THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCA Article

Rapid, Accurate Classification of Single Emitters in Various Conditions and Environments for Blinking-Based Multiplexing

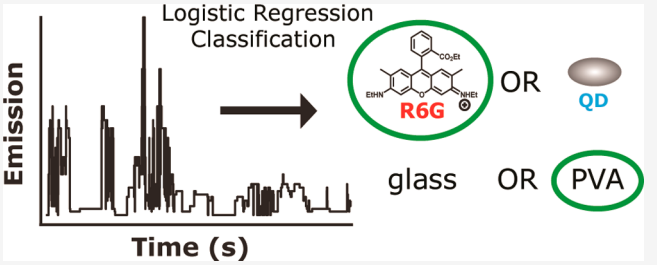
Published as part of The Journal of Physical Chemistry virtual special issue "Early-Career and Emerging Researchers in Physical Chemistry Volume 2".

Grayson R. Hoy,[†] Grace A. DeSalvo,[†] Sophia H. Haile, Emma N. Smith, and Kristin L. Wustholz*

Cite This: https://doi.org/10.1021/acs.jpca.3c00917 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Although single-molecule imaging is widely applied in biology and materials science, most studies are limited by their reliance on spectrally distinct fluorescent probes. We recently introduced blinking-based multiplexing (BBM), a simple approach to differentiate spectrally overlapped single emitters based solely on their intrinsic blinking dynamics. The original proof-of-concept study implemented two methods for emitter classification: an empirically derived metric and a deep learning algorithm, both of which have significant drawbacks. Here, a multinomial logistic regression (LR) classification is applied to rhodamine 6G (R6G) and CdSe/ZnS quantum dots (QDs) in various experimental conditions (i.e., excitation power and bin time) and environments (i.e., glass versus polymer). We demonstrate that LR analysis is rapid and generalizable, and classification accuracies of 95% are routinely observed, even within a complex polymer environment where multiple factors contribute to blinking heterogeneity. In doing so, this study (1) reveals the experimental conditions (i.e., P_{exc} = 1.2 μW and t_{bin} = 10 ms) that optimize BBM for QD and R6G and (2) demonstrates that BBM via multinomial LR can accurately classify both emitter and environment, opening the door to new opportunities in single-molecule imaging.



INTRODUCTION

Multicolor single-molecule imaging is traditionally achieved by labeling targets within a sample of interest with probes (e.g., organic dyes, quantum dots, fluorophore-coupled oligonucleotides) that emit at different wavelengths.[1−5] Although relatively straightforward to implement, this approach is fundamentally limited by its reliance on spectrally distinct emitters. We recently demonstrated blinking-based multiplexing (BBM),[6] a simple and versatile strategy to achieve multiplexed detection of individual emitters without probing their spectral color. Under continuous illumination, single emitters exhibit time-dependent fluctuations in emission intensity due to the population and depopulation of optically bright and dark states in a phenomenon called "blinking". BBM harnesses the intrinsic fluctuations between emissive "on" and nonemissive "off" periods to multiplex or classify the emitters. In our initial proof-of-concept study, we demonstrated that two commercially available and spectrally overlapped probes, rhodamine 6G (R6G) and core−shell CdSe/ZnS quantum dots (QD), are classified with ∼93% accuracy using BBM.[6]

Accurate classification of individual emitters based on their blinking dynamics requires measuring hundreds of emission-time traces and analyzing them using a statistically principled method. Previous studies have shown that the change point detection (CPD) method developed by Yang and co-workers provides an accurate determination of the statistically significant emission intensities, corresponding event durations, and event types (i.e., so-called segments and intervals) within a blinking trace.[7−13] From CPD analysis, a number of blinking statistics can be determined for each emitter: the number of distinct emission intensities (N_I), minimum and maximum intensities (I_{min} and I_{max}), time-averaged emission intensity ($\langle I \rangle_t$), average on and off interval times ($\langle t_{on,int} \rangle$ and $\langle t_{off,int} \rangle$), average on and off segment times ($\langle t_{on,seg} \rangle$ and $\langle t_{off,seg} \rangle$), the number of on and off segments per 200 s observation period ($N_{on,seg}$ and $N_{off,seg}$), the number of on and off intervals ($N_{on,int}$ and $N_{off,int}$), and corresponding event frequencies (e.g., $\nu_{E,int}$ = ($N_{on,int}$ + $N_{off,int}$)/200 s). Whereas traditional blinking studies use one or more of these statistics to probe photophysics or -chemistry within a complex environment (e.g., electron

Received: February 10, 2023
Revised: March 24, 2023

transfer events between eosin Y sensitizers and TiO_2),[14] BBM harnesses this information for emitter classification.

Indeed, we previously developed an empirically derived blinking metric (M) for BBM given by

$$M = N_I I_{\min} \nu_{E,int}{}^2 \langle I \rangle_t{}^2 \langle t_{on,seg} \rangle \quad (1)$$

which is a product of the CPD-derived blinking statistics that best differentiate R6G and QD. Although emitter classification using M yielded an overall accuracy of 92.5%, this metric method is not readily generalizable. That is, M is empirically determined after a substantial time investment and only for a particular set of emitters and experimental conditions (i.e., on glass substrates, 1 μW excitation power, 10 ms bin time). Depending on the emitters, conditions, and corresponding blinking behavior, M might be improved by incorporating different blinking statistics into eq 1 and adjusting the corresponding magnitude and sign of their exponents. The possible combinations are endless, such that defining and optimizing M for different sets of BBM experiments would require an experimentalist to perform countless trial-and-error iterations. Thus, M-based classification limits the practical application of BBM. An alternative approach to emitter classification is to employ machine learning (ML). For example, we demonstrated that a deep learning algorithm based on a 1D ResNet model yielded an average classification accuracy of 93.5%.[6] However, deep learning methods are computationally expensive, require enormous data sets, and often have limited interpretability.[15]

Expanding the utility and applicability of BBM beyond a proof-of-concept study requires demonstrating a rapid, accurate, inexpensive, and generalizable classification method. Furthermore, the efficacy of BBM under various experimental and environmental conditions is unknown. Since blinking dynamics and their associated photophysical distributions are known to broaden in heterogeneous environments,[13,16] it is critical to examine the efficacy of BBM in complex environments and under different experimental conditions. To address these limitations, we implement logistic regression (LR) classification, a popular ML algorithm that produces probability outputs for binary or multiclass classification without requiring enormous data input or computational time.[17−19] To evaluate LR classification for BBM, we measure and analyze the blinking statistics of hundreds of R6G and QD emitters as a function of laser excitation power (P_{exc}), experimental bin time (t_{bin}), and surrounding environment [i.e., on glass substrates and in poly(vinyl alcohol) (PVA) matrix]. Ultimately, we demonstrate that in all cases, LR analysis is generalizable to various conditions and environments, yielding >95% classification accuracy for BBM, even for emitters within a complex PVA environment.

METHODS

Materials and Sample Preparation. QD (Invitrogen, Qdot 565 ITK carboxyl quantum dots, 8 μM solution in borate buffer) and R6G (Acros Organics, 99%) were used as received. Emitter stock solutions were prepared in ultrapure water (ThermoScientific, Easy Pure II, 18.2 MΩ cm) in base bathed glassware (1 M KOH for 12−24 h) and sonicated for 1 min before use. Glass coverslips (Fisher Scientific, 12-545-102) were cleaned in a base bath for 12−24 h, then rinsed with ultrapure water and dried with clean air (Wilkerson, X06-02-000). Solutions of 10 wt % PVA were prepared by gradually adding powdered PVA (MP Biomedicals, 88% hydrolyzed, MW ∼15 kDa) to ultrapure water heated to 40 °C. For thorough mixing, the solution was gently stirred at 65 °C for several hours before being transferred to a rotator (Fisher Scientific, Multi-Purpose Tube Rotator, 88861049) for 12−24 h at 5 rpm.

For single-molecule experiments on glass, solutions were prepared in ultrapure water. R6G and QD stock solutions were diluted to ∼1 nM to achieve appropriate labeling densities for single-molecule studies. Samples were prepared by spin-coating 35 μL of the resulting QD or R6G aqueous solutions onto a clean coverslip using a spin coater (Laurell Technologies, WS-400-6NPP-LITE) operating at 3000 rpm with 5 s acceleration and 30 s spin times, respectively. For experiments in PVA, dye solutions were prepared in 10% wt PVA solutions, consistent with previous single-molecule studies. To image single emitters in thin films (i.e., ∼250 nm thickness)[20] of the polymer, approximately micromolar solutions of the emitters were diluted to final concentrations of 1×10^{-8} M R6G and 4×10^{-9} M QD in PVA in order to achieve appropriate labeling densities, consistent with previous single-molecule studies.[16,20] The resulting R6G/PVA and QD/PVA solutions were mixed for 12−24 h at 5 rpm. Next, samples for single-molecule microscopy were prepared by spin coating 100 μL of QD/PVA or R6G/PVA solution onto a coverslip at 3000 rpm with a 5 s acceleration time and 30 s spin time. After spin coating, emitter/PVA samples were stored under vacuum overnight before use in order to minimize the amount of adsorbed water in the sample, which is known to diminish emission intensity via vibrational relaxation.[21]

Single-Molecule Microscopy. Single-molecule samples were secured in a custom aluminum flow cell (approximately 1.5 × 3 × 0.5 in. and equipped with Tygon tubing for N_2 flow) and placed on a nanopositioning stage (Physik Instrumente, LP E-545) atop an inverted confocal microscope (Nikon, TiU). A high numerical aperture (NA) 100× oil-immersion objective (Nikon Plan Fluor, NA = 1.3) is used to focus laser excitation at 532 nm (Spectra Physics, Excelsior) to a diffraction-limited spot. Excitation powers of 0.8, 1.0, and 1.2 μW were used for single-emitter experiments on glass, and P_{exc} = 1.0 μW was used for experiments in PVA. Sample emission was collected through the objective, passed through an edge filter (Semrock, LP03-532RS-2S), and focused to an avalanche photodiode detector (APD) with a 50 μm aperture (MPD, PDM050CTB) to provide confocal resolution. Focus during raster scans was maintained using a z-axis microscope lock (Applied Science Instruments, MFC-2000). A custom LabView program was used to control the nanopositioning stage and collect corresponding emission intensity with a 30 ms dwell time. Blinking dynamics of individual emitters were captured using a 10 ms integration time for 200 s. The observation of diffraction-limited spots, blinking dynamics, irreversible single-step photobleaching, and concentration dependence of spot density were used to establish single-emitter detection. Spots that did not meet the criteria for single emitters (i.e., exceeded the diffraction limit of the instrument) were discarded.

Blinking Analysis and Classification. Change point detection (CPD) method was used to analyze blinking dynamics and parse the trajectories into statistically significant intensities and corresponding temporal durations as described elsewhere.[14] Briefly, the durations of the first and last events are not included in the calculation of the blinking statistics

https://doi.org/10.1021/acs.jpca.3c00917
J. Phys. Chem. A XXXX, XXX, XXX−XXX

since they are artificially set by the observation period. The lowest deconvolved intensity level is designated as nonemissive (off), and levels with intensities greater than one standard deviation above the rms noise are denoted as emissive (on). Blinking events are further parsed into two categories: segments and intervals. A segment corresponds to the duration of an event at a particular intensity. An interval is the duration of successive segments that occur prior to a switch between on and off (i.e., intervals are generally longer than segments). The event frequency is calculated as the number of intervals (i.e., number of switches between on and off per 200 s observation window). Unless otherwise specified, CPD-derived intensities correspond to counts per 10 ms bin. Details on the calculation of the blinking metric, M, are presented elsewhere.[6]

Like most ML techniques, LR requires standardization of input predictors for efficient optimization and interpretability. The blinking statistics generated by CPD vary in size by several orders of magnitude. Thus, the regression coefficients output from LR are meaningless without a common scale. CPD-derived blinking statistics were standardized using the z-score normalization approach. For each statistic, we find $z = (x - \mu)/\sigma$, where μ and σ are the statistic's mean and standard deviation, respectively. The resulting z values and their associated identities (i.e., P_{QD} equal to 0 or 1) are input to the LR analysis software (Matlab, mnrfit and mnrval) as predictors for training. In practice, LR uses maximum likelihood estimation (MLE) to determine the set of fit parameters that maximizes the log-likelihood given by the log(odds) (i.e., logit function):

$$\text{logit}(A) = \log \frac{(P_A)}{(1 - P_A)} = mx + ny + ...b \quad (2)$$

Here, P_A is the probability of being class A and the fit parameters are b and a set of regression coefficients (m, n, ...) associated with each predictor (x, y, ...). Each data set (i.e., collection of CPD-analyzed blinking traces for one set of conditions) is split into training (90%) and testing (10%) sets and 10-fold cross-validation is performed to minimize bias in the model's predictions.

RESULTS AND DISCUSSION

Single Emitter Classification Using LR. To circumvent the limitations of an empirically derived classification metric or deep learning algorithm, we implemented logistic regression (LR), a supervised learning algorithm that produces probability outputs [e.g., the probability that an emitter is class A (P_A)], which can then be thresholded for classification (e.g., in binary classification, a default value of $P_A > 0.50$ is used to classify emitters as type A versus type B). Importantly, this threshold is a tunable parameter that should be optimized for a particular problem or application. In the case of emitter classification for BBM, LR works by modeling P_A as a logistic (sigmoid) function applied to an input predictor (x) according to

$$P_A(x) = [1 + e^{-(mx+b)}]^{-1} \quad (3)$$

where x is one of the CPD-derived blinking statistics and the fit parameters are a y-intercept (b) and regression coefficient (m). In other words, LR analysis produces plots of the blinking statistics alongside their identities expressed as binary probabilities (i.e., known QD are assigned to $P_{QD} = 1$ and known R6G are assigned to $P_{QD} = 0$). Next, LR determines the fit parameters that best model the data.

Figure 1A shows the N_I distributions for 148 R6G and 143 QD emitters on glass plotted against their binary P_{QD} values.

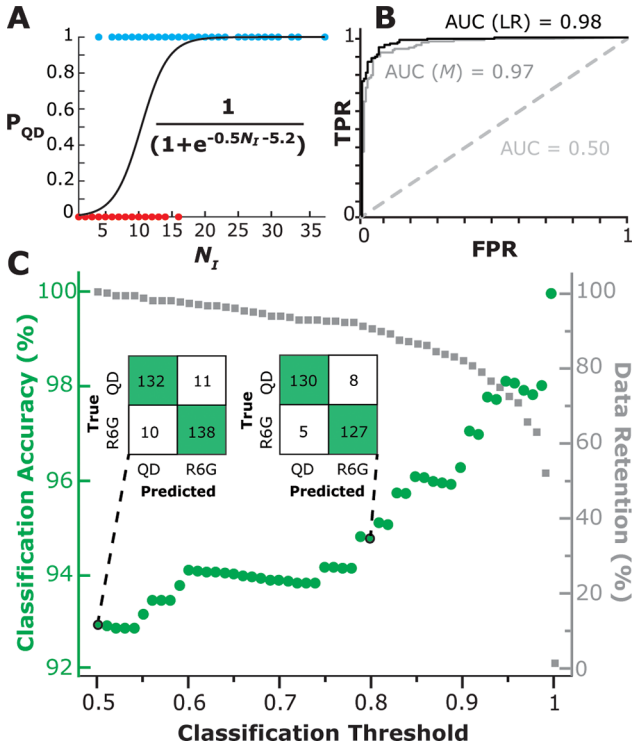


Figure 1. (A) LR analysis of the blinking dynamics from 148 R6G and 143 QD emitters obtained using 532 nm laser excitation, $P_{exc} = 1$ μW, and $t_{bin} = 10$ ms, with N_I used as the sole predictor yields 85.9% classification accuracy. (B) ROC curves for (gray) M classifier and (black) multinomial LR (i.e., P_{QD} classifier) yield AUC values of 0.97 and 0.98, respectively, as compared to the hypothetical case of random guessing (i.e., AUC = 0.5). (C) (green) Overall classification accuracy of multinomial LR increases as the classification threshold is raised from 0.5 to 1, consistent with (inset) representative confusion matrices corresponding to thresholds of 0.5 and 0.8. As classification threshold is raised and more "uncertain" emitters are discarded, the (gray) data retention (i.e., fraction of data that remains in the data set) is decreased.

Consistent with our previous study, the N_I distribution for R6G contains individual values ranging from 1 to 16, with an average value, $\langle N_I \rangle$, of 6.0 ± 0.3, where the error is the standard deviation of the mean. The corresponding distribution for 143 QD ranges from 4 to 38, with $\langle N_I \rangle = 17.1 \pm 0.6$. If N_I values are used by LR as the classification predictor, the resulting best fit to a sigmoid function corresponds to $b = -5.2$ and $m = 0.5$. In this framework, an unknown emitter with a N_I value of 12 would yield a P_{QD} of 69.0%. If the default threshold for binary classification (i.e., $P_{QD} > 50.0\%$ is classified as QD) is applied, then the emitter is classified as QD. When LR classification using N_I as the sole predictor is applied to the entire data set of 291 emitters, the overall classification accuracy (i.e., fraction of correct predictions) is 85.9%.

That LR analysis produces a relatively high classification accuracy even when significant overlap exists in the N_I distributions (Figure 1A) is quite promising. However, as described in the Supporting Information, attempts to use LR with the other CPD-derived blinking statistics as sole

https://doi.org/10.1021/acs.jpca.3c00917
J. Phys. Chem. A XXXX, XXX, XXX−XXX

predictors also produced insufficient classification accuracies. These observations are consistent with our previous findings that individual blinking statistics do not provide for sufficient multiplexing capability. To improve classification accuracy, we exploit one of the main advantages of LR—its ability to generate classification probabilities given a set of independent variables. In so-called multinomial LR, P_A is modeled as a logistic function applied to a set of independent input predictors (x, y, ...) according to

$$P_A(x, y, ...) = [1 + e^{-(mx+ny+...b)}]^{-1} \quad (4)$$

with fit parameters b and a set of regression coefficients (m, n, ...) associated with each predictor (x, y, ...). For BBM, the set of 10 independent blinking statistics from CPD (i.e., N_I, I_{min}, I_{max}, $\langle I\rangle_t$, $\langle t_{on,int}\rangle$, $\langle t_{off,int}\rangle$, $\langle t_{on,seg}\rangle$, $\langle t_{off,seg}\rangle$, $N_{on,seg}$, and $N_{off,seg}$) are used as input predictors. Other blinking statistics are highly correlated (e.g., ν_E values are directly proportional to the number of events) and are therefore discarded from LR analysis to minimize collinearity, which can lead to unreliable estimates of the regression coefficients.[22]

When multinomial LR is applied to the entire data set of 291 emitters on glass, the resulting best fit to a sigmoid function corresponds to $P_{QD} = [1 + e^{-z}]^{-1}$, where

$$\begin{aligned} z = {} & 8.5I_{min} + 4.0N_I - 3.8I_{max} + 3.2\langle I\rangle_t - 3.0\langle t_{off,seg}\rangle \\ & + 1.2\langle t_{off,int}\rangle + 0.7N_{off,seg} + 0.6\langle t_{on,int}\rangle - 0.1N_{on,seg} \\ & + 0.04\langle t_{on,seg}\rangle + 2.2 \end{aligned} \quad (5)$$

Equation 5 highlights the power of multinomial LR for BBM: the output sigmoid represents the sensitivity of emitter classification to each blinking statistic. That is, for the case of R6G and QD on glass at P_{exc} = 1 μW and t_{bin} = 10 ms, the regression coefficients for I_{min}, N_I, I_{max}, and $\langle I\rangle_t$ are largest in magnitude (i.e., 8.5, 4.0, −3.8, and 3.2, respectively) and therefore govern classification. On the other hand, $\langle t_{on,seg}\rangle$ has a regression coefficient close to zero and is relatively insignificant for classification. If the default threshold for binary classification (i.e., P_{QD} > 50.0% is classified as QD) is applied to the entire data set, then 270/291 (92.7%) of emitters are accurately classified, a process that is complete in just a few seconds. However, accuracy is insufficient to assess a classification model.

Consistent with prior studies, we performed receiver operating characteristic (ROC) curve analysis to characterize the true positive rate (i.e., occurrence of values that are actually positive and predicted positive, TPR = $\frac{TP}{TP + FN}$) and false positive rate (i.e., occurrence of values that are actually negative but predicted positive, FPR = $\frac{FP}{FP + TN}$) of the binary classification system. Here, TP, TN, FP, and FN refer to true positives, true negatives, false positives, and false negatives, respectively, and TP is designated as a QD that is correctly classified. An optimal binary classification system maximizes TPR and minimizes FPR. In ROC curve analysis, the TPR and FPR are plotted as the threshold for binary classification is varied. Whereas "perfect" classification corresponds to a point at (0, 1) that represents no false positives or negatives, random guessing will yield a diagonal line bisecting the ROC space. Thus, the shape of a ROC curve and its corresponding area under the curve (AUC) are used to establish a threshold and measure classification quality. That is, random guessing (i.e., poor classification) corresponds to AUC = 0.5 and classification quality is increased as AUC approaches 1.

Figure 1B presents the ROC curves and associated AUC values for the previously reported empirically derived blinking metric (M, eq 1) as compared to multinomial LR-based classification. The ROC plots for M-based and LR classification yield AUC values of 0.97 and 0.98, respectively. While both methods provide for excellent classification, consistent with AUC values close to unity, LR yields superior results: rapid classification without user input, a higher AUC value, and usable probability outputs (i.e., $P_{QD/R6G}$ values). Indeed, ROC curve analysis demonstrates that these probability outputs, which are thresholded for classification, can be tuned to maximize TPR and minimize FPR for a particular application.

Figure 1C depicts how the overall classification accuracy, TPR, and FPR of the multinomial LR vary as classification threshold is increased from 0.5 to 0.8. When the default classification threshold (i.e., emitters with P_{QD} or P_{R6G} > 0.5 are classified as QD or R6G, respectively) is applied to the entire data set, 270/291 (92.7%) of emitters are accurately classified. The corresponding confusion matrix, which depicts predicted versus true values, shows that TPR = 132/143 (0.92) and FPR = 10/148 (0.07). However, if the classification threshold is increased (e.g., emitters with $P_{QD/R6G}$ > 0.8 are classified as QD or R6G, while the remaining "less certain" emitters are discarded), then 257/270 (94.8%) of emitters are accurately classified, with TPR = 0.94 and FPR = 0.04. In the latter case, 21 (7.2%) of emitters with $P_{QD/R6G}$ < 0.8 have been discarded to yield a data retention of 92.8%. Figure 1C summarizes the trade-off between classification accuracy, classification threshold, and data retention, which can be tuned for a particular application. For example, in data-dense experiments where ∼25% of the data can be readily discarded, multinomial LR analysis yields ∼98% classification accuracy. The stepped appearance of this plot comes from abrupt decreases in accuracy associated with rare, misidentified emitters, a characteristic that would be diminished by increasing the sample size. Collectively, these results demonstrate that increasing the LR classification threshold yields improved performance as compared to the previously reported M analysis—in a rapid and fully automated manner.

LR Classification of Emitters in Various Experimental Conditions. To determine if BBM using multinomial LR is readily generalizable to other experimental conditions and environments, we first investigated the blinking dynamics of R6G and QD emitters at various P_{exc} and t_{bin}. The blinking dynamics of 102, 148, and 88 R6G molecules as well as 98, 143, and 83 QDs on glass were obtained at 0.8, 1.0, and 1.2 μW, respectively. Figure 2 shows representative blinking dynamics of R6G and QD on glass obtained using 532 nm laser excitation and t_{bin} of 10 ms as P_{exc} is increased from 0.8 to 1.2 μW (i.e., ∼650−1000 W/cm^2 based on the diffraction-limited spot size of the instrument). Consistent with previous studies, under the same conditions QD blinking is brighter, more persistent, and exhibits more intensity fluctuations as compared to R6G. CPD analysis of the blinking traces in Figure 2 demonstrate that for both emitters, I_{max} increases with P_{exc}, consistent with an increase in the photoexcitation rate constant before saturation. At $P_{exc} \geq$ 1.2 μW, photobleaching of R6G occurs within ∼10−50 s and a concomitant decrease in time-averaged emission intensity ($\langle I\rangle_t$) is observed. On the other hand, QD emitters are relatively photostable and $\langle I\rangle_t$ is observed to increase with P_{exc}. Consistent with the blinking

https://doi.org/10.1021/acs.jpca.3c00917
J. Phys. Chem. A XXXX, XXX, XXX−XXX

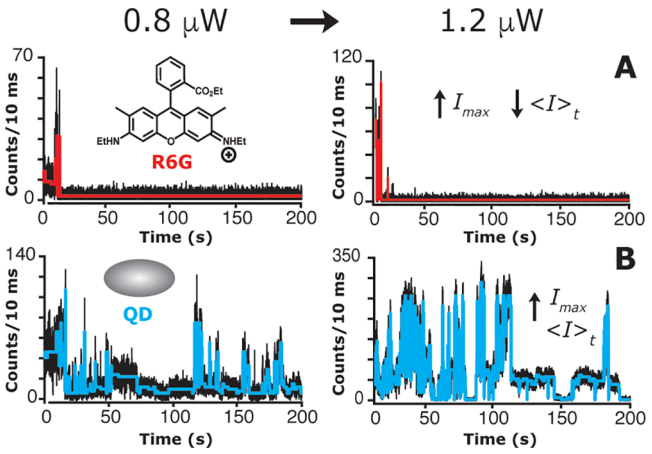


Figure 2. Representative P_{exc}-dependent blinking dynamics of (red) R6G and (blue) QD on glass obtained using 532 nm excitation and t_{bin} = 10 ms at 0.8 and 1.2 μW. (A) R6G blinking shown with (red line) CPD analysis demonstrates an increase in I_{max} and decrease in $\langle I \rangle_t$ with P_{exc}. (B) Corresponding (blue line) CPD analysis of QD blinking exhibits more intense and persistent emission as P_{exc} is increased.

measurements of R6G and QD on glass at P_{exc} = 1.0 μW, the differences between the average CPD-derived blinking statistics of R6G and QD are modest and the corresponding distributions appear overlapped at all powers.

To test the generalizability of multinomial LR and the extent to which P_{exc} optimizes BBM, we analyzed the blinking dynamics of R6G and QD at 0.8 and 1.2 μW using CPD and then performed both *M*-based and multinomial LR classification. As expected, *M*-based classification without additional time investment is insufficient (i.e., using eq 1 to classify R6G and QD at different P_{exc} yields classification accuracies of 82.0% and 79.5% for 0.8 and 1.2 μW, respectively). However, multinomial LR is both rapid and accurate. For example, multinomial LR analysis of 102 R6G and 98 QD emitters obtained at P_{exc} of 0.8 μW yield the best fit to a sigmoid function corresponding to $P_{QD} = [1 + e^{-z}]^{-1}$, where $z = 1.4I_{min} + 0.5N_I - 0.8I_{max} + 0.5\langle I \rangle_t + 0.1\langle t_{off,seg} \rangle - 0.4\langle t_{off,int} \rangle + 1.6N_{off,seg} - 0.09\langle t_{on,int} \rangle + 1.7N_{on,seg} + 0.1\langle t_{on,seg} \rangle + 0.9$. The regression coefficients in the above sigmoid are entirely different from those in eq 5, demonstrating that the blinking dynamics and associated classification are indeed dependent on P_{exc}. This comparison also highlights another strength of multinomial LR for BBM. Since each fit parameter represents its own axis in multidimensional space, the resulting sigmoids are quite distinct, even for relatively similar experimental conditions (see Table S1). When the default classification threshold is applied to the entire data set, 166/200 (83.0%) of emitters are accurately classified with a TPR and FPR of 0.73 and 0.08, respectively. A higher classification accuracy of 93.3% corresponding to TPR = 0.94 and FPR = 0.07 can be achieved for this data set using a classification threshold of 0.86, though the resulting ~50% data retention is less than ideal. However, this is the only data set for which classification accuracy does not reach 95%, suggesting that BBM of R6G and QD is not optimal at these conditions.

For blinking measurements of 88 R6G and 83 QD emitters obtained at P_{exc} = 1.2 μW, multinomial LR produced a sigmoid function corresponding to $z = 69I_{min} + 2.3N_I - 4.3I_{max} + 2.6\langle I \rangle_t - 1.1\langle t_{off,seg} \rangle - 0.0009\langle t_{off,int} \rangle - 1.1N_{off,seg} - 1.5\langle t_{on,int} \rangle + 1.4N_{on,seg} + 0.3\langle t_{on,seg} \rangle + 17.4$. Again, the regression coefficients for different P_{exc} are distinct. In this case, when the default classification threshold is applied to the entire data set, 270/291 (93.0%) of emitters are accurately classified with TPR = 132/143 (0.92) and FPR = 10/148 (0.06). A 95% classification accuracy and 82.5% data retention are reached using a classification threshold of 0.87. Thus, at all powers measured here, BBM via multinomial LR classification yields excellent binary classification. To investigate the impact of experimental integration time on multinomial LR classification and BBM, we computationally binned the blinking dynamics of 143 R6G and 148 QD emitters on glass obtained at P_{exc} = 1.0 μW and t_{bin} = 10 ms up to 100 ms and analyzed the resulting traces using CPD. Multinomial LR analysis of the 100 ms data produced a default classification accuracy of 92.1%. If a classification threshold of 0.79 is used, then the 95% classification accuracy benchmark is met with >90% data retention.

Table 1 summarizes the P_{exc}- and t_{bin}-dependent multinomial LR classification results, including the classification accuracies, TPR and FPR values associated with the default classification threshold of $P > 0.50$, as well as the thresholds needed to achieve 95% accuracy in each experiment. Collectively, the P_{exc}- and t_{bin}-dependent blinking studies of R6G and QD emitters demonstrate that multinomial LR provides for high classification accuracy and usable probability outputs that can be tuned to various experimental conditions. This analysis also shows that BBM via multinomial LR is optimized for R6G and QD emitters at a P_{exc} of 1.2 μW and t_{bin} of 10 ms. That is, these experimental conditions on glass produce the highest default classification accuracies and minimal data loss. The corresponding best fits to sigmoid functions output by multinomial LR for all conditions and environments examined in this study are summarized in the Supporting Information (Table S1).

LR Classification in PVA. After establishing the efficacy of multinomial LR to classify and assess various experimental

Table 1. Multinomial LR Classification of R6G and QD Emitters in Various Conditions

		using default threshold			for 95% accuracy			
classification		accuracy (%)	TPR	FPR	threshold	TPR	FPR	data retention (%)
R6G vs QD	0.8 μW, 10 ms	83.0	0.73	0.08	0.86[a]	0.94	0.07	52.0
	1.0 μW, 10 ms	92.7	0.92	0.07	0.80	0.94	0.04	92.8
	1.0 μW, 100 ms	92.1	0.92	0.08	0.79	0.96	0.05	91.1
	1.2 μW, 10 ms	93.0	0.92	0.06	0.87	0.97	0.06	82.5
	1.0 μW, 10 ms, PVA	86.7	0.85	0.12	0.87	0.95	0.05	58.5
glass vs PVA	1.0 μW, 10 ms, R6G	88.1	0.91	0.16	0.91	0.98	0.08	64.2
	1.0 μW, 10 ms, QD	76.1	0.71	0.18	0.88	0.95	0.08	29.0

[a]95% accuracy is not achieved under these conditions. This threshold corresponds to a maximum 93.3% accuracy that also minimizes data loss.

https://doi.org/10.1021/acs.jpca.3c00917
J. Phys. Chem. A XXXX, XXX, XXX−XXX

conditions for BBM, we turned our focus to studying blinking in more complex environments. To determine if BBM can be translated to multicolor imaging of biological samples, for example, we sought to examine emitter classification within a heterogeneous environment that is more representative of a complex biological system. Previous work has shown that poly(vinyl alcohol) (PVA) serves as an effective model biological environment in single-molecule experiments.[23] In particular, Gibbs and co-workers validated PVA as a model environment to efficiently quantify and correlate the photoswitching properties of eight different fluorophores to image quality in single-molecule localization microscopy (SMLM). Single-molecule spectroscopy studies of fluorophores and quantum dots in polymers also have a long history,[21,24−30] and the blinking dynamics of R6G and QD in PVA are known to be complex.[28,30] Therefore, to assess BBM in a more complex and relevant environment, we measured the blinking dynamics of hundreds of R6G and QD emitters in PVA at 532 nm excitation at 1.0 μW P_{exc}.

Figure 3A shows a representative false-colored emission image of 10^{-8} M R6G-doped PVA (R6G/PVA). Single

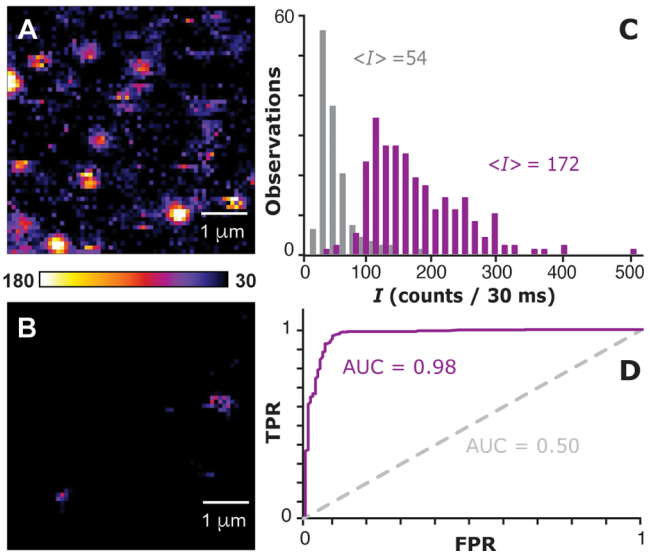


Figure 3. (A) Representative false-colored emission image of 10^{-8} M R6G/PVA at P_{exc} = 1 μW, where the color scale represents photon counts per 30 ms. (B) Corresponding image for blank PVA obtained using the same conditions demonstrates residual blank PVA signal that cannot be attributed to R6G or QD. (C) Histogram of I values compiled from (gray) 138 blank PVA and (purple) 261 emitter/PVA measurements. (D) ROC curve for (purple) I classifier to distinguish between background PVA and emitter/PVA yields an AUC value of 0.98, as compared to the hypothetical case when a classifier cannot distinguish (i.e., AUC = 0.5).

emitters are typically identified as bright diffraction-limited spots atop a dim background in emission images, but Figure 3A also shows an increase in background signal that is not observed for corresponding measurements on glass. To determine if the polymer is the source of the increased background signal, we measured 30 emission scans of blank PVA. Indeed, a representative emission image of blank PVA (Figure 3B) displays dim diffraction-limited spots and pixels of relatively high intensity scattered across the image. To ensure that background signal from the polymer, presumably due to impurities and/or scattering from large aggregates, are excluded from the emitter data set, we measured the distribution of emission intensity (I) values for 138 blank PVA spots and a total of 261 QD/PVA and R6G/PVA spots (termed "emitter/PVA"). The $\langle I \rangle$ of 54 ± 2 counts per 30 ms for blank PVA is much lower than the corresponding value for emitter/PVA of 172 ± 4 counts per 30 ms. As shown in Figure 3C, the I distribution for blank PVA is also well separated from the corresponding emitter/PVA distribution.

Since the distributions are not substantially overlapped, we implemented an I threshold to exclude blank PVA from true emitters using a ROC curve analysis (Figure 3D). In this case, a true positive is defined as QD or R6G being correctly identified as an emitter, whereas a false positive refers to the case where signal from blank PVA is incorrectly attributed to an emitter. The AUC value of 0.98 indicates that I is an effective parameter for differentiating blank PVA from emitters embedded in PVA. Using the ROC curve, a threshold is carefully selected to maximize TPR (i.e., include emitter data) while minimizing FPR (i.e., exclude blank PVA). Additionally, to prevent large quantities of true emitter data from being discarded, the threshold should not be set too high. With these factors in mind, ROC curve analysis yields a TPR of 0.92 and FPR of 0.06 for an I threshold value of 100 counts per 30 ms, where $I \geq 100$ corresponds to emitter and $I < 100$ to PVA. In this framework, 20/261 (<8%) of spots from the emitter/PVA data set are attributed to blank PVA and discarded. This low data loss rate is a small trade-off to ensure that a vast majority (>94%) of background PVA is excluded from further analysis. For the rest of this article, we focus on the remaining 129 QD/PVA and 112 R6G/PVA spots and associated blinking dynamics, which have been categorized as emitters.

The blinking dynamics and corresponding CPD-derived blinking statistics of QD and R6G are significantly modified in PVA as compared to glass substrates (see Supporting Information, Table S2), consistent with previous observations.[6,10,26,30] The distributions of on- and off-interval times are also broader in PVA (data not shown), consistent with the observation of complex electron transfer in PVA.[25,26,30] The observations that the blinking behavior of single emitters is highly dependent on local environment and that the dynamics in PVA appear more broad as compared to glass confirm that PVA is a good model environment for BBM. To investigate the efficacy of BBM in PVA, we performed multinomial LR classification on 129 QD/PVA and 112 R6G/PVA emitters. Figure 4 shows the resulting classification accuracy and data retention of QD and R6G embedded in PVA as a function of classification threshold. Classification of the full data set using the default classification threshold of $P_{QD/R6G} > 0.5$ yields 86.7% classification accuracy, indicating that BBM performs well but not up to the standards reported for glass substrates. However, the 95% classification accuracy benchmark is met with a classification threshold of 0.87, though only ∼60% of the data is retained. Nevertheless, in cases where the experimentalist can afford to exclude data (e.g., in SMLM), multinomial LR classification produces exceptional accuracy, even within the complex PVA environment.

As a final test of both multinomial LR classification and the BBM concept, we tested the accuracy of not just emitter classification (i.e., QD vs R6G) but also environment classification (i.e., glass vs PVA). Table 1 shows the environmental classification results for R6G and QD emitters. Multinomial LR classification can successfully differentiate between R6G on glass and in PVA. A 95% classification

https://doi.org/10.1021/acs.jpca.3c00917
J. Phys. Chem. A XXXX, XXX, XXX−XXX

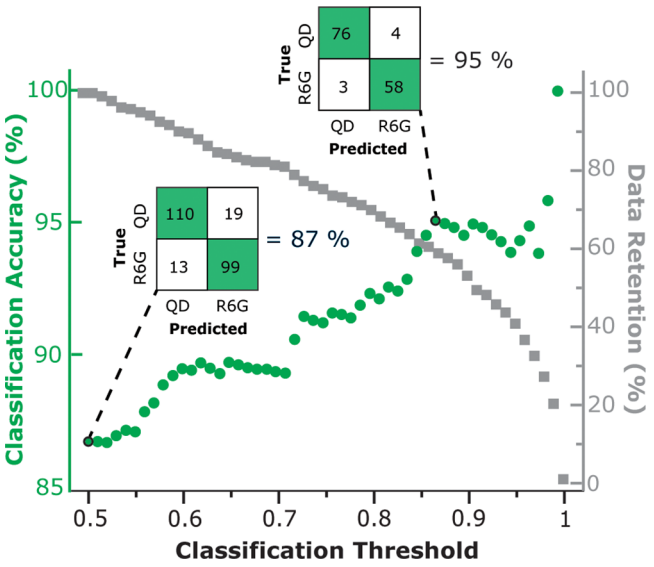


Figure 4. Multinomial LR classification of QD/PVA and R6G/PVA. (green) Classification accuracy and (gray) data retention is plotted as a function of classification threshold. (insets) Confusion matrices for the default classification versus a 0.87 classification threshold. Using the default threshold retains 100% the blinking data, but yields an 87% accuracy, TPR = 0.85, and FPR = 0.12. If the classification threshold is increased to 0.87, then ~40% of "uncertain" emitters are discarded to yield a 95% accuracy, TPR = 0.95, and FPR = 0.05.

accuracy, with TPR of 0.98 and FPR of 0.08, is achieved using a threshold of 0.91 and data retention of 64.2%. Interestingly, the model is less robust for distinguishing between QD on glass and PVA, though a 95% accuracy is attained with a sufficiently high classification threshold. The observation that BBM based on LR analysis can classify both the emitter and its environment is extremely promising. For instance, we envision that this approach can be harnessed for multiplexed imaging of multicomponent systems—providing not just spectral multiplexing but environmental characterization at the same time.

CONCLUSIONS

Multinomial LR classification represents a powerful, new way to classify spectrally overlapped single emitters and their environments for BBM. In this study, we demonstrate that unlike an empirically derived blinking metric, LR classification is rapid, generalizable across various conditions, and accurate. Indeed, the results summarized in Table 1 show that classification accuracies of 95% are routinely achieved for QD and R6G emitters as excitation power and bin time are modified. In doing so, this study also demonstrates that optimizing experimental conditions can yield enhanced multiplexing capability. For instance, BBM of QD and R6G emitters is optimal (i.e., the highest default classification accuracies are observed) at P_{exc} = 1.2 μW and t_{bin} = 10 ms. With multinomial LR classification, the full potential of BBM is now within reach. For example, single emitters are classified with high accuracy even in a complex PVA environment, where multiple factors are at play to impact blinking. These results have inspired us to expand BBM studies to include multiple spectrally overlapped emitters and even structurally analogous organic fluorophores. Our preliminary results show that multiclass multinomial LR readily provides for >90% classification accuracy among three different probes with high data retention. Finally, the observation that BBM can classify both the emitter and its environment leads to intriguing new research directions, including BBM imaging experiments that decode both emitter identity and environment or location. Ultimately, BBM via multinomial LR classification has extraordinary potential to expand the palette of single-molecule probes and unlock new opportunities in single-molecule imaging. To further advance its capabilities, both fundamental blinking and applied BBM studies of organic probes under imaging conditions are underway.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.3c00917.

Additional blinking statistics, distributions, and fit parameters (PDF)

AUTHOR INFORMATION

Corresponding Author

Kristin L. Wustholz − *Department of Chemistry, William & Mary, Williamsburg, Virginia 23187, United States;* orcid.org/0000-0001-7705-4240; Email: kwustholz@wm.edu

Authors

Grayson R. Hoy − *Department of Chemistry, William & Mary, Williamsburg, Virginia 23187, United States;* orcid.org/0000-0002-4105-0456

Grace A. DeSalvo − *Department of Chemistry, William & Mary, Williamsburg, Virginia 23187, United States*

Sophia H. Haile − *Department of Chemistry, William & Mary, Williamsburg, Virginia 23187, United States*

Emma N. Smith − *Department of Chemistry, William & Mary, Williamsburg, Virginia 23187, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.3c00917

Author Contributions

†G.R.H. and G.A.D. contributed equally to this paper

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (CHE-2102099). G.R.H. acknowledges the NASA Virginia Space Grant Consortium for support through an Undergraduate Research Fellowship and the Charles Center at William & Mary for Honors research funding.

REFERENCES

(1) Möckl, L.; Moerner, W. E. Super-Resolution Microscopy with Single Molecules in Biology and Beyond-Essentials, Current Trends, and Future Challenges. *J. Am. Chem. Soc.* **2020**, *142* (42), 17828−17844.

(2) Willets, K. A.; Wilson, A. J.; Sundaresan, V.; Joshi, P. B. Super-Resolution Imaging and Plasmonics. *Chem. Rev.* **2017**, *117* (11), 7538−7582.

(3) Patterson, G.; Davidson, M.; Manley, S.; Lippincott-Schwartz, J. Superresolution Imaging Using Single-Molecule Localization. *Annu. Rev. Phys. Chem.* **2010**, *61* (1), 345−367.

(4) Bittel, A. M.; Saldivar, I. S.; Dolman, N. J.; Nan, X.; Gibbs, S. L. Superresolution Microscopy with Novel BODIPY-Based Fluorophores. *PLoS One* **2018**, *13*, e0206104.

(5) Gallina, M. E.; Xu, J.; Dertinger, T.; Aizer, A.; Shav-Tal, Y.; Weiss, S. Resolving the Spatial Relationship between Intracellular Components by Dual Color Super Resolution Optical Fluctuations Imaging (SOFI). *Opt Nanoscopy* **2013**, *2* (1), 2.
(6) DeSalvo, G.; Hoy, G.; Kogan, I.; Li, J.; Palmer, E.; Luz-Ricca, E.; Scemama de Gialluly, P.; Wustholz, K. L. Blinking-Based Multiplexing: A New Approach for Differentiating Spectrally Overlapped Emitters. *J. Phys. Chem. Lett.* **2022**, *13* (22), 5056−5060.
(7) Li, H.; Yang, H. Statistical Learning of Discrete States in Time Series. *J. Phys. Chem. B* **2019**, *123* (3), 689−701.
(8) Song, N.; Yang, H. Parallelization of Change Point Detection. *J. Phys. Chem. A* **2017**, *121* (27), 5100−5109.
(9) Watkins, L. P.; Yang, H. Detection of Intensity Change Points in Time-Resolved Single-Molecule Measurements. *J. Phys. Chem. B* **2005**, *109* (1), 617−628.
(10) Wong, N. Z.; Ogata, A. F.; Wustholz, K. L. Dispersive Electron-Transfer Kinetics from Single Molecules on TiO_2 Nanoparticle Films. *J. Phys. Chem. C* **2013**, *117* (41), 21075−21085.
(11) Lynch, P. G.; Richards, H.; Wustholz, K. L. Unraveling the Excited-State Dynamics of Eosin y Photosensitizers Using Single-Molecule Spectroscopy. *J. Phys. Chem. A* **2019**, *123* (13), 2592−2600.
(12) Tan, J.; Garakyaraghi, S.; Tagami, K.; Frano, K.; Crockett, H.; Ogata, A.; Patterson, J.; Wustholz, K. L. Contributions from Excited-State Proton and Electron Transfer to the Blinking and Photobleaching Dynamics of Alizarin and Purpurin. *J. Phys. Chem. A* **2017**, *121* (1), 97−106.
(13) Wustholz, K. L.; Bott, E. D.; Kahr, B.; Reid, P. J. Memory and Spectral Diffusion in Single-Molecule Emission. *J. Phys. Chem. C* **2008**, *112* (21), 7877−7885.
(14) Kopera, K. M.; Tuckman, H. G.; Hoy, G. R.; Wustholz, K. L. Origin of Kinetic Dispersion in Eosin-Sensitized TiO_2: Insights from Single-Molecule Spectroscopy. *J. Phys. Chem. C* **2021**, *125* (43), 23634−23645.
(15) Mater, A. C.; Coote, M. L. Deep Learning in Chemistry. *J. Chem. Inf. Model* **2019**, *59*, 2545−2559.
(16) Riley, E. A.; Bingham, C.; Bott, E. D.; Kahr, B.; Reid, P. J. Two Mechanisms for Fluorescence Intermittency of Single Violamine R Molecules. *Phys. Chem. Chem. Phys.* **2011**, *13* (5), 1879−1887.
(17) Boni, Y. T.; Cammarota, R. C.; Liao, K.; Sigman, M. S.; Davies, H. M. L. Leveraging Regio- and Stereoselective C(Sp3)−H Functionalization of Silyl Ethers to Train a Logistic Regression Classification Model for Predicting Site-Selectivity Bias. *J. Am. Chem. Soc.* **2022**, *144* (34), 15549−15561.
(18) Kirasich, K.; Smith, T.; Sadler, B. Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets. *SMU Data Science Review* **2018**, *1* (3), 9.
(19) Hosmer, D. W.; Lemeshow, S. *Applied Logistic Regression*, 2nd ed.; 2000.
(20) Sluss, D.; Bingham, C.; Burr, M.; Bott, E. D.; Riley, E. A.; Reid, P. J. Temperature-Dependent Fluorescence Intermittency for Single Molecules of Violamine R in Poly(Vinyl Alcohol). *J. Mater. Chem.* **2009**, *19* (40), 7561−7566.
(21) Ishitobi, H.; Kai, T.; Fujita, K.; Sekkat, Z.; Kawata, S. On Fluorescence Blinking of Single Molecules in Polymers. *Chem. Phys. Lett.* **2009**, *468* (4−6), 234.
(22) Stoltzfus, J. C. Logistic Regression: A Brief Primer. *Acad. Emer. Med.* **2011**, *18* (10), 1099−1104.
(23) Bittel, A. M.; Nickerson, A.; Saldivar, I. S.; Dolman, N. J.; Nan, X.; Gibbs, S. L. Methodology for Quantitative Characterization of Fluorophore Photoswitching to Predict Superresolution Microscopy Image Quality. *Sci. Rep.* **2016**, *6* (1), 29687.
(24) Clifford, J. N.; Bell, T. D. M.; Tinnefeld, P.; Heilemann, M.; Melnikov, S. M.; Hotta, J.; Sliwa, M.; Dedecker, P.; Sauer, M.; Hofkens, J.; et al. Fluorescence of Single Molecules in Polymer Films: Sensitivity of Blinking to Local Environment. *J. Phys. Chem. B* **2007**, *111* (25), 6987−6991.
(25) Zondervan, R.; Kulzer, F.; Kol'chenko, M. A.; Orrit, M. Photobleaching of Rhodamine 6G in Poly(Vinyl Alcohol) at the Ensemble and Single-Molecule Levels. *J. Phys. Chem. A* **2004**, *108* (10), 1657−1665.
(26) Zondervan, R.; Kulzer, F.; Mathies, G.; Orrit, M. Non-Exponential Kinetics of Photoblinking and Photobleaching of Rhodamine 6G in Polyvinylalcohol. *Meeting Abstract, Single Molecules* **2002**, *3* (5-6), 317.
(27) Schuster, J.; Cichos, F.; von Borczyskowski, C. Blinking of Single Molecules in Various Environments. *Opt. Spec.* **2005**, *98*, 712−717.
(28) Issac, A.; Krasselt, C.; Cichos, F.; von Borczyskowski, C. Influence of the Dielectric Environment on the Photoluminescence Intermittency of CdSe Quantum Dots. *ChemPhysChem* **2012**, *13* (13), 3223−3230.
(29) Kuno, M.; Fromm, D. P.; Hamann, H. F.; Gallagher, A.; Nesbitt, D. J. Nonexponential "Blinking" Kinetics of Single CdSe Quantum Dots: A Universal Power Law Behavior. *J. Chem. Phys.* **2000**, *112* (7), 3117−3120.
(30) Zondervan, R.; Kulzer, F.; Orlinskii, S. B.; Orrit, M. Photoblinking of Rhodamine 6G in Poly(Vinyl Alcohol): Radical Dark State Formed through the Triplet. *J. Phys. Chem. A* **2003**, *107* (35), 6770−6776.

https://doi.org/10.1021/acs.jpca.3c00917
J. Phys. Chem. A XXXX, XXX, XXX−XXX