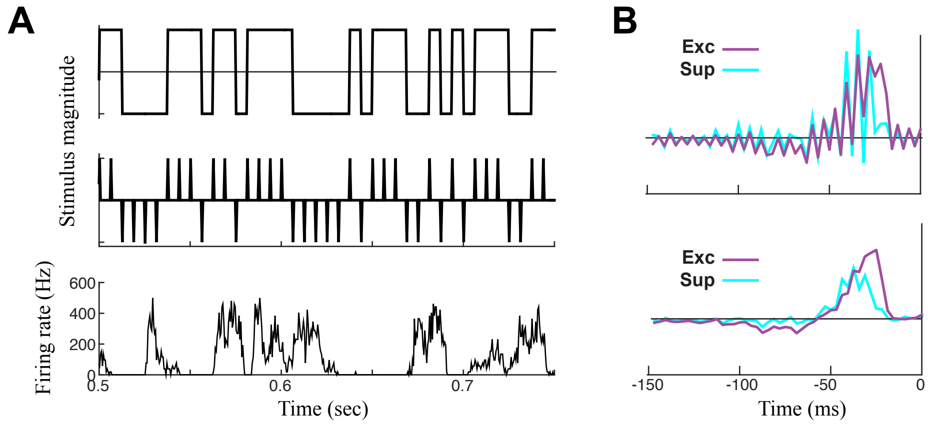
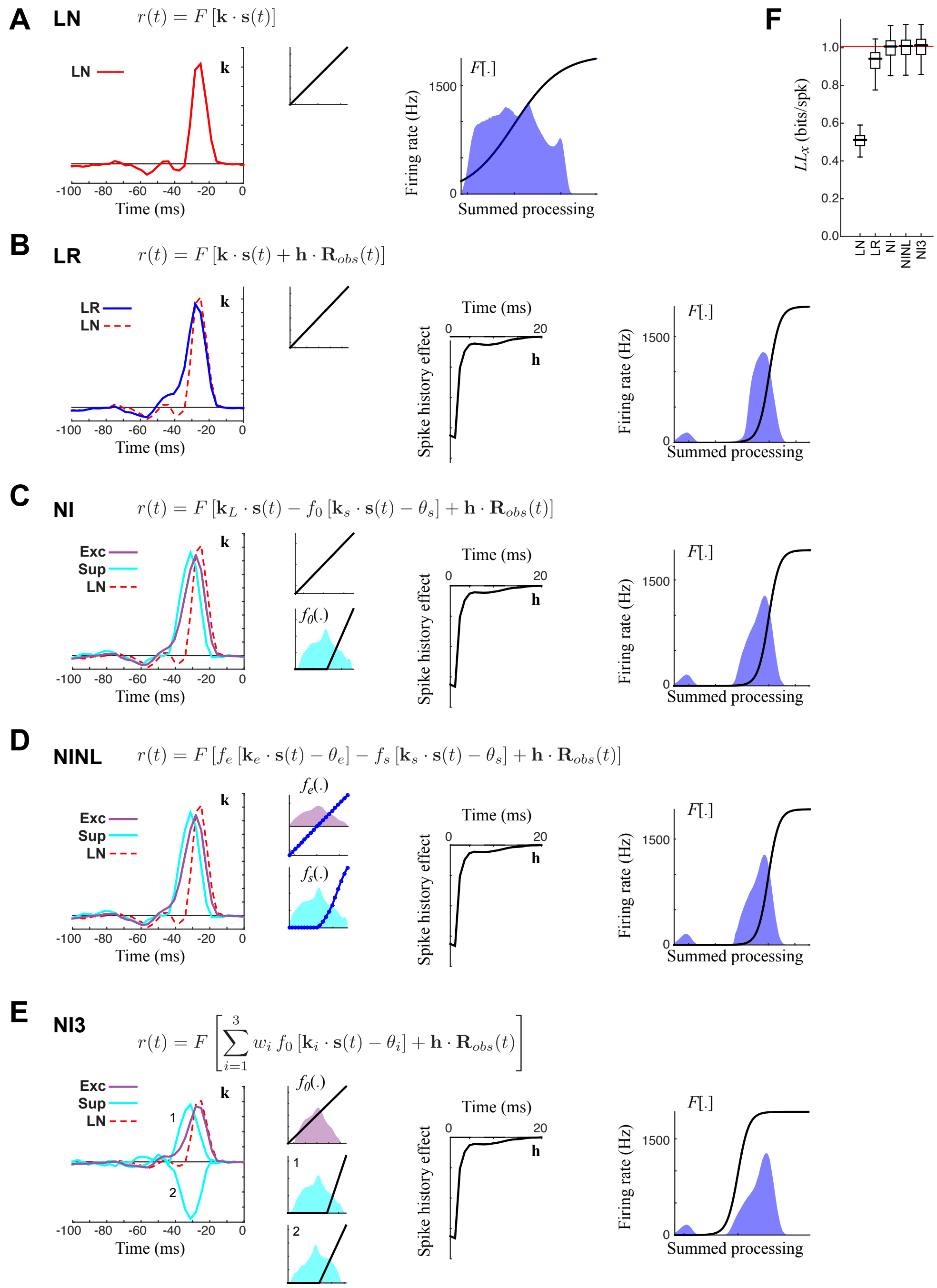
Supplemental Figure 2



Stimulus pre-processing at high time resolution. A. The stimulus was produced by a cathode ray tube (CRT) with a monitor refresh of 160 Hz (6.3 ms frame duration). We considered two possibilities in modeling the resulting visual input at the 0.5 ms resolution used for analysis. *Top*: one common approach (e.g., Butts et al., 2011) is to assume the stimulus is constant over the frame duration. The result is shown for 250 ms sample of the binary m-sequence stimulus (one of three used in the experiments in this paper). *Middle*: we used an alternative approach that assumed the stimulus was only on in one analysis bin, of the form shown. While this might be a closer approximation to the actual CRT signal, the point is to make it at least as fine as the time resolution of the actual monitor signal (see below). *Bottom:* for reference, we show the PSTH of an example ON LGN neuron (aligned with the stimulus), showing the fine temporal resolution present. **B.** The excitatory and suppressive filters of the NIM for the example ON neuron given the constant-stimulus assumption (*top)* and the delta-function assumption (bottom). These filters were optimally smoothed (to maximize the likelihood of the model), so the jaggedness in the top-row filters actually helps in predicting the response given the stimulus assumption. This can be explained by the fact that these top-row filters are trying to deconvolve the overly smooth stimulus assumed. By comparison, assuming the stimulus has too-fine temporal resolution (*bottom*) will make the resulting filters include a smoothing convolution. However, if the filters themselves have longer time scales than 6 ms (as they do), such smoothing will have negligible effect, and thus this assumption is much more appropriate for fitting these filters in neurons that are sensitive to such high temporal resolution of the stimulus.



Range of models for example OFF RGC. A. The LN model for an example OFF RGC, organized as follows: *Left*: temporal filter; *Center:* upstream nonlinearity, in this case there is no upstream nonlinearity, so a linear function is pictured; *Right*: spiking nonlinearity (black), shown with respect to the underlying range of summed processing of the model components (blue). **B.** LR model, involving linear stimulus processing and refractoriness implemented by a spike-history term. Same organization as (A), except a column 3 now shows the spike-history term. **C.** NIM with a linear "excitatory" term and nonlinear suppression, which is the form used to fit RGCs throughout the paper. The suppressive upstream nonlinearity $f_0(.)$ is parametric (rectified linear) and specified by a single threshold parameter. **D.** The NIM where the upstream nonlinearities are non-parametric and both fit (NINL model). The resulting form of the non-parametric nonlinearities is what motivates the parametric form used in this paper. **E.** Models could be fit with more nonlinear terms, in this case NIM3 shows the best model fit to the data with 3 nonlinear terms. **F.** Cross-validated log-likelihoods of all models shown on repeat data. This shows the NIM (C) has nearly identical performance as the more complex models, motivating its use throughout the paper.