

Goal: assess the likelihood of a given model for single photon responses based on a collection of experimental responses. Use this approach to determine most likely model parameters (e.g. trajectory of Rh\* activity, degree of saturation in cascade, number of shutoff steps, ...).

This could be done in (at least) two broad ways: (1) use simulation of large collection of responses to characterize probability of observing a given response from model, and evaluate likelihood of each experimental response using this characterization; (2) use mean and SEM of parameters of response from experiment and use these to evaluate likelihood of given set of simulated responses. In principle, option (1) seems preferable because it does not involve any assumptions about what matters about responses, and because measured responses can be evaluated against a large set of simulated responses (as large as needed to eliminate finite sample size effects). In practice, however, this approach seems difficult for reasons given below.

Consider a set of measured single photon responses  $r_n(t)$ , which consist of a single photon response with added continuous noise. We can write these responses compactly as

$$r_n(t) = \sum_i a_{i,n} e_i(t) + \eta(t) \quad (1)$$

where  $a_{i,n}$  are weights specific to response  $n$ ,  $e_i$  are a set of basis functions shared by all  $r_n$  (e.g. principle components), and  $\eta$  represents continuous dark noise. We want to compare these to a set of simulated responses  $s_n(t)$ :

$$s_n(t) = \sum_i b_{i,n} e_i(t) \quad (2)$$

where  $e_i$  are the same as those for the measured responses.

The ideal approach would seem to be to use the simulated responses to generate distributions  $P(b_{i,n})$ . These distributions could then be used to evaluate the likelihood of each measured single photon response — i.e. the likelihood of  $a_{i,n}$  for each single photon response. Combining likelihoods across all measured single photon responses would tell us how likely a given set of measured responses is given a particular model. We could then search across model parameters to find those that maximize the likelihood.

A problem with this approach is that we cannot obtain the  $a_{i,n}$  directly because we cannot separate, on a single trial, the single photon response and the continuous noise. It would seem we are limited to making this separation on average, when we do not need to know the noise trajectory but instead only its covariance (or other statistics). Thus we can estimate the distribution  $P(a_{i,n})$  by projecting both single photon responses and failures along each  $e_i$  and correcting the single photon response projections for that failures projections (e.g. if we assume things are gaussian we can subtract the variances). This seems to preclude evaluating the likelihood of individual experimental responses against a set of simulated responses.

One alternative would seem to switch things around — evaluate the likelihood of individual simulated responses against the set of measured responses. There are several issues with this

approach. One is whether we have a sufficient number of experimental responses to really estimate the full distributions (i.e.  $P(a_{i,n})$ ). Another more fundamental issue is that maximizing the likelihood of the simulated responses can be achieved by placing each response at the peak of  $P(a_{i,n})$  — i.e. removing all variability in the simulated responses.

The issues above would appear to complicate approaches based on single responses.

An alternative is to evaluate likelihood based on several parameters of the collected responses. Those parameters could be the distributions of coefficients  $P(a_{i,n})$  and  $P(b_{i,n})$  or other quantities such as  $CV_{area}$ ,  $t_{peak,\sigma^2}/t_{peak,\mu}$ , ... . We can estimate the mean and SEM of these parameters from experiment (e.g. estimate each parameter for each rod and take mean and SEM of those estimates). Then we can use the mean and SEM of each parameters to evaluate likelihood of given set of simulated responses. This approach in practice seems to work quite well (meaning likelihood changes substantially for 10-20% changes in model parameters).

Practical issues: we can collect a few thousand single photon responses and a few thousand failures. These come from 10-20 rods. Responses of different rods differ in amplitude and kinetics, and in addition the properties of the dark noise vary from rod to rod. Some, but not all, of this rod-to-rod variation can be eliminated by normalizing the peak amplitude and time to peak of the single photon response to 1 before combining data across rods.

8/30/2011 FMR

Here's another possibility. Use PCA to identify low dimensional representation of combined experimental single photon responses. For each cell measure how much of total variability of single photon responses is captured by each dimension. Use these fractional variance measures, along with total variance (CV) in likelihood analysis.