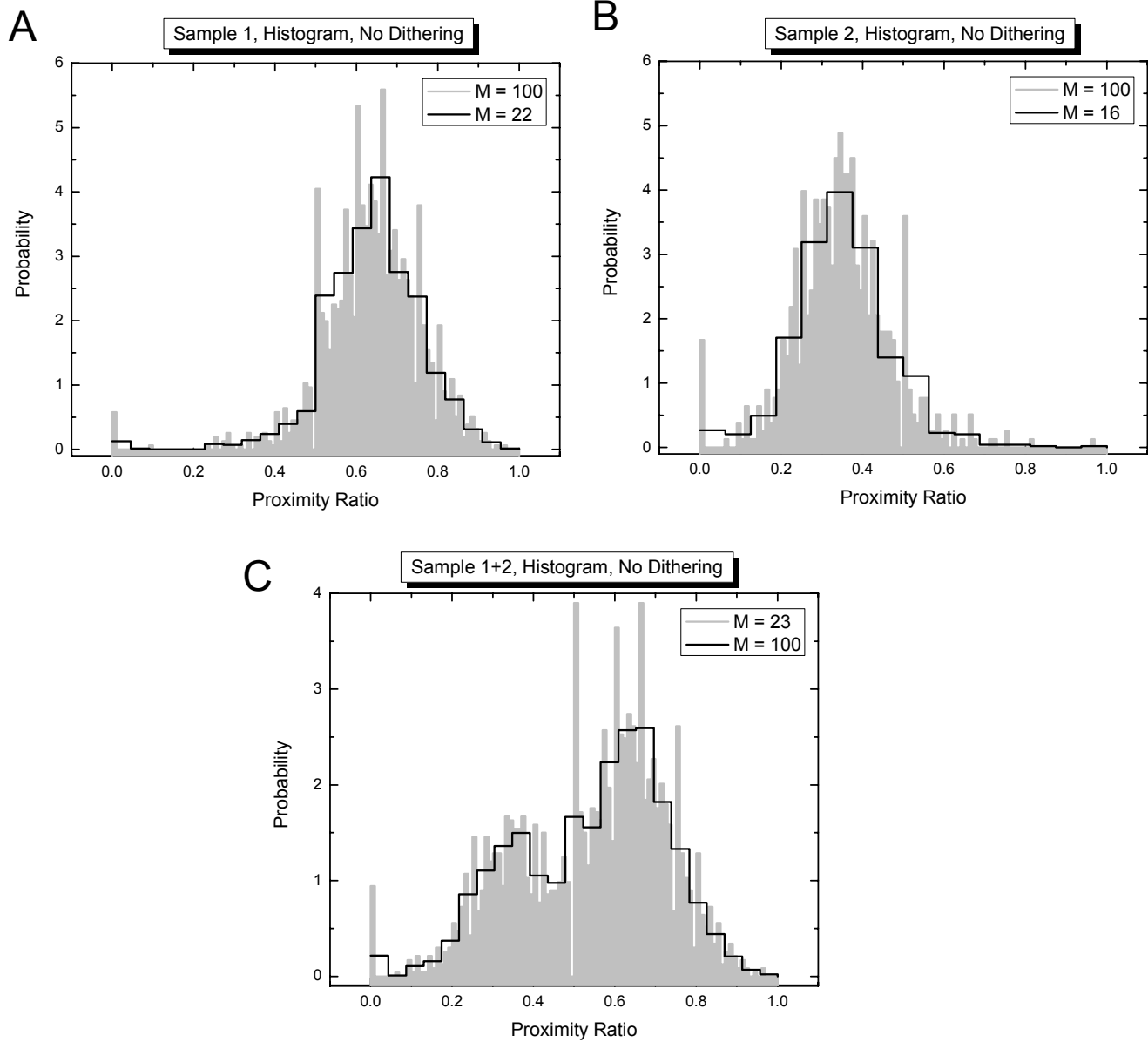


Fig. S9

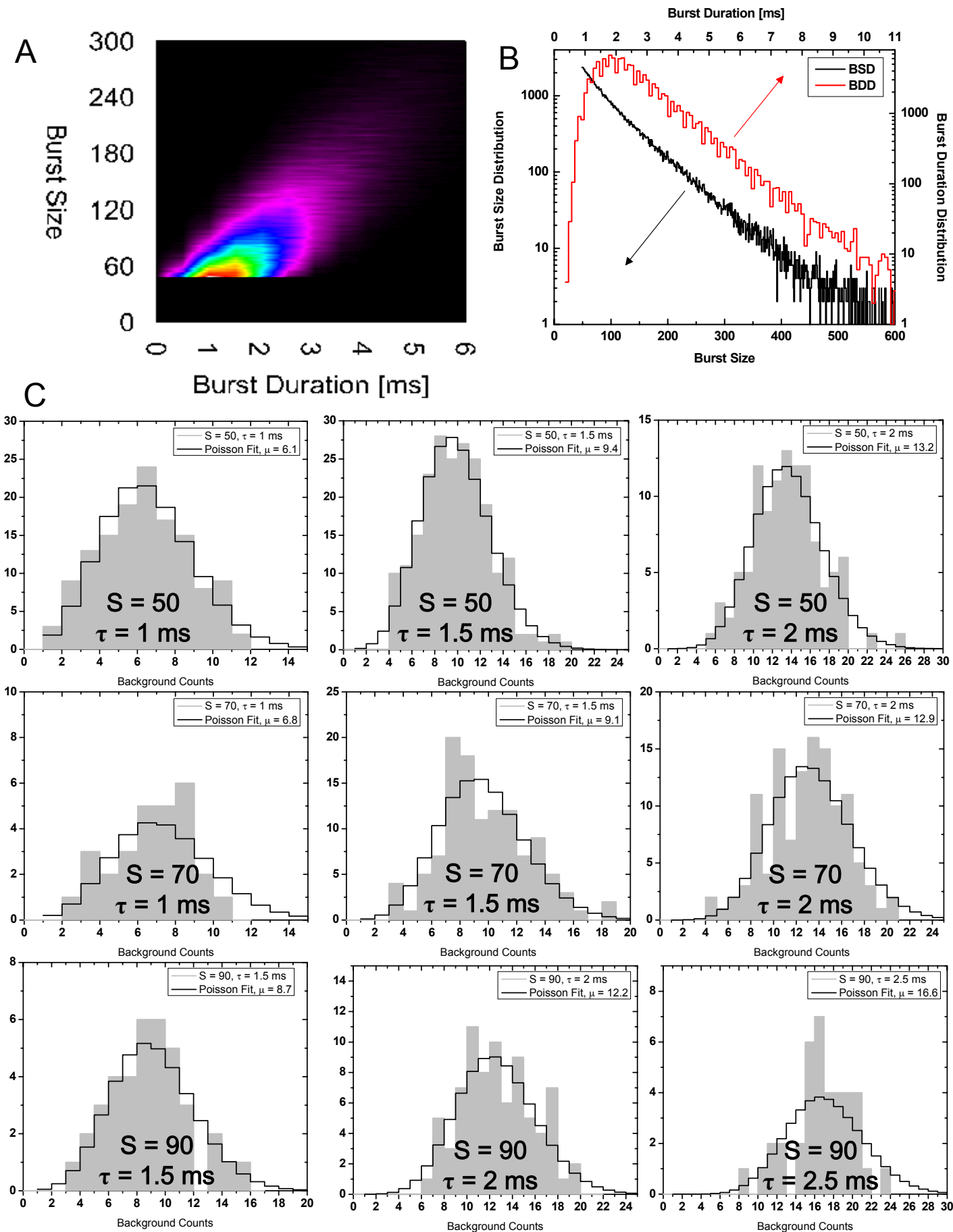


Fig. S10