
Research Article: New Research | Sensory and Motor Systems

Dynamics of temporal integration in the lateral geniculate nucleus

<https://doi.org/10.1523/ENEURO.0088-22.2022>

Cite as: eNeuro 2022; 10.1523/ENEURO.0088-22.2022

Received: 25 February 2022

Revised: 21 June 2022

Accepted: 27 July 2022

This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.eneuro.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Copyright © 2022 Alexander et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

1

2

3 Dynamics of temporal integration in the lateral geniculate nucleus

4

5 Prescott C. Alexander^{1,2} Henry J. Alitto^{1,3} Tucker G. Fisher^{1,4}6 Daniel L. Rathbun^{1,5} Theodore G. Weyand⁶ W. Martin Usrey^{1,2,3,*}

7

8 ¹ Center for Neuroscience, University of California, Davis, Davis, California, USA9 ² Center for Vision Science, University of California, Davis, Davis, California, USA10 ³ Department of Neurobiology, Physiology, and Behavior, University of California, Davis, Davis, California,
11 USA12 ⁴ Department of Neurobiology, Stanford University School of Medicine, Stanford, California, USA13 ⁵ Department of Ophthalmology, Henry Ford Health System, Detroit, Michigan, USA14 ⁶ Department of Cell Biology and Anatomy, Louisiana State University Health Sciences Center, New Orleans,
15 Louisiana, USA.

16 * Correspondence: wmusrey@ucdavis.edu

17

18

19

20

21 Acknowledgments: We thank K.E. Neverkovec, D.J. Sperka, J. Johnson, and R. Oates for expert technical
22 assistance. Supported by NIH grants EY013588 (WMU), P30 EY12576 (WMU), EY015387 (PCA).

23 **Abstract**

24

25 Before visual information from the retina reaches primary visual cortex, it is dynamically filtered by the
26 lateral geniculate nucleus (LGN) of the thalamus, the first location within the visual hierarchy at which
27 non-retinal structures can significantly influence visual processing. To explore the form and dynamics of
28 geniculate filtering we used data from monosynaptically connected pairs of retinal ganglion cells (RGCs)
29 and LGN relay cells in the cat that, under anesthetized conditions, were stimulated with binary white noise
30 and/or drifting sine-wave gratings to train models of increasing complexity to predict which RGC spikes
31 were relayed to cortex, what we call “relay status”. In addition, we analyze and compare a smaller data set
32 recorded in the awake state to assess how anesthesia might influence our results. Consistent with previous
33 work, we find that the preceding retinal inter-spike interval is the primary determinate of relay status with
34 only modest contributions from longer patterns of retinal spikes. Including the prior activity of the LGN cell
35 further improved model predictions, primarily by indicating epochs of geniculate burst activity in recordings
36 made under anesthesia, and by allowing the model to capture gain control-like behavior within the awake
37 LGN. Using the same modeling framework, we further demonstrate that the form of geniculate filtering
38 changes according to the level of activity within the early visual circuit under certain stimulus conditions.
39 This finding suggests a candidate mechanism by which a stimulus specific form of gain control may operate
40 within the LGN.

41

42 **Significance**

43

44 The LGN is a dynamic, tunable filter, transforming information as it flows from the retina to primary visual
45 cortex. In this work we utilize a large data set of monosynaptically connected RGC and LGN cell pairs to
46 model the filtering function performed by individual LGN neurons in the anesthetized or awake state. We
47 demonstrate that, while much of the filtering that the LGN performs can be accounted for by temporal
48 summation, other factors, such as the bursting activity of relay cells, also play a role. Additionally, we show
49 that the time scale of summation is dynamic under certain stimulus and network conditions and that the
50 integration dynamics are largely similar between the anesthetized and awake states.

51 Introduction

52

53 There are two primary dimensions along which relay cells of the lateral geniculate nucleus (LGN) might
54 transform the visual information that they receive from the retina, namely space and time. In the spatial
55 dimensions, a substantial body of evidence suggests a limited transformation, most notably an increase in the
56 strength of the antagonistic surround of the center/surround receptive field (Usrey et al., 1998, 1999; Wang et
57 al., 2010). On the other hand, data demonstrating substantial temporal transformations by LGN relay cells
58 of their direct retinal inputs abound (Usrey et al., 1998; Carandini et al., 2007; Sincich et al., 2007; Sincich
59 et al., 2009; Babadi et al., 2010; Wang et al., 2010; Rathbun et al., 2016). Prior work has demonstrated that
60 the temporal transformation performed by the LGN results in an increased encoding efficiency in the signals
61 sent by the LGN to primary visual cortex (V1) compared to the signals received from the retina (Sincich
62 et al., 2009; Uglesich et al., 2009; Wang et al., 2010), and that this increased efficiency can be explained by
63 temporal summation within relay cells (Carandini et al., 2007; Sincich et al., 2007; Casti et al., 2008) and a
64 selective filtering out of less informative retinal spikes (Rathbun et al., 2010). Furthermore, it has recently
65 been shown that temporal summation within the LGN changes with stimulus contrast (Alitto et al., 2019a),
66 suggesting that geniculate filtering is dynamic and can adapt to the statistics of the visual environment. The
67 aim of this work is to investigate this filtering process by modeling the input-output relation of LGN cells
68 using generalized linear models (GLMs), and to further examine whether the input-output relation changes
69 under different stimulus or network conditions.

70 In order to investigate the input-output relation of LGN relay cells, we first assembled a large database
71 of simultaneous, extracellular recordings of monosynaptically connected retinal ganglion cell (RGC) - LGN
72 cell pairs from previously published work in anesthetized cats (Usrey et al., 1998; Rathbun et al., 2010;
73 Fisher et al., 2017). Although these data offer a near optimal level of spatial and temporal resolution with
74 which to examine input-output relations in single neurons, they only capture a single RGC input to each
75 relay cell, which are thought to receive input from between two and five RGCs in the cat (Cleland et al.,
76 1971; Usrey et al., 1999). Thus, instead of focusing our analyses on the full spike train produced by relay
77 cells, which contains contributions from all RGC inputs, we instead focus specifically on trying to model
78 the process that determines which spikes from the recorded RGC are relayed, that is elicit a spike in their
79 geniculate partner, and which are not. We begin by considering the simplest model of temporal summation,
80 the often used inter-spike interval (ISI) model (Usrey et al., 1998; Sincich et al., 2007; Weyand, 2007; Wang
81 et al., 2010; Rathbun et al., 2016; Alitto et al., 2019a) whereby the relay probability of each retinal spike
82 is predicted based on the elapsed time since the last retinal spike. We then show how the ISI model can

83 be conceptually extended using GLMs, allowing the full pattern of retinal spikes, within a given window of
84 time, to be used in the predictions. We then introduce a two component GLM that includes the pattern
85 of LGN spikes preceding each retinal spike to investigate whether the LGN spike train contains additional
86 information about the relay probability of future retinal spikes. Finally, we explore whether high levels
87 of activity within the retino-thalamo-cortical circuit influences how LGN relay cells integrate their retinal
88 inputs, and whether this change might explain the dynamic temporal filtering within relay cells that has
89 been previously reported (Rathbun et al., 2016; Alitto et al., 2019a).

90 While this approach allows the computations performed by individual LGN relay cells to be examined with
91 a level of detail unmatched by any existing method, it does require anesthesia to record the spiking activity
92 of individual RGCs within the eye. In order to complement this approach, and to offer more general findings,
93 we additionally analyze a smaller data set recorded from awake cats in which S-potentials, the extracellular
94 record of excitatory post-synaptic potentials driven by the dominant retinal input (Kaplan and Shapley,
95 1984), were recorded simultaneously with the LGN spikes that they frequently elicit (Weyand, 2007). Given
96 the small size of the awake data set, we cannot make quantitative comparisons between the anesthetized
97 and awake state. However, we can use the awake data set to qualitatively confirm, or refute, whether our
98 findings from the anesthetized state are generally applicable.

99 **Methods**

100

101 **Data sources**

102

103 The data analyzed in this study contributed to previous reports on the retinogeniculate pathway in both
104 anesthetized (Usrey et al., 1998, 1999; Rathbun et al., 2010, 2016; Fisher et al., 2017; Alitto et al., 2019b)
105 and awake (Weyand, 2007) cats. All experimental procedures conformed to NIH and USDA guidelines and
106 were approved by the Institutional Animal Care and Use Committee at the University of California, Davis
107 or LSU Health Sciences Center.

108

109 **Code Accessibility**

110

111 All data and code used in this study are available at: https://github.com/scottiealexander/relayglm_paper.

112 The code is also available as Extended Data.

113

114 **Computing and software resources**

115

116 All analyses were performed on a Dell Precision T3610 desktop with an Intel Xenon processor (E5-1620)
117 running the Ubuntu 18.04.6 operating system.

118 All analyses were performed using custom written code in the Julia programming language version 1.6.1

119 (Bezanson et al., 2017). Visualizations were created using the Julia interface (Johnson, 2020) to the Mat-

120 plotlib graphics package (Hunter, 2007).

121

122 **Anesthetized recordings**

123

124 **Surgery and preparation**

125

126 Twenty-three adult cats of either sex contributed to this data set. As previously described, anesthesia
127 was initiated with ketamine (10 mg/kg, i.m.) or ketamine and thiopental sodium (20 mg/kg, i.v.) and
128 maintained with either sodium pentothal (2-3 mg/kg/h, i.v.), or isoflurane (0.7-2%). Administration rate
129 of the anesthetic agent was increased when physiological monitoring indicated low levels of anesthesia. A
130 tracheotomy was performed and animals were placed in a stereotaxic apparatus and mechanically respired.
131 Body temperature, ECG, EEG, and expired CO₂ were monitored for the duration of the experiment. All
132 wound margins were infused with lidocaine. The cortical surface overlying the LGN was exposed by a

133 craniotomy and durotomy and then protected with a layer of agarose. To minimize eye movements and
134 facilitate retinal recordings, the sclera beneath the lateral margin of each eye was glued to a rigid ring that
135 was mounted to the stereotaxic frame. The posterior chamber of each eye was accessed via a trans-scleral
136 guide tube inserted through the ring. Upon completion of surgical procedures, animals were paralyzed with
137 either vecuronium bromide (0.2 mg/kg/hr, i.v.) or gallium triethiodide (6-8 mg/kg/h). The nictitating
138 membranes of the eye were retracted with 10% phenylephrine and pupils were maintained in a dilated state
139 with 1% atropine sulfate and flurbiprofen sodium (1.5 mg/h). The eyes were then refracted, fitted with
140 contact lenses, and focused on a tangent screen in front of the animal.

141

142 **Electrophysiological recording and visual stimuli**

143

144 Extracellular recordings of RGCs were made using single, parylene-coated microelectrodes (AM Systems)
145 inserted through the trans-scleral guide tube into the posterior chamber of the eye via a custom-made
146 manipulator. Extracellular recordings of LGN cells in the A laminae were made using a seven-channel
147 multielectrode array (Thomas Recording). Neural signals were amplified, filtered (AM Systems, Thomas
148 Recording) and recorded by either a computer running Brainwave software (Datawave Systems) or a 1401
149 data acquisition system connected to a computer running the Spike2 software package (Cambridge Electronic
150 Design). Single-neuron isolation was based on waveform analysis and the presence of a refractory period in
151 the auto-correlogram.

152 Visual stimuli were generated by either a Pepper Graphics System video card (Number Nine Computer
153 Corporation) and presented on a CRT monitor at 80 or 100 Hz (NEC Multisync), or a VSG 2/5 visual
154 stimulus generator (Cambridge Research Systems) and presented on a gamma-calibrated CRT monitor at
155 140 Hz (Sony). Drifting sinewave gratings that varied in either contrast or diameter were presented at a
156 temporal frequency of 4 Hz and at the optimal spatial frequency for the RGC-LGN pair under study. Binary
157 white-noise stimuli were comprised of a 16x16 grid of squares where the brightness of each square (black
158 or white) on each stimulus frame was governed by a 215-1 frame long pseudorandom sequence (the “m-
159 sequence”, (Sutter, 1987; Reid et al., 1997)). The stimulus frame was updated either on every or every other
160 monitor frame (7 - 25 ms stimulus frame duration).

161

162 **Awake recordings**

163

164 Four adult cats of either sex contributed to this data set. Details of surgical procedures, training, and
165 recording have been described previously (Weyand and Gafka, 1998; Weyand, 2007). In brief, animals

166 underwent an initial implant surgery to allow for head-fixed training and eye tracking, followed by a training
167 period in which animals learned to maintain fixation to within 1.5° of a small spot (0.2°) for 1-3 seconds
168 to receive a food reward. Following the training period animals underwent a second surgery in which a
169 canula was introduced into the brain (~6 mm deep) through a small craniotomy and fixed in place allowing a
170 microelectrode to access the LGN for awake, extracellular, recordings (the orientation of the canula could be
171 adjusted, see (Weyand, 2007) for details). Signals from microelectrodes (1 - 1.5 M Ω at 1 kHz) were amplified
172 ($\times 100$ - 1,000), filtered (0.001 - 10 kHz), and digitized at 22.5 kHz by a modified VCR (A. R. Vetter,
173 Rebersberg, PA) and transferred to a computer for storage using hardware and software from National
174 Instruments (Austin, TX). S-potentials and action potentials were identified and sorted offline using Mini-
175 Analysis (Athens, GA) (for details see (Weyand, 2007)). As S-potentials are thought to be the extracellular
176 record of excitatory post-synaptic potentials driven by the dominant retinal input (Cleland et al., 1971;
177 Kaplan and Shapley, 1984; Weyand, 2007), the delay between a successful S-potential (reflecting a relayed
178 RGC spike) and the triggered LGN spike is substantially shorter than the analogous delay between an RGC
179 spike recorded within the eye and the LGN spike that it triggers. Thus, for the analyses presented in this
180 paper the timing of the S-potentials for a given pair were shifted “backwards” in time relative to the paired
181 LGN spikes by 2.4 ms which ensured that the median delay between S-potentials and triggered LGN spikes
182 (which was 0.4 ms prior to shifting) matched the median delay observed between RGC and triggered LGN
183 spikes in the anesthetized data set (2.8 ms). This shift helps to minimize any contribution from the different
184 recording approaches to any differences in timing that may be observed between the awake and anesthetized
185 data sets and allows S-potentials to be identified as relayed or not using the same criteria as those used
186 for RGC spikes recorded within the eye (see Identification of monosynaptically connected pairs and relayed
187 RGC spikes). For simplicity, throughout this paper we refer to both RGC spikes recorded within the eye as
188 well as time-shifted S-potentials as “RGC spikes”.

189 Given the difficulty of recording S-potentials in an awake animal, the duration over which individual pairs
190 could be recorded was often quite limited and thus most of the pairs analyzed in this study (7 of 8) were not
191 presented with a controlled stimulus but were instead stimulated by whatever features of the well-lit room
192 fell within their receptive field (see (Weyand, 2007) for details). The one exception, pair 200001250, was
193 stimulated with a sinewave grating (see Extended Data Figure 7-2).

194 **Data analysis**

195

196 **Identification of monosynaptically connected pairs and relayed RGC spikes**

197

198 Simultaneously recorded RGC and LGN cells that showed a prominent, short-latency peak in their spike
 199 time cross-correlograms were considered to be monosynaptically connected pairs (Mastrorarde, 1987; Usrey
 200 et al., 1999; Rathbun et al., 2010; Fisher et al., 2017). All S-potential-LGN pairs from (Weyand, 2007) met
 201 this criterion by definition. Cross-correlograms, LGN spiking relative to each RGC spike, were constructed
 202 for all pairs (from both the anesthetized and awake data sets) using 0.1 ms bins. Peaks were considered
 203 prominent if at least one bin exceeded a threshold of $\mu_{\text{baseline}} + 3\sigma_{\text{baseline}}$, where μ_{baseline} is the mean of the
 204 baseline period spanning 30-50 ms on either side of the peak bin, and σ_{baseline} is the standard deviation
 205 of the baseline period. All bins adjacent to the peak bin that also exceeded the threshold were considered
 206 part of the peak. Peaks were considered short latency if they occurred within 2-6 ms of $t = 0$, the time of
 207 each retinal spike. All retinal spikes that were followed by a LGN spike that fell within the peak bins of
 208 cross-correlograms were considered “relayed”, all other retinal spikes were considered “non-relayed”. Retinal
 209 efficacy (or simply efficacy) is the number of relayed spikes divided by the total number of retinal spikes.
 210 Likewise, all LGN spikes that were preceded by a retinal spike within the monosynaptic window (defined as
 211 above) were considered “triggered”. Retinal contribution (or simply contribution) is the number of triggered
 212 LGN spikes divided by the total number of LGN spikes.

213

214 **Modeling framework**

215

216 All models discussed in the paper generally take the form:

217

218

219

$$\lambda = \sigma(f(t|\theta)) \quad (1)$$

220

221 where t are retinal spike times, θ are the model parameters, and λ are the predicted relay probabilities in
 222 (0, 1). For the ISI model, f is a nonlinear map between the interval $t_i - t_{i-1}$ and a conditional intensity. For
 223 GLMs, f is a linear function of t represented as a binary vector over an n millisecond interval prior to each
 224 t_i . For two component GLMs, f also takes as input the LGN spike times t_{LGN} , $f(t; t_{\text{LGN}}|\theta)$. For all models,
 225 σ is the logistic function:

226

227

228

229

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \quad (2)$$

236 that maps a conditional intensity to a relay probability.

237

238 **Assessing model performance**

239

240 All models were assessed in a train-on-90%, test-on-10% ten-fold cross-validation procedure. In each fold, 90%
 241 of the data was used to fit the model and the remaining 10% was used only to assess model performance. This
 242 procedure was performed ten times such that all data appear in the test set exactly once. Data partitioning
 243 across folds was performed such that all test sets contained approximately the same number of relayed spikes.
 244 This balancing helped reduce the variability in mean efficacy across folds for a given pair, which serves to
 245 stabilize the performance metric that we used (see below) especially for pairs with relatively low mean efficacy.
 246 Model performance is the mean performance across folds.

247 As all models presented in this paper produce a relay probability (0,1) for each retinal spike, we use the
 248 cross-validated single-event Bernoulli information ($J_{\text{Bernoulli}}$) to assess model performance. $J_{\text{Bernoulli}}$ is the
 249 Bernoulli analog of the cross-validated single-spike information used for Poisson GLMs (Williamson et al.,
 250 2015) and can be calculated from the Bernoulli log-likelihood function \mathcal{L} (Truccolo et al., 2005; Williamson
 251 et al., 2015):

252

$$253 \quad \mathcal{L}(\lambda; y(t)) = \sum y(t) \log(\lambda) + (1 - y(t)) \log(1 - \lambda) \quad (3)$$

254

255 where λ are the predicted relay probabilities (as above), and $y(t)$ indicates whether each retinal spike was
 256 relayed as {0,1}, what we call “relay status”. Using \mathcal{L} we can calculate $J_{\text{Bernoulli}}$:

257

$$258 \quad J_{\text{Bernoulli}} = \frac{1}{n_{\text{test}} \log(2)} (\mathcal{L}(\lambda_{\text{train}}; y_{\text{test}}) - \mathcal{L}(y_{\text{test}})) \quad (4)$$

259

260 where $\lambda_{\text{train}} = \lambda(t_{\text{test}} | \theta_{\text{train}})$ are the predicted relay probabilities for test-set retinal spikes (t_{test}) given
 261 the parameters (θ_{train}) learned from the training-set. $y_{\text{test}} = y(t_{\text{test}})$ is the observed relay status for t_{test} ,
 262 and $n_{\text{test}} = \text{length}(t_{\text{test}})$ is the number of retinal spikes in t_{test} . $\mathcal{L}(y_{\text{test}})$ represents the log-likelihood of a
 263 homogeneous Bernoulli model where the mean efficacy of the test-set is predicted for every spike:

264

$$265 \quad \mathcal{L}(y_{\text{test}}) = r_{\text{test}} \log(\lambda_{\text{test}}) + (n_{\text{test}} - r_{\text{test}}) \log(1 - \lambda_{\text{test}}) \quad (5)$$

266

267 where $r_{\text{test}} = \sum y_{\text{test}}$ and λ_{test} is the mean efficacy across t_{test} (i.e., $\frac{r_{\text{test}}}{n_{\text{test}}}$).

272 In this construction, $J_{\text{Bernoulli}}$ has units of bits/spike and for well fit models will take on values between ~ 0
 273 (no better than a homogeneous model) and 1 (perfect performance). In practice, poorly fit models can result
 274 in negative $J_{\text{Bernoulli}}$ due to separate training and testing data sets (i.e., cross-validation). Conceptually,
 275 $J_{\text{Bernoulli}}$ quantifies how informative model predictions are about the relay status of the test-set relative to
 276 a homogeneous model with the same mean efficacy as the test-set. While somewhat elaborate compared
 277 to metrics like accuracy, for a binary process like relay status it is important to take into account the
 278 fact that as λ_{test} approaches 0 or 1, correctly predicting the outcome becomes trivial. Quantifying model
 279 performance relative to a homogeneous model ensures that as λ_{test} approaches 0 or 1, $J_{\text{Bernoulli}}$ approaches
 280 0 for a model with perfect predictions (and values less than 0 for lesser performing models). While this
 281 behavior is necessary to accurately quantify model performance on this kind of classification task (i.e., where
 282 the number of relayed and non-relayed spike cannot be matched), it entails that the maximum achievable
 283 $J_{\text{Bernoulli}}$ depends in part on the mean efficacy of the RGC-LGN pair being modeled. For example, for a
 284 pair with an efficacy of 0.05 the maximum $J_{\text{Bernoulli}}$ for a perfect performing model is 0.286 bits/spike.

285

286 **Inter-spike interval models**

287

288 Usrey et al. (Usrey et al., 1998) described the effect of retinal inter-spike interval (ISI) on efficacy, using
 289 the term “paired spike enhancement”. They observed that retinal spikes following short ISIs have a higher
 290 efficacy than those following long ISIs. Following (Wang et al., 2010), we recast that observation as a simple
 291 model for predicting which retinal spikes were relayed based on the elapsed time since the last retinal spike.
 292 This model was constructed by creating a histogram of the ISIs preceding all relayed retinal spikes and
 293 dividing the count in each bin by the total number retinal ISIs that fell within that bin. We used a bin width
 294 of 1 ms and the resulting histograms were smoothed with a unit-area Gaussian (the standard deviation of
 295 which was chosen separately for each pair, see [Hyperparameter optimization](#)) to produce a function relating
 296 ISI to efficacy (ISI-efficacy function) which we denote as $P(t|t_{\text{ISI}})$ where t is the time of a retinal spike and
 297 $t_{\text{ISI}} = t_i - t_{i-1}$ is the ISI preceding t for ISIs up to a maximum (ISI_{MAX}) that was chosen separately for
 298 each pair (see [Hyperparameter optimization](#)). For any retinal spikes with ISIs greater than ISI_{MAX} , the
 299 model predicted the average efficacy across all ISIs in the corresponding data set. For example, if the ISI of
 300 a retinal spike within a given test-set is greater than ISI_{MAX} , the model would predict the mean efficacy of
 301 that test-set (i.e., $\frac{r_{\text{test}}}{n_{\text{test}}}$).

302

303

304

305

After building $P(t|t_{\text{ISI}})$ (abbreviated as P below for clarity) for a given pair, the fitting algorithm then found
 a linear transform $f(P) = \beta P + \alpha$ such that the Bernoulli log-likelihood of the resulting predictions

306
$$\lambda_{\text{ISI}} = \sigma(f(P)) \tag{6}$$

307
308 was maximized. This allows a shifting and rescaling of the predictions such that the mean of λ_{ISI} matches the
309 mean efficacy of the data being used for model fitting. Omitting this step would penalize the ISI model quite
310 significantly in the calculation of $J_{\text{Bernoulli}}$ because λ_{ISI} may be incorrectly scaled relative to the mean efficacy
311 (due to the ISI cutoff) and thus the likelihood of the homogeneous model, $\mathcal{L}(\lambda)$ above, would be expected to
312 be large relative to $\mathcal{L}(\lambda_{\text{ISI}})$, yielding potentially negative values for $J_{\text{Bernoulli}}$ that would incorrectly indicate
313 poor performance.

314 As with all models discussed herein, for quantifying performance all parameters and hyperparameter were
315 determined from an independent subset of the data from that used to assess performance (see [Assessing](#)
316 [model performance](#)).

317
318 **Generalized linear models**

319
320 **General** In order to generalize the ISI based model to consider all activity within a period of time preceding
321 each retinal spike, we used a generalized linear model (GLM) framework ([Truccolo et al., 2005](#); [Paninski et](#)
322 [al., 2007](#); [Pillow et al., 2008](#); [Babadi et al., 2010](#)). GLMs are a generalization of ordinary linear regression
323 in which the to-be-predicted, or “response”, variable need not be normally distributed, and the predictor
324 variables and response variable need not be linearly related ([Nelder and Wedderburn, 1972](#)). Similarly, GLMs
325 can be thought of as a particular class of linear-nonlinear (LN) cascade models in which the nonlinearity,
326 or activation function, is fixed and invertible ([Chichilnisky, 2001](#); [Paninski, 2004](#)). GLMs generally take the
327 form:

328
329
$$\mathbf{y} = g(\mathbf{X}\boldsymbol{\theta}) \tag{7}$$

330
331 where y is the response variable, \mathbf{X} is a matrix of predictors, $\boldsymbol{\theta}$ is a vector of model parameters, and g is
332 the activation function (formally, the inverse link function). Given an assumed or known error distribution
333 of y and an appropriate choice of g , the parameters $\boldsymbol{\theta}$ can be efficiently fit by maximum likelihood methods
334 ([Paninski, 2004](#); [Babadi et al., 2010](#)).

335 In the present context, the response (y) that we are trying to predict is the (binary) relay status of each
336 retinal spike. Thus, a natural choice for the error distribution of y is the Bernoulli distribution, and a natural

337 choice of activation function is the Logistic function (i.e., logistic regression). The Bernoulli-Logistic GLM
 338 is given by:

$$339 \quad \mathbf{y} = \lambda(t|\boldsymbol{\theta}) = \sigma(\mathbf{X}\boldsymbol{\theta}) \quad (8)$$

341 where t are the retinal spike times, and the predictor matrix \mathbf{X} is derived from the retinal spike times alone
 342 (retinal history model) or using both the retinal and LGN spikes times (combined history model). The relay
 343 status, y , of a set of retinal spikes, t , is then modeled as:
 344

$$345 \quad y(t) \sim \text{Bernoulli}(\lambda(t|\boldsymbol{\theta})) \quad (9)$$

347 The parameter vector $\boldsymbol{\theta}$ that minimized the negative log-likelihood (i.e., $-\mathcal{L}$) for each model instance was
 348 found using Newton's method (Nocedal and Wright, 2006) as implemented in (Mogensen and Riseth, 2018).
 349

350 **Retinal history models** Within the GLM framework used here, \mathbf{X} is an m by $n + 1$ matrix where m is
 351 the number of retinal spikes being used to fit the model (typically 90% of the retinal spikes recorded under
 352 a given stimulus condition, see [Assessing model performance](#)) and n is the number of temporal components.
 353 The additional column is the additive offset or "y-intercept" term. In the "retinal history only" version (RH)
 354 of the model, whose predictor matrix, sans-offset, we will refer to as \mathbf{X}_R , the "temporal components" are
 355 simply n 1 ms time bins representing the retinal spike train, as a binary vector, during the n milliseconds
 356 preceding each retinal spike. In this form, summing over the m rows of \mathbf{X}_R would yield the autocorrelogram
 357 of the retinal spike train over an n millisecond window. The hyperparameter n was optimized separately for
 358 each pair (see [Hyperparameter optimization](#)). The n parameters corresponding to the n time bins of a fitted
 359 model can be thought of as a linear kernel or filter that reflects the extent to which retinal spikes occurring
 360 at a given time prior to the "target spike" influence the likelihood that the target spike will be relayed.
 361

362 Given previous work suggesting that LGN temporal filters are likely to be smooth functions in this context
 363 (Usrey et al., 1998; Rathbun et al., 2010), and to help prevent overfitting, we introduce a smoothing prior
 364 on $\boldsymbol{\theta}$ (excluding the y-intercept term), yielding a maximum *a-posteriori* (MAP) estimator for $\boldsymbol{\theta}$:

$$365 \quad \mathcal{L}_{\text{MAP}}(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}) - \eta \sum (\theta_i - \theta_{i-1})^2 \quad (10)$$

367 where the prior weighting term η is optimized separately for each pair (see [Hyperparameter optimization](#)).

368 **Combined history models** The combined history (CH) model extends the RH model by introducing
 369 another set of predictors derived from the activity of the LGN cell. In the RH model, the LGN cell only
 370 contributes by classifying each retinal spike as relayed or non-relayed, whereas in the CH model the recent
 371 activity of the LGN cell can also contribute to the relay prediction (via spikes not “caused” by the recorded
 372 RGC). For the CH model, the predictor matrix \mathbf{X}_C can be thought of as the column-wise concatenation
 373 of \mathbf{X}_R with an analogous binary matrix, \mathbf{X}_L , of size m by p where each row of length p is the LGN cell’s
 374 binary spike train (1 ms bin size) during the p milliseconds preceding each retinal spike. Thus, summing
 375 over the rows of \mathbf{X}_L would yield the cross-correlogram of the LGN activity relative to the RGC spike times
 376 for negative time lags. Importantly, the time window in which the LGN cell could respond to a given RGC
 377 spike was not included; the model could only consider events preceding a retinal spike in predicting whether
 378 or not it was relayed.

379 To help prevent overfitting we introduce a Gaussian prior on the coefficients of the CH model (θ_C) to penalize
 380 large coefficient values (i.e., ridge regression), yielding a MAP estimator for θ_C :

$$381 \quad \mathcal{L}_{\text{MAP}}(\theta_C) = \mathcal{L}(\theta_C) - \eta \sum \theta_C^2 \quad (11)$$

382 where, as above, η is optimized separately for each pair (see [Hyperparameter optimization](#)).

383 As the CH model is an extension of the RH model, the time window spanned by \mathbf{X}_R was fixed, for each pair,
 384 at the value derived from RH model fitting (see [Retinal history models](#) above). The time window spanned
 385 by \mathbf{X}_L was optimized for each pair in an analogous manner (see [Hyperparameter optimization](#)).

386 In order to help mitigate the cost of increasing the number of free parameters (which would otherwise increase
 387 quite dramatically), for CH models \mathbf{X}_R and \mathbf{X}_L were represented in a basis of raised-cosine functions following
 388 common practice ([Pillow et al., 2005, 2008](#); [Ghanbari et al., 2017](#)):

$$389 \quad b_k(t) = \frac{\cos(q_k(t)) + 1}{2} \quad (12)$$

$$391 \quad q_k(t) = (\log(t + \Psi) - \log(\phi_k + \frac{\pi}{2\gamma})) \quad (13)$$

392 such that $q_k(t) \in [-\pi, \pi]$, where ϕ_k is the center of the “raised bump” of the k -th basis vector, Ψ is a constant
 393 hyperparameter (see [Hyperparameter optimization](#)) that controls the linearity of the spacing between bumps,
 394 and γ is a scaling factor that controls the width of the bumps such that they tile the time axis (i.e., γ is a
 395 function of the number of basis vectors and the duration they need to cover). This representation greatly
 400

401 reduces the number of parameters while still allowing good temporal resolution around the time of the retinal
 402 spike (by setting Ψ closer to ~ 1) at the cost of forcing the kernels to be smooth. However, this smoothness
 403 assumption is well supported by Usrey et al. (1998) and Rathbun et al. (2010) and loosely resembles, in its
 404 effects, the smoothing prior used in fitting RH models.

405 In practice \mathbf{X}_C is a m by $n_R + n_L + 1$ matrix where n_R and n_L are the number of basis vectors used to
 406 represent \mathbf{X}_R and \mathbf{X}_L respectively. Here, n_R was set to 16 and n_L was optimized separately for each pair
 407 (see [Hyperparameter optimization](#)).

408 In a manner analogous to thinking of \mathbf{X}_C as $[\mathbf{X}_R \ \mathbf{X}_L]$, we can separate the retinal and LGN filters learned
 409 by the model as $\boldsymbol{\theta}_C = [\boldsymbol{\theta}_R, \boldsymbol{\theta}_L]$ (ignoring the additive offset term). For clarity, throughout this paper when
 410 we refer to $\boldsymbol{\theta}_R$ (or $\boldsymbol{\theta}_L$) we are referring to $\boldsymbol{\theta}_R$ transformed back into the time-domain:

$$411 \quad \boldsymbol{\theta}_R = \mathbf{B}_R \boldsymbol{\theta}_R^* \quad (14)$$

$$413 \quad \mathbf{B}_R = [b_{R,1}(t) \ b_{R,2}(t) \ \dots \ b_{R,k}(t)] \quad (15)$$

414 where $b_{R,k}(t)$ is the k -th basis for \mathbf{X}_R (as above) and $\boldsymbol{\theta}_R^*$ are the coefficients on \mathbf{X}_R learned by the model.
 415 The analogous set of relations apply to $\boldsymbol{\theta}_L$, \mathbf{B}_L etc.

416
 417
 418 **Optimization error** For RH models, in which data used for fitting were represented in the standard
 419 temporal basis, the standard error of the estimate for each parameter was computed from the Hessian of the
 420 log-likelihood function ($\nabla^2 \mathcal{L}(\boldsymbol{\theta}_{ML})$) at the maximum likelihood estimate ($\boldsymbol{\theta}_{ML}$) following standard practice
 421 (Paninski, 2004; Truccolo et al., 2005; Paninski et al., 2007; Babadi et al., 2010).
 422

$$423 \quad \text{stderr}(\boldsymbol{\theta}_{ML}) = \text{diag}([\nabla^2 \mathcal{L}(\boldsymbol{\theta}_{ML})]^{-1})^{1/2} \quad (16)$$

424
 425
 426
 427 The standard error of parameter estimates for models fit to data represented in the raised-cosine basis are
 428 omitted from visualizations as the standard error cannot be validly transformed back into the time-domain
 429 as the parameter estimates can.

430
 431 **Hyperparameter optimization**

432
 433 Hyperparameter values that could be chosen based on the literature or reasonable assumptions (in cases
 434 where a hyperparameter has little impact on the model overall) were fixed for all pairs at the values specified

435 below. For quantifying model performance (Figures 4 & 5), all other hyperparameters were chosen for each
436 pair from a predefined set based on which value yielded the highest cross-validated $J_{\text{Bernoulli}}$ in a nested
437 cross-validation procedure. On each fold of the main cross-validation loop the training-set (consisting of 90%
438 of the total data from a pair) was further partitioned into sub-training and sub-testing sets (again a 90-10
439 split); the combination of model parameters and hyperparameters that yielded the highest cross-validated
440 $J_{\text{Bernoulli}}$ on the sub-testing set across sub-folds were then used to quantify the model's performance on the
441 main testing-set.

442
443 **ISI efficacy model** The maximum ISI for which the ISI-efficacy model would predict a value other than
444 the mean was chosen for each pair from a set of eight logarithmically spaced values (base 10) between 0.03
445 and 0.5 seconds. The standard deviation of the Gaussian kernel used to smooth ISI-efficacy functions was
446 chosen from a set of seven logarithmically spaced values (base 10) between 0.002 and 0.03 seconds and the
447 value 0 (i.e., no smoothing).

448
449 **RH model** The RH model contains two hyperparameters: the temporal span, which is the length of the
450 time window preceding each target retinal spike that is used to train the model, and the prior weighting term,
451 η , that controls the magnitude of the smoothness constraint (see [Retinal history models](#)). The temporal
452 span was chosen from a set of eight logarithmically spaced values (base 10) between 0.03 and 0.5 seconds
453 (rounded to the nearest millisecond). The prior weighting term was chosen from a set of five logarithmically
454 spaced values (base 2) between 4 and 4096.

455
456 **CH model** As the CH model is an augmented version of the RH model, the temporal span of the retinal
457 component of each pair's CH model was fixed at the value derived from RH model fitting. Thus, seven
458 hyperparameters remained: the temporal span of the LGN component, the number of basis vectors (in the
459 raised-cosine basis, see [Combined history models](#)) used to represent each component (one hyperparameter
460 per component), the weight given to the l^2 penalty for each component, and the linearity of the basis vector
461 spacing, Ψ , (see [Combined history models](#), again one per component). The temporal span of the LGN
462 component was chosen from a set of eight logarithmically spaced values (base 10) between 0.04 and 0.6
463 seconds (rounded to the nearest millisecond). The number of basis vectors for retinal components was fixed
464 at 16 for all pairs. For LGN components the number of basis vectors was chosen from the set {8, 12, 18,
465 24, 32}. The weight of the l^2 penalty was chosen from a set of five logarithmically spaced values (base 2)
466 between 0.125 and 8.0. The linearity of basis vector spacing, Ψ , was fixed at 10 and 8 for retinal and LGN

467 components respectively. The six-dimensional grid defined by the specified sets of values for the six non-fixed
468 hyperparameters was searched exhaustively.

469

470 **Filter visualization**

471

472 For visualizing (Figures 2 & 3) and analyzing (Figures 6-8) temporal profiles of the filters learned by the
473 models, data from all pairs were fit with a fixed set of hyperparameters: temporal span for both RH and
474 CH components was fixed at 200 ms, and the number of basis vectors in RH (CH) components was fixed
475 at 16 (24). As they more directly affect the shape of the learned filters, prior weighting terms were chosen
476 individually for each pair using ten-fold cross-validation from the same range specified in **Hyperparameter**
477 **optimization**. When displaying averaged filters for a population or condition filters for each pair were scaled
478 to have unit norm prior to averaging and, unless specified otherwise, error shading reflects the 95% confidence
479 interval of the mean (see **Statistics**).

480

481 **Burst spike definition**

482

483 Geniculate bursts were identified by the criteria established by (Lu et al., 1992): a geniculate burst must be
484 preceded by at least 100 ms of quiescence and contain two or more spikes each separated by no more than
485 4 ms. The relaxed definition reduced the quiescence duration to 50 ms and increased the maximum ISI to
486 6 ms (Figure 4C & D). Non-cardinal burst spikes were defined as all spikes that were part of an identified
487 burst, except the first or “cardinal” spike of each burst.

488

489 **Classification of retinal spikes by activity level**

490

491 In order to assess how the level of activity of the early visual network might alter the integration dynamics
492 of LGN cells we partitioned all retinal spikes from a given pair into four quartiles based on the LGN spike
493 count in a 100 ms window preceding each retinal spike. RH models were then fit separately to data from
494 each quartile. Differences between filters learned from data from distinct quartiles were quantified by taking
495 the integral of the absolute difference between the two filters: $\int |\theta_N - \theta_M|$. Where θ_N and θ_M are the filters
496 learned the N'th and M'th quartiles respectively. We refer to this metric as the “absolute difference” metric.

497 **Simulating GLMs**

498
499 Given a retinal spike train and a set of learned filter coefficients θ , a GLM can be used to simulate the relay
500 status of the retinal spike train by constructing a predictor matrix \mathbf{X} from the retinal spike train as described
501 above (see [Retinal history model](#)), multiplying \mathbf{X} by the learned coefficients and passing the result through
502 the logistic function $\sigma(\mathbf{X}\theta)$ to attain the predicted relay probability for each retinal spike. Relay status \mathbf{y} can
503 then be simulated by drawing a random number for each retinal spike from a uniform distribution on (0,1);
504 if the random number is less than the predicted probability for a given retinal spike the spike is considered
505 relayed (this is equivalent to flipping a coin whose probability of heads, or in this case “relayed”, is given
506 by the predicted relay probability). A GLM can then be fit to the retinal spike train and simulated relay
507 status just as is done for real data (see [Retinal history model](#)). Due to the stochastic nature of simulating
508 relay status, for all simulations presented here the final two steps (simulate relay status and fit GLM) are
509 repeated 50 times for each pair and the resulting coefficients are then averaged.

510

511 **Statistics**

512

513 Unless otherwise noted in the text, data are reported as the median (or paired median difference) and the
514 median absolute deviation (MAD) defined as: $\text{median}(|\mathbf{x} - \text{median}(\mathbf{x})|)$. Confidence intervals are derived
515 from bootstrap estimation with 5000 re-samples, and are bias corrected and accelerated ([Efron, 1987](#)) using
516 the Bootstrap.jl software package ([Gehring, 2020](#)). For the awake data set, the small sample size prevents
517 the valid use of bootstrap-derived confidence intervals; thus we report the range of values ([min, max])
518 instead. For model comparisons, p -values are calculated from paired samples permutation tests with 5000
519 re-samples, where the permutation is performed within pair. For example, if comparing model A to model
520 B, on each iteration the model performance values for each pair are randomly reassigned (i.e., \hl{swapped
521 or not between A and B with probability 0.5}) and the resulting paired median difference is calculated. After
522 5000 iterations the observed paired median difference is compared to the permuted differences distribution
523 to yield a p -value. Computed p -values are then corrected so that they cannot be exactly zero (which would
524 otherwise be possible given the discrete nature of the permuted differences distribution) using the method
525 proposed in Phipson and Smyth ([2010](#)).

526 Results

527

528 To investigate the factors that contribute to how the LGN filters retinal spike trains, we analyzed data from
529 45 monosynaptically connected RGC-LGN cell pairs from anesthetized cats and 8 pairs from awake cats. For
530 the recordings under anesthesia, neurons were stimulated with binary white noise (n=40) and/or drifting
531 sinewave gratings (n=33) and connectivity was assessed by cross-correlation of the spike times from the two
532 simultaneously recorded neurons. Figure 1 shows data from an example pair. The top row (Figure 1, A & B)
533 shows receptive field (RF) maps of the RGC (left) and LGN neuron (right) derived from the spike-triggered
534 average of the binary white-noise frames. The one standard deviation contour of a circularly symmetric
535 Gaussian fit to the LGN (RGC) RF is overlaid in white (black) on the RGC (LGN) RF, demonstrating
536 the high degree of spatial overlap between the two RFs. The bottom row (Figure 1, C & D) shows the
537 cross-correlograms, LGN spike times relative to each RGC spike, for the two stimulus conditions for this
538 pair. Using a monosynaptic latency derived from the time lag at which the cross-correlogram peaks, we
539 identified each retinal spike as being relayed (i.e., evoked a spike in its LGN partner) or not, and each LGN
540 spike as being triggered (i.e., was evoked by a RGC spike) or not. Retinal efficacy, the proportion of retinal
541 spikes that were relayed (see [Identification of monosynaptically connected pairs and relayed RGC spikes](#)), for
542 this example pair during binary white noise (drifting grating stimuli) was 0.316 (0.473); retinal contribution,
543 the proportion of LGN spikes that were triggered, was 0.812 (0.760). Across the population, for binary
544 white noise data median retinal efficacy was 0.097 (median absolute deviation (MAD) 0.070, 95% confidence
545 interval (CI) [0.054, 0.173]) and median retinal contribution was 0.247 (MAD 0.156, 95% CI [0.136, 0.394]);
546 for drifting grating data median retinal efficacy was 0.161 (MAD 0.098, 95% CI [0.088, 0.229]) and median
547 retinal contribution was 0.347 (MAD 0.197, 95% CI [0.205, 0.490]).

548 Additionally, we analyzed data from a smaller set of eight RGC-LGN cell pairs from awake cats in which the
549 spike train of the connected RGC was inferred from the presence of S-potentials that could be isolated, along
550 with the LGN cell's spikes, from the extracellular voltage trace recorded within the LGN ([Weyand, 2007](#)).
551 Across the population, median retinal efficacy was 0.519 (MAD 0.133, range [0.154, 0.724]) and median
552 retinal contribution was 0.935 (MAD 0.037, range [0.604, 0.997]).

553 Importantly, these data confirm the well documented finding that not every retinal spike is relayed by the
554 LGN (([Cleland et al., 1971](#); [Kaplan and Shapley, 1984](#); [Kaplan et al., 1987](#); [Usrey et al., 1998](#); [Sincich et al.,
555 2007](#); [Weyand, 2007](#)) among others). Taken together with the generally accepted notion that every non-burst
556 relay cell spike is triggered by the retina ([Kaplan and Shapley, 1984](#); [Sincich et al., 2007](#); [Weyand, 2007](#)),
557 this finding suggests that the primary role of LGN relay cells is to edit the incoming retinal spike train by

558 selective deletion. Thus, we sought to investigate the factors that determine which retinal spikes are relayed
559 and which are not, what we term “relay status”. Given this goal, we consider models of retinogeniculate
560 transmission that focus specifically on predicting the relay status of retinal spikes rather than trying to
561 predict the LGN spike train directly (i.e., we do not attempt to predict LGN spikes that were not triggered
562 by the recorded RGC).

563

564 **ISI efficacy model**

565

566 Previous work has clearly demonstrated that one of the primary factors that determines which retinal spikes
567 are relayed is the elapsed time since the last retinal spike (i.e., retinal interspike interval (ISI), (Usrey et al.,
568 1998; Carandini et al., 2007; Sincich et al., 2007; Casti et al., 2008; Sincich et al., 2009; Wang et al., 2010)).
569 This is often visualized by plotting retinal efficacy as a function of the preceding retinal ISI ((Usrey et al.,
570 1998), see *Inter-spike interval models*). Figure 2, left column, shows the ISI-efficacy relation for an example
571 pair of cells from the anesthetized data set (A, pair ID 208), the population as a whole (C, anesthetized
572 data set), and the relations for each pair in the awake data set (E), where the data from each pair in C
573 and E were normalized to their mean before averaging. The ISI-efficacy functions follow the typical decay
574 pattern (shorter ISIs in general show higher efficacies) that has been reported previously (Usrey et al., 1998;
575 Weyand, 2007; Rathbun et al., 2010). Interestingly, the drifting grating data (Figure 2C, red line) do show
576 a slight increase in efficacy for ISIs greater than 150 ms, potentially caused by the release from a slow acting
577 suppressive influence such as synaptic depression. Implicitly, the ISI-efficacy relation is a simple model for
578 predicting which retinal spikes were relayed based on the preceding retinal ISI (Wang et al., 2010), thus
579 we formalized the model to quantitatively assess its decoding performance. We utilized a 10-fold cross-
580 validation procedure in which ISI-efficacy functions were constructed using 90% of retinal spikes (training
581 set), and performance was assessed on the remaining 10% (test set) by looking up the expected efficacy of
582 each spike in the test set from the training-set-derived ISI-efficacy function (see *Inter-spike interval models*).
583 This procedure was repeated ten times such that each retinal spike was included in the test set once and
584 model performance was evaluated by the cross-validated, single-spike Bernoulli information ($J_{\text{Bernoulli}}$) which
585 quantifies how informative model predictions are about the relay status of test-set retinal spikes relative to a
586 homogeneous model that always predicts the mean efficacy (see *Assessing Model Performance*). For binary,
587 white-noise data, median $J_{\text{Bernoulli}}$ was 0.019 bits/spike (MAD 0.018, 95% CI [0.004, 0.041]). For drifting
588 grating data, median $J_{\text{Bernoulli}}$ was 0.026 bits/spike (MAD 0.017, 95% CI [0.010, 0.030]). For the awake
589 data set, median $J_{\text{Bernoulli}}$ was 0.177 bits/spike (MAD 0.085, range [0.075, 0.439]). This demonstrates that
590 the ISI-efficacy model was able to predict the relay status of retinal spikes significantly better than the

591 homogeneous model regardless of the stimulus condition or the state of the animal (anesthetized or awake).

592

593 **Retinal history model**

594

595 While retinal ISI is a strong predictor of relay status, its influence is a natural consequence of the temporal
596 integration that occurs within relay cells. This fact suggests that the history dependence of relay probability
597 is likely to extend beyond the most recent spike and might be better captured by considering all retinal
598 spikes that occur within a given window of time. Thus, we sought to extend the ISI-efficacy model by using
599 generalized linear models (GLM) to predict the relay status of retinal spikes based on the patterns of retinal
600 activity preceding each spike, what we call the retinal history (RH) model. Historically, GLMs have been
601 used to predict the activity of visual neurons based on the changing pattern of a visual stimulus (Chichilnisky,
602 2001; Paninski, 2004; Truccolo et al., 2005; Pillow et al., 2008; Babadi et al., 2010); here we instead use the
603 pattern of activity recorded simultaneously from a monosynaptic input (see **Generalized linear models**). In
604 brief, the GLM predicts the relay status of a retinal spike by convolving the pattern of recent activity with
605 a learned temporal filter, the output of which is then passed through a static nonlinearity to produce a relay
606 probability. Specifically, we use Bernoulli-Logistic GLMs (i.e., logistic regression) to model retinogeniculate
607 transmission as a binary parsing (Wang et al., 2010) or coin-flip process where the probability of a positive
608 outcome (relayed retinal spike) varies continuously over time as a function of the pattern of recent retinal
609 spikes (see **Retinal history models**).

610 Figure 2, right column, shows the temporal filters learned from drifting grating (red) and binary, white-noise
611 data (blue) for an example pair (Figure 2B, pair ID 208), the population recorded under anesthesia (Figure
612 2D), and the population recorded in the awake state (Figure 2F), where filters from each pair were scaled to
613 have unit norm before averaging in D and F. For visualization purposes, the time span preceding each retinal
614 spike that the model could consider (temporal span) was set to 0.2 seconds for all pairs (see [Visualizing
615 filters] and **Hyperparameter optimization**). Much like the ISI-efficacy functions, the temporal filters show
616 relatively large positive values in the time window just prior to the target retinal spike (at $t=0$), indicating
617 that retinal spikes falling within this time window increase the likelihood that the target retinal spike will
618 be relayed. Retinal spikes that occurred earlier relative to the target spike (> 0.02 - 0.04 seconds) were less
619 informative about relay status, as shown by the smaller magnitude of the filter values, and in general tended
620 to slightly decrease the probability that the target retinal spike would be relayed (i.e., filter values slightly $<$
621 0) for pairs recorded under anesthesia. Interestingly, the filters learned from drifting grating data tended to
622 have larger negative values during earlier pre-spike time windows ($> \sim 0.08$ - 0.18 seconds pre-spike) and show

623 a slight oscillation at ~10Hz, which is unlikely to be due solely to the periodic nature of the drifting grating,
624 which had a temporal frequency of 4Hz (see [Electrophysiological recording and visual stimuli](#)). As with the
625 ISI-efficacy model, the performance of the RH model was assessed using a train-on-90% test-on-10%, 10-fold
626 cross-validation procedure, in which overall performance was computed as the average $J_{\text{Bernoulli}}$ across folds.
627 For binary, white-noise data median $J_{\text{Bernoulli}}$ was 0.024 bits/spike (MAD 0.022, 95% CI [0.007, 0.042]). For
628 drifting grating data median $J_{\text{Bernoulli}}$ was 0.068 bits/spike (MAD 0.047, 95% CI [0.036, 0.099]). For the
629 awake data set $J_{\text{Bernoulli}}$ was 0.154 bits/spike (MAD 0.072, range [0.061, 0.452]).

630

631 **Combined history model**

632

633 Although the recorded RGC could account for the majority of LGN spikes in many cell pairs (i.e., a retinal
634 contribution > 0.5), a considerable number of LGN spikes could not be directly attributed to (i.e., were not
635 triggered by) the recorded RGC. These non-triggered spikes likely represent the activity of other RGC inputs
636 to the recorded relay cell ([Usrey et al., 1999](#)), and might provide a non-redundant source of information that
637 could aid predictions about which RGC spikes were relayed and which were not. Thus, we built an augmented
638 version of the RH model that included an additional filter that acted on the spiking history of the recorded
639 LGN relay cell, what we call the combined history (CH) model. Two attributes of this additional filter are
640 worth noting: 1) the activity of the LGN cell only contributes to the RH model by identifying which retinal
641 spikes were relayed. Thus, for any pair with a retinal contribution less than one, the LGN activity may
642 contribute additional information that the model can take advantage of, and 2) the LGN filter is aligned
643 relative to the time of the target retinal spike just as the retinal filter is, so only LGN spikes that occurred
644 prior to target retinal spike are included (see [Combined history models](#)). This construction is distinct from
645 those commonly used to represent spike history effects in GLM models ([Pillow et al., 2008](#); [Babadi et al.,](#)
646 [2010](#)) and reflects our focus on predicting the relay status of retinal spikes and not the activity of the LGN
647 cell per se. As a result, the LGN filter can capture some features of LGN activity, such as bursting in certain
648 circumstances, but not others, such as a refractory period, which is not relevant for predicting retinal relay
649 status.

650 Figure 3 shows the retinal (Figure 3A) and LGN (Figure 3B) filters for an example pair (pair ID 208) and the
651 population as a whole (Figure 3C & D, filters from each pair were scaled to have unit norm before averaging).
652 For visualization purposes, the temporal span of both retinal and LGN filters was set to 0.2 seconds for all
653 pairs (see [Filter visualization](#)). Two aspects of the filters learned by the CH model are worth noting. First,
654 the shape of the retinal filters are nearly identical to the shape of the retinal filters learned by the RH model

655 as expected (compare Figure 2D and Figure 3C), despite using far fewer parameters (see **Combined history**
656 **models**), suggesting that the addition of the LGN filter has not fundamentally changed how the model is
657 weighting retinal spikes in making predictions. Second, much like the retinal filters, the LGN filters show
658 a strong positive component immediately preceding the target spike that rapidly declines (~ 0.015 seconds)
659 followed by a lower amplitude negative component that decays to near zero fairly quickly for drifting grating
660 data (~ 0.04 seconds, Figure 3C red) and more slowly for binary white noise data (~ 0.1 seconds, Figure 3C
661 blue). The strong, positive weights assigned by the model to the time window immediately preceding the
662 target spike suggests that retinal spikes that follow LGN spikes at very short latencies are more likely to be
663 relayed. This pattern of LGN-RGC-LGN spiking is expected to be particularly likely when a retinal spike
664 arrives during a geniculate burst (Llinás and Jahnsen, 1982; Huguenard and McCormick, 1992; Alitto et
665 al., 2019b). To test whether this filter component was in fact due to LGN bursting, we repeated the CH
666 model fitting procedure after removing all non-cardinal burst spikes (i.e., removing all spikes that comprise
667 a burst except the first spike, see **Burst spike definition**). Interestingly, while the resulting filters do show a
668 strongly attenuated early positive component for the drifting grating data, removing all non-cardinal burst
669 spikes only minimally altered the LGN filters learned from binary white noise data (Figure 4B). However,
670 relaxing the definition of bursts somewhat to include more high-frequency events reduced the early positive
671 component for binary white noise data (Figure 4D), suggesting that the early positive component of LGN
672 filters may reflect both burst as well as high-frequency, non-burst events (Alitto et al., 2019b).

673 The retinal filters learned by the CH model from the awake data (Figure 3E) closely resembled those learned
674 from the anesthetized data, as expected from the RH model results (Figure 2, D & F). However, the LGN
675 filters learned from the awake data show a very different pattern. Instead of the short latency, positive
676 component that appears to be due in large part to LGN bursting (see above), the LGN filters for seven of
677 the eight pairs of the awake data set show a clear, negative component over the same time span ($\sim -0.03 - 0.0$
678 seconds preceding the target spike). Two aspects of this observation are worth noting. First, LGN cells in the
679 awake data set produced very few bursts. Averaged across pairs only 0.235% (median 0%, range [0.0, 1.52])
680 of LGN spikes were part of bursts, with five of the eight producing no bursts at all by the accepted definition
681 ((Lu et al., 1992), see **Burst spike definition**). In comparison, across pairs from the anesthetized data sets the
682 median percentage of spikes that were part of bursts was much higher: 14.203% (MAD 9.698, 95% CI [9.223,
683 17.496]) for the binary white noise data set, and 18.264% (MAD 14.805, 95% CI [9.792, 27.777]) for the
684 drifting grating data set. Thus, the lack of the positive component seen in the anesthetized data is expected.
685 Second, the negative component of the LGN filters suggests that some form of gain control or normalization
686 is occurring. This follows from the construction of the model, negative LGN filter weights over some time

687 interval indicate that LGN spikes that occur during that interval will push the model towards predicting that
688 the target spike will not be relayed, thus lowering the activity of the LGN cell itself and producing a gain
689 control or normalization-like effect (i.e., the same retinal input produces a smaller magnitude response when
690 the LGN has just been active compared with when it has just been quiescent (Shapley and Enroth-Cugell,
691 1984)).

692 As with previously discussed models, the performance of the CH model was assessed using 10-fold cross-
693 validation procedure. For binary white noise data median $J_{\text{Bernoulli}}$ across pairs was 0.033 bits/spike (MAD
694 0.030, 95% CI [0.015, 0.064]), and for drifting grating data median $J_{\text{Bernoulli}}$ was 0.073 bits/spike (MAD
695 0.051, 95% CI [0.051, 0.134]). For the awake data set, median $J_{\text{Bernoulli}}$ was 0.263 bits/spike (MAD 0.083,
696 range [0.086, 0.489]). Consistent with the idea that CH-LGN filters may be capturing the effect of LGN
697 bursts in the anesthetized data set, we observed that the gain in performance of CH models compared to
698 RH models across pairs was fairly well correlated with the “burstiness” of the LGN cell of each pair. The
699 Spearman correlation between $J_{\text{Bernoulli}}$ difference (CH - RH) and percentage of LGN spikes that were
700 part of bursts (not including cardinal spikes, see **Burst spike definition**) was 0.500 (95% CI [0.182, 0.726],
701 $p < 0.01$) for binary white noise data, and 0.286 (95% CI [-0.087, 0.569], $p \approx 0.1$) for drifting grating data
702 (Extended Data Figure 7-1C).

703

704 **Model comparison**

705

706 In order to illustrate how well each model performed relative to the others we first examined how well
707 the model-predicted efficacies correlated with the observed efficacies. To do this we grouped the retinal
708 spikes from each pair according to their predicted efficacy (normalized by the mean efficacy of that pair),
709 calculated the observed efficacy for each group (also normalized within-pair by that pair’s mean efficacy),
710 and then plotted the normalized, observed efficacy against the normalized, predicted efficacy. Efficacies, both
711 predicted and observed, for each pair were normalized by the observed mean efficacy of that pair (across all
712 spikes) to account for the large difference in efficacy across pairs as is typically done (e.g (Alitto et al., 2019a,
713 2019b)). In such a framework, a well performing model will produce a “unity” line with a slope of one and
714 y-intercept of zero (i.e., predicted efficacy and observed efficacy match). The left column of Figure 5 shows,
715 for each data set, the median relationship between observed and predicted efficacy for each model (error
716 bars represent the MAD across pairs). While all models appear to perform quite well within this framework,
717 there is a systematic trend for the ISI-efficacy model to perform worse for the spikes that it predicts to have
718 the highest efficacy within the drifting grating and binary white noise data sets. Given that the highest

719 efficacy spikes should follow short ISIs (see e.g., Figure 2), this suggests that the ISI-efficacy model may be
720 performing worse than the GLMs specifically for short ISI spikes. Consistent with this suggestion, the right
721 column of Figure 5 shows that the difference in $\mathcal{J}_{\text{Bernoulli}}$ between the GLM and ISI-efficacy models is most
722 pronounced for retinal spikes with the shortest ISIs within the drifting grating and binary white noise data
723 sets. Interestingly, within this comparison framework the ISI model appears to perform as well as the GLMs
724 on the awake data set.

725 While Figure 5 provides a helpful overview of model performance, given the present context the most rigorous
726 way to assess the performance of the models presented here is using $\mathcal{J}_{\text{Bernoulli}}$, the cross-validated single-spike
727 Bernoulli information, which quantifies the accuracy of model predictions on a spike-by-spike basis. Figures
728 6 and 7 summarize the results of a direct model comparison analysis for the binary white noise and drifting
729 grating data respectively, in which all hyperparameters for all models were optimized individually for each
730 pair (see [Hyperparameter optimization](#)). The top row of each figure shows the cross-validated $\mathcal{J}_{\text{Bernoulli}}$ for
731 each pair and each model, where points corresponding to the same pair are connected, and the bottom row
732 shows a bootstrap estimation of the paired median difference in $\mathcal{J}_{\text{Bernoulli}}$ between models (see [Statistics](#)).
733 For the binary white noise data (Figure 6), the paired median difference between ISI-efficacy and RH models
734 was 0.002 bits/spike (MAD 0.003 95% CI [0.000, 0.003], $p \approx 0.0092$)^a, between ISI-efficacy and CH models
735 was 0.009 bits/spike (MAD 0.008 95% CI [0.004, 0.015], $p \approx 0.0002$)^b, and between RH and CH models was
736 0.004 bits/spike (MAD 0.004 95% CI [0.003, 0.009], $p \approx 0.0002$)^c. For the drifting grating data (Figure 7),
737 the paired median difference between ISI-efficacy and RH models was 0.030 bits/spike (MAD 0.020 95% CI
738 [0.012, 0.047], $p \approx 0.0002$)^d, between ISI-efficacy and CH models was 0.049 bits/spike (MAD 0.033 95% CI
739 [0.032, 0.080], $p \approx 0.0002$)^e, between RH and CH models was 0.020 bits/spike (MAD 0.014 95% CI [0.006,
740 0.027], $p \approx 0.0002$)^f.

741 While the small size of the awake data set precludes a statistical comparison of model performance, a
742 qualitative assessment shows largely the same pattern as seen in the anesthetized data. Extended Data
743 Figure 7-2A illustrates the pairwise difference in model performance between the three models (ISI, RH
744 and CH) which suggests that although no difference between the performance of the ISI and RH models is
745 evident, the inclusion of the LGN filter in the CH model may substantially improve performance (median
746 pairwise difference in $\mathcal{J}_{\text{Bernoulli}}$ between RH and CH models was 0.058 bits/spike, range [-0.004, 0.170]).

747 Overall, while RH models do show significantly better performance than ISI-efficacy models, and CH models
748 significantly outperform RH models, the magnitude of the performance gain is rather modest, suggesting
749 that, overall, retinal ISI is the dominant factor in determining which retinal spikes are relayed. However,
750 while both stimulus conditions showed this trend, the magnitude of the performance gain associated with RH

751 and CH models over the ISI-efficacy model was substantially larger when pairs were stimulated with drifting
752 gratings, suggesting that some subtler aspects of LGN integration may differ between the two stimulus
753 conditions (e.g., Figure 2D & 3C, (Usrey et al., 1998)).

754

755 **Integration dynamics depend on firing rate**

756

757 One potential drawback of using GLMs in the present context is that by fitting a single set of filters to
758 all spikes (or a random subset), we are asking the fitting algorithm to find what amounts to the average
759 integration behavior of relay cells during the recording period. The analysis is, by design, insensitive to any
760 changes in relay cell integration that may occur within a stimulus condition. While this implicit assumption
761 of stationarity may be largely valid for the binary white noise stimulus, it may not hold during drifting
762 grating stimulation due to the high degree of spatial and temporal correlations present in drifting gratings,
763 which are of course absent from the binary white noise. The strong correlations present in drifting gratings
764 may result in larger fluctuations in activity for both the RGC-LGN cell pair being recorded as well as the
765 wider network (including e.g., the thalamic reticular nucleus, V1, etc.) and thus may alter LGN integration
766 dynamics in a more significant manner. Consistent with this idea, we observed higher RGC firing rate
767 variability during drifting grating stimulation in the 200 ms period immediately preceding each retinal spike
768 (the same time period that the model could consider): median pairwise difference in firing rate standard
769 deviation (gratings minus binary white noise) was 3.549 spikes/second (MAD 4.581, 95% CI [0.152, 5.844])
770 (median 16.768 and 11.587 spikes/second for gratings and binary white noise respectively). The models
771 presented thus far are not sensitive to these potential within-condition changes, as each model is fit to all
772 spikes (or a randomly selected subset) from a single stimulus condition. Thus, we sought to investigate
773 specifically whether LGN integration dynamics might differ based on the level of activity by assigning each
774 retinal spike to one of four “quartile” subsets (Q1 - Q4) of the data based on the quartile into which the
775 LGN spike count in a 100 ms window preceding each retinal spike fell (see [Classification of retinal spikes by](#)
776 [activity level](#)). We then fit separate GLMs to the data from each quartile for each stimulus type. For this
777 analysis we consider only RH models, as the quartile partitioning results in too few LGN spikes in the lowest
778 activity quartile to reliably fit CH models. Additionally, the binary white noise data from one pair (pair
779 ID 102) did not contain enough spikes to reliably fit RH models for each quartile and was excluded from
780 activity level analyses. Figure 8 shows the filters learned by the model for each activity level and stimulus
781 condition averaged across pairs (filters from each pair were scaled to have unit norm before averaging) where
782 the shaded regions represent the 95% confidence interval (CI) across pairs (see [Filter visualization](#)). For
783 binary white noise (Figure 8A), there is an apparent trend towards a small difference between approximately

784 40 and 120 ms preceding the target spike (at time=0) such that retinal spikes during that window may have
785 a somewhat stronger negative influence on relay probability (i.e., push the model to predict “not relayed”)
786 during epochs of heightened activity (Q3 and Q4); however, the magnitude and variability of this effect
787 (as seen in the overlapping CI shading) suggest little qualitative difference between activity levels. On the
788 other hand, filters learned from drifting grating data show a much clearer difference between activity levels,
789 specifically within a time window approximately 5 ms to 20 ms before the target retinal spike (Figure 8C and
790 inset), such that the filters learned from high activity data (Q3 & Q4) show a faster decay towards zero from
791 the initial positive peak immediately preceding the target retinal spike. This difference suggests a narrowing
792 of the effective integration window of LGN relay cells during epochs of elevated activity. Importantly, this
793 difference is unlikely to be due to differences in the ability of the model to fit the different data sets (Figure
794 8D), as median paired difference in $J_{\text{Bernoulli}}$ between models fit to data from the highest (Q4) and lowest
795 (Q1) activity levels was -0.005 bits/spike (Q4 - Q1, MAD 0.041, 95% CI [-0.047, 0.003], $p \approx 0.353$)^g. Model
796 performance was also not significantly different between Q4 and Q1 subsets for the binary white noise data
797 set: median paired difference in $J_{\text{Bernoulli}}$ was 0.001 bits/spike (MAD 0.008, 95% CI [-0.001, 0.004], $p \approx$
798 0.396)^h.

799 One potential concern with the above analysis is that the data used to train the model differed considerably
800 between quartiles. Although the quartiles are defined based on LGN firing rates, retinal firing rates will of
801 course be highly correlated. Thus, the observed difference in LGN integration dynamics could be due entirely
802 to differences in the training data. To control for this possibility we use a single, fixed filter learned from all
803 the data from a given pair (i.e., the filters shown in Figure 2) to simulate the relay status of each retinal spike
804 (i.e., the pattern of retinal spikes preceding each target spike is convolved with the learned filter, the output
805 of which is passed through the logistic function and relay status is determined by a coin flip, see [Simulating](#)
806 [GLMs](#)). We then performed the quartile subsetting and model fitting exactly as for Figure 8. The learned
807 filters for each stimulus type and activity quartile are shown in Extended Data Figure 8-1. Importantly,
808 in this case the training data have exactly the same quartile related differences as for the original analysis,
809 the only difference is that the integration dynamics of the LGN cell are fixed via the simulation. Thus, the
810 fact that the filters learned from all the quartile subsets are highly overlapping suggests that the differences
811 observed in Figure 8C are not due to differences in the training data alone. The overlap in the learned filters
812 is especially apparent through the first ~30-50 ms where the most striking difference in Figure 8C can be
813 seen.

814 The finding that LGN integration dynamics depend on firing rate proved to be robust to the precise time
815 window used to classify activity levels (tested over a range spanning 50 - 200 ms, see Extended Data Figure

816 8-2C-D); however, using time windows close to the cycle duration of the drifting grating (i.e., around 250
817 ms) is likely to produce a severe underestimate of the real difference as it would effectively average over the
818 preferred and non-preferred phases of the drifting grating (which is the likely cause of the higher variability
819 in firing rate observed during drifting grating stimulation). Consistent with this idea, repeating the analysis
820 using a 250 ms time window to partition the data into quartiles substantially reduced the difference between
821 filters learned from the drifting grating data (filters learned from binary white noise data continued to show
822 no difference, see Extended Data Figure 8-2A-B).

823 The awake data set did not contain a sufficient number of spikes to perform the quartile subsetting procedure
824 that we used for the anesthetized data set (median number of retinal spikes per-pair in the awake data set was
825 2,017.0 (MAD 427.5), while anesthetized data sets had a median of 12,303.5 (MAD 5,624.0) and 38,425.0
826 (MAD 18,150.0) for the binary white noise and drifting grating data sets respectively). Thus, we used a
827 median split to assign each retinal spike from each pair to a low or high activity subset. The filters learned
828 from low and high subsets showed little difference (Extended Data Figure 7-2B), similar to what was seen
829 in the binary white noise (anesthetized) data although lacking the prolonged negative component (between
830 approximately -90 to -120 ms). Interestingly, the one pair that does appear to show a more substantial
831 difference between filters learned from low and high activity data (pair 200001250) was stimulated with
832 gratings during recording (see [Discussion](#)).

833 To quantify the apparent differences in filters learned from the highest (Q4) and the lowest (Q1) activity
834 data (Figure 8C), we calculated the integral of the absolute difference between the Q1 and Q4 filters for
835 each pair (see [Classification of retinal spikes by activity level](#)). The distribution of the paired absolute
836 differences, along with kernel density estimates, for each stimulus condition are shown in Figure 9A with
837 the corresponding estimation of the median of each distribution show in Figure 9B. For the binary white
838 noise data set the median absolute difference between Q4 and Q1 was 0.024 (MAD 0.010, 95% CI [0.018,
839 0.028]), and for drifting grating it was 0.055 (MAD 0.026, 95% CI [0.033, 0.067]). For the filters learned from
840 simulated data (Extended Data Figure 8-1), the median absolute difference was 0.002 (MAD 0.001, 95% CI
841 [0.002, 0.003]) and 0.005 (MAD 0.003, 95% CI [0.003, 0.007]) for binary white noise and drifting grating
842 data respectively.

843 A paired permutation test including only pairs for which both binary white noise and drifting grating
844 data were available (N=27) confirmed the differences between the two stimulus conditions: paired median
845 difference (drifting gratings minus binary white noise) in absolute difference was 0.031 (MAD 0.016 95% CI
846 [0.018, 0.039], $p \approx 0.0002$)ⁱ. Repeating the analysis when only including the 30 ms period preceding the
847 target retinal spike yielded similar results (paired median difference of 0.008, MAD 0.007 95% CI [0.002,

848 0.012], $p \approx 0.0004$).

849 Discussion

850

851 The aim of this study was to investigate how LGN relay cells integrate their retinal inputs over time, and
852 how the integration process changes under different stimulus and network conditions, by using computational
853 models to predict which retinal spikes were relayed on to V1 and which were not. We model retinogeniculate
854 transmission as a coin flip (or Bernoulli) process where the primary quantity of interest is the probability,
855 p , that each incoming retinal spike will be relayed. In the simplest possible model p is a constant given by
856 the mean efficacy across all retinal spikes recorded from a given RGC-LGN cell pair. This constant p model
857 (or homogeneous Bernoulli model) forms the basis of comparison for all other models that we considered, as
858 the constant p model captures the fact that as mean efficacy approaches the extremes (0 or 1) predicting
859 relay status becomes trivial (simply guessing the mean will approach perfect performance). Thus, we chose
860 to quantify model performance in terms of the cross-validated single-spike Bernoulli information ($J_{\text{Bernoulli}}$)
861 which quantifies how informative model predictions are about the relay status of retinal spikes (that were
862 not “seen” during model fitting) relative to a homogeneous model. In our construction, $J_{\text{Bernoulli}}$ has units
863 of bits/spike and can take on values between ~ 0 and 1, where 0 represents performance no better than a
864 constant p model and 1 represents perfect performance (see [Assessing model performance](#) for details).

865 The fact that $J_{\text{Bernoulli}}$ quantifies model performance relative to a homogeneous model is critical given the
866 present context of trying to predict the relay status of retinal spikes. This follows from the fact that the
867 difficulty of predicting relay status varies with mean efficacy: relay status is trivially easy to predict for pairs
868 with a mean efficacy close to zero or one, and is maximally difficult for pairs with a mean efficacy of 0.5.
869 Thus, an optimal performance metric needs to take into account both the quality of the predictions as well
870 as the difficulty of the task for a given pair. $J_{\text{Bernoulli}}$ does exactly this. However, as a result the maximum
871 $J_{\text{Bernoulli}}$ achievable for pairs with very low or very high mean efficacy is substantially less than one. This
872 fact accounts in part for the low $J_{\text{Bernoulli}}$ values achieved by the models considered here, especially on the
873 anesthetized data sets where many pairs have low mean efficacies (9% and 25% of pairs from the drifting
874 grating and binary white noise data sets, respectively, have a mean efficacy less than 0.05). It should be
875 noted that this behavior is not a deficiency in the $J_{\text{Bernoulli}}$ metric, rather it reflects an inherent difficulty in
876 predicting relay status.

877 We first considered a model where p varies in time according to the elapsed interval since the last retinal
878 spike (ISI) based on extensive evidence that retinal spikes following shorter ISIs are more likely to be relayed
879 due to temporal summation ([Usrey et al., 1998](#); [Sincich et al., 2007](#); [Weyand, 2007](#); [Casti et al., 2008](#); [Sincich
880 et al., 2009](#); [Wang et al., 2010](#); [Rathbun et al., 2016](#); [Alitto et al., 2019a](#)). Following the framework of Wang

881 et al. (2010), we formalize this observation as a simple model, $p = f(ISI)$, where the relation between ISI
882 and relay probability (i.e., efficacy) is learned from a subset of the data (training set) and the performance
883 of the model is tested on a separate subset (testing set, see [Assessing model performance](#)).

884 We further considered a model where p is a function of the pattern of retinal spikes that an LGN cell receives
885 within a given window of time, what we call the retinal history (RH) model. Conceptually, this can be seen
886 as an extension of the ISI-efficacy model that additionally takes into account the notion that the influence
887 of retinal activity on the current state of a relay cell (i.e., its propensity to relay a retinal input should one
888 arrive) is unlikely to be limited to just the most recent retinal spike. Thus, allowing a model to consider the
889 full pattern of recent spikes from the recorded RGC should improve predictions of relay status and provide
890 a less constrained view of the temporal integration dynamics of retinogeniculate interactions. To that end
891 we utilized Bernoulli-Logistic generalized linear models to predict the relay status of each retinal spike based
892 on the convolution of a learned temporal filter (retinal filter) with the pattern of recent retinal activity, the
893 output of which is then mapped to a predicted relay probability (or equivalently, predicted efficacy).

894 In comparing the parameters learned by the ISI-efficacy and RH models, one critical difference between the
895 models is worth nothing. For the ISI-efficacy model, relay probability is modeled as a univariate, nonlinear
896 function of ISI, while the RH model is a linear function of the multivariate pattern of retinal spikes over
897 a given time window (which is then passed through a logistic nonlinearity). Thus, the similarity of the
898 ISI-efficacy functions and RH retinal filters presented in Figure 2 should be interpreted carefully. However,
899 the rapid decay of both functions does tell a consistent story, namely that the time windows over which
900 retinal spikes positively interact (i.e., promote a relay probability above the mean) is approximately 20 - 30
901 ms regardless of the stimulus (gratings or binary white noise) or the state of the animal (anesthetized or
902 awake). This likely accounts for the observation that the RH model only outperforms the ISI-model by the
903 smallest of margins in the anesthetized data (Figures 6 & 7), and not at all in the awake data (though the
904 small size of the awake data set should be noted).

905 The final model that we considered was a further augmented version of the RH model that included a second,
906 learned temporal filter (LGN filter) that operated on the recent activity history of the LGN cell, what we
907 call the combined history (CH) model. As stated previously, for RH models the LGN activity is only used to
908 identify the relay status of each RGC spike, and thus the LGN spike train (and, in particular the LGN spikes
909 not triggered by the recorded RGC) may provide additional information that can help predict the relay
910 status of retinal spikes. While the CH model did outperform the other two models for all data sets tested
911 here, further analysis of the correlates of performance and consideration of the shape of the learned filters
912 suggests that the improvement may be based on different features within the anesthetized and awake data

913 sets. In particular we found that, for the anesthetized data set the improvement in performance between RH
914 and CH models was correlated with the degree of “burstiness” (i.e., the percentage of LGN spikes that were
915 part of bursts) of the LGN cells of the pairs (Extended Data Figure 7-2). Furthermore, the shape of the LGN
916 filters, large positive values at very short pre-target-spike latencies, suggests that the model is capturing the
917 increase in retinal efficacy that occurs during geniculate bursts (Alitto et al., 2019b), and this component
918 of the LGN filters was specifically attenuated when non-cardinal burst spikes were removed from the data
919 prior to CH model fitting (Extended Data Figure 7-1). In contrast, the LGN filters learned from the awake
920 data cannot be accounted for by bursts, as burst were extremely rare in the awake data set. Instead, the
921 negative component seen between ~40 and 0 ms (Figure 3) likely reflects the influence of a gain control
922 or normalization mechanism that could result from intrathalamic negative feedback through the thalamic
923 reticular nucleus (TRN) (or perhaps the longer LGN → V1 → TRN → LGN loop). Across the analyses that
924 we performed, this was the only clear difference between the awake and anesthetized data sets.

925 Lastly, we asked whether relay cell temporal integration dynamics might differ depending on the level of
926 activity within the retinogeniculate circuit, and whether that difference is seen for both stimulus conditions
927 in the anesthetized data. To that end we assigned each retinal spike to one of four data subsets based on the
928 quartile of LGN activity during the preceding 100 ms (see [Classification of retinal spikes by activity level](#))
929 and fit RH models separately to each data subset. We specifically chose to use LGN activity to partition
930 retinal spikes as, although retinal and geniculate activity levels are highly correlated, LGN activity is likely
931 to be more indicative of the activity level of the wider retino-thalamo-cortical circuit. For binary white
932 noise data, learned temporal filters showed little difference between subsets (Figure 7A), while for drifting
933 grating data a substantial difference is observed between approximately 5 and 20 ms (Figure 7C) such that
934 filters learned from the highest activity subsets (Q3 and Q4) show a shorter effective temporal integration
935 window (i.e., the duration of time preceding a target spike where the arrival of another retinal spike will
936 increase the likelihood that the target spike is relayed). For the awake data set, most pairs showed little
937 difference between epochs of higher and lower activity when analyzed in a similar manner (albeit using a
938 simpler median split as there were not enough spikes to reliably fit model to quartile subsets). Interestingly,
939 the one apparent exception (pair 200001250, Extended Data Figure 7-2B) was also the only pair that was
940 stimulated with gratings during recordings. While this is a single example and so should be considered only
941 the slimmest of evidence, it is nonetheless consistent with the idea that the effective integration window of
942 LGN cells, in both the awake and anesthetized states, is dynamically regulated in a manner that is inversely
943 proportional to the ongoing firing rate (i.e., shorter integration windows during periods of higher activity).
944 While there are several cellular and circuit mechanisms that could underlie the shortening of the temporal

945 integration window, such as spike rate adaptation within relay cells, short-term depression at the retinogeniculate synapse, feedforward inhibition from geniculate interneurons, feedback inhibition (direct or indirectly
946 via cortex) from the thalamic reticular nucleus, or a change in oscillatory activity coming from the retina
947 (Koeppell et al., 2009), the functional consequence of this process is a form a gain control wherein the specificity of geniculate filtering scales with activity level. The idea being that, under lower levels of activity the
948 LGN behaves more permissively and relays patterns of retinal spikes that under higher activity conditions,
949 where the LGN is less permissive, would not be relayed. This process might offer an explanation for several
950 observations about retinogeniculate transmission, such as the finding from Alitto et al. (2019a) that retinal
951 efficacy following ISIs in the ~5 to ~25 ms range is higher under low contrast (and thus low activity) than
952 high contrast (and thus high activity) stimulus conditions. Likewise it could potentially explain the finding
953 from Rathbun et al. (2016) that as the contrast of a drifting grating stimulus increases, responses of LGN
954 cells shift to progressively earlier phases of the stimulus cycle and that the rate of this “phase advance” is
955 higher in relay cells compared to their direct retinal inputs. Further work is needed to address whether the
956 magnitude of the integration window shortening that we observe here quantitatively matches the observations
957 listed above.

960

961 **Relationship to previous work**

962

963 A considerable amount of effort has been put into modeling the computations performed by relay cells of the
964 LGN, due in large part to the fact that simultaneous recordings of both a dominant input (from RGCs) and
965 the output (LGN spiking) is possible. Prior work on modeling retinogeniculate interactions can be coarsely
966 grouped into two approaches: those that focus on LGN processing of retinal spike trains in the absence
967 (Casti et al., 2008; Heiberg et al., 2013), or presence (Norheim et al., 2012) of extra-retinal input, and those
968 that include an additional channel for processing the visual stimulus directly (Babadi et al., 2010; Butts
969 et al., 2016). The logic of including the additional stimulus channel is that it enables models to capture
970 stimulus driven effects that are not mediated by the direct retinal input, so that “indirect” effects (e.g., from
971 cortical or TRN feedback) might be uncovered. While this is a powerful approach to studying geniculate
972 computations generally, we instead chose to focus our efforts more narrowly on modeling how LGN cells
973 process individual retinal inputs by trying to predict which retinal spikes were relayed and which were not.
974 This approach is particularly well suited to our data, which consists primarily of recordings of RGC-LGN cell
975 pairs in which the RGC spikes were recorded within the eye. This entails that 1) we can be confident that
976 few, if any, RGC spikes went undetected, and 2) that most of our recordings were made from non-dominant
977 RGC inputs. The second point follows from the observation that most relay cells in the cat receive input

978 from two to five RGCs (Cleland et al., 1971; Hamos et al., 1987; Usrey et al., 1999; Martinez et al., 2014),
979 and thus landing an extracellular electrode in the vicinity of the dominant input should be somewhat rare.
980 Conversely, S-potential recordings are likely to reflect just the dominant input (Kaplan and Shapley, 1984;
981 Weyand, 2007). Consistent with this idea, we observed considerably higher mean efficacies in the awake data
982 set (on average ~ 0.52) compared to either the drifting grating (~ 0.16) or binary white noise (~ 0.1) data sets
983 from the anesthetized animal. Given the above, we reasoned that the most fruitful approach would be to
984 focus on predicting the relay status of the retinal spikes that we did record and avoid making predictions
985 about LGN spikes that were not triggered by the RGC under study.

986 Overall, this approach emphasizes the computations being performed by relay cells on individual retinal
987 inputs. Previous work has proposed that the core of these computations is well approximated by linear
988 filtering with an exponential kernel (Casti et al., 2008; Heiberg et al., 2013) as suggested by the strong
989 relationship between retinal efficiency and retinal ISI (Usrey et al., 1998; Carandini et al., 2007; Sincich et
990 al., 2007; Casti et al., 2008; Sincich et al., 2009; Uglesich et al., 2009; Rathbun et al., 2010; Wang et al.,
991 2010). The strength of taking a statistical approach, as we do here, is that the form of the linear filter is
992 directly learned by the model. Our results confirm that an exponential filter is indeed a good model of relay
993 cell temporal integration and, given the relatively short apparent time constants (on the order of 10 - 20ms,
994 consistent with Casti et al. (2008)), suggest that the retinal ISI is likely to be the strongest single influence
995 on whether a given retinal spike is relayed or not.

996

997 **Conclusion**

998

999 Overall, our results suggest that the dominant factor that determines whether or not a given RGC spike is
1000 relayed to cortex by the LGN is the retinal ISI, confirming previous findings (Usrey et al., 1998; Carandini
1001 et al., 2007; Sincich et al., 2007; Casti et al., 2008; Sincich et al., 2009; Uglesich et al., 2009; Rathbun et
1002 al., 2010; Wang et al., 2010). However, quantitatively smaller, yet still likely important, contributions were
1003 observed for retinal activity further into the past, as well as LGN activity patterns indicative of periods
1004 of burst firing. Furthermore, we have demonstrated that the time scale over which the LGN integrates its
1005 retinal inputs changes as a function of the level of activity within the retino-thalamo-cortical circuit. This
1006 finding raises the possibility that gain control (Shapley and Enroth-Cugell, 1984), a core visual function of
1007 the LGN (Alitto et al., 2019a), could be achieved in part by modulating the temporal integration window of
1008 LGN relay cells. The source of this modulation remains an open question for future work to explore.

1009 **References**

- 1010 Alitto HJ, Rathbun DL, Fisher TG, Alexander PC, Usrey WM (2019a) Contrast gain control and retino-
1011 geniculate communication. *European Journal of Neuroscience* 49:1061–1068.
- 1012 Alitto HJ, Rathbun DL, Vandeleeest JJ, Alexander PC, Usrey WM (2019b) The Augmentation of Retino-
1013 geniculate Communication during Thalamic Burst Mode. *Journal of Neuroscience* 39:5697–5710.
- 1014 Babadi B, Casti A, Xiao Y, Kaplan E, Paninski L (2010) A generalized linear model of the impact of direct
1015 and indirect inputs to the lateral geniculate nucleus. *Journal of Vision* 10:22.
- 1016 Bezanson J, Edelman A, Karpinski S, Shah VB (2017) Julia: A Fresh Approach to Numerical Computing.
1017 *SIAM Review* 59:65–98.
- 1018 Butts DA, Cui Y, Casti ARR (2016) Nonlinear computations shaping temporal processing of precortical
1019 vision. *Journal of Neurophysiology* 116:1344–1357.
- 1020 Carandini M, Horton JC, Sincich LC (2007) Thalamic filtering of retinal spike trains by postsynaptic sum-
1021 mation. *Journal of Vision* 7:20.
- 1022 Casti A, Hayot F, Xiao Y, Kaplan E (2008) A simple model of retina-LGN transmission. *Journal of Com-
1023 putational Neuroscience* 24:235–252.
- 1024 Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. *Network (Bristol, England)*
1025 12:199–213.
- 1026 Cleland BG, Dubin MW, Levick WR (1971) Simultaneous recording of input and output of lateral geniculate
1027 neurones. *Nature: New Biology* 231:191–192.
- 1028 Efron B (1987) Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association*
1029 82:171–185.
- 1030 Fisher TG, Alitto HJ, Usrey WM (2017) Retinal and Nonretinal Contributions to Extraclassical Surround
1031 Suppression in the Lateral Geniculate Nucleus. *Journal of Neuroscience* 37:226–235.
- 1032 Gehring J (2020) [Juliangehring/Bootstrap.jl](#).
- 1033 Ghanbari A, Malyshev A, Volgushev M, Stevenson IH (2017) Estimating short-term synaptic plasticity from
1034 pre- and postsynaptic spiking. *PLOS Computational Biology* 13:e1005738.
- 1035 Hamos JE, Van Horn SC, Raczkowski D, Sherman SM (1987) Synaptic circuits involving an individual
1036 retinogeniculate axon in the cat. *The Journal of Comparative Neurology* 259:165–192.
- 1037 Heiberg T, Kriener B, Tetzlaff T, Casti A, Einevoll GT, Plesser HE (2013) Firing-rate models capture
1038 essential response dynamics of LGN relay cells. *Journal of Computational Neuroscience* 35:359–375.
- 1039 Huguenard JR, McCormick DA (1992) Simulation of the currents involved in rhythmic oscillations in thalamic
1040 relay neurons. *Journal of Neurophysiology* 68:1373–1383.

- 1041 Hunter JD (2007) Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering* 9:90–95.
- 1042 Johnson SG (2020) JuliaPy/PyPlot.jl.
- 1043 Kaplan E, Purpura K, Shapley RM (1987) Contrast affects the transmission of visual information through
1044 the mammalian lateral geniculate nucleus. *The Journal of Physiology* 391:267.
- 1045 Kaplan E, Shapley R (1984) The origin of the S (slow) potential in the mammalian Lateral Geniculate
1046 Nucleus. *Experimental Brain Research* 55:111–116.
- 1047 Koepsell K, Wang X, Vaingankar V, Wei Y, Wang Q, Rathbun DL, Usrey WM, Hirsch JA, Sommer FT
1048 (2009) Retinal oscillations carry visual information to cortex. *Frontiers in Systems Neuroscience* 3:1–18.
- 1049 Llinás R, Jahnsen H (1982) Electrophysiology of mammalian thalamic neurones in vitro. *Nature* 297:406–408.
- 1050 Lu SM, Guido W, Sherman SM (1992) Effects of membrane voltage on receptive field properties of lat-
1051 eral geniculate neurons in the cat: Contributions of the low-threshold Ca²⁺ conductance. *Journal of*
1052 *Neurophysiology* 68:2185–2198.
- 1053 Martinez LM, Molano-Mazón M, Wang X, Sommer FT, Hirsch JA (2014) Statistical Wiring of Thalamic
1054 Receptive Fields Optimizes Spatial Sampling of the Retinal Image. *Neuron* 81:943–956.
- 1055 Mastronarde DN (1987) Two classes of single-input X-cells in cat lateral geniculate nucleus. II. Retinal
1056 inputs and the generation of receptive-field properties. *Journal of Neurophysiology* 57:381–413.
- 1057 Mogensen P, Riseth A (2018) Optim: A mathematical optimization package for Julia. *Journal of Open*
1058 *Source Software* 3:615.
- 1059 Nelder JA, Wedderburn RWM (1972) Generalized Linear Models. *Journal of the Royal Statistical Society*
1060 *Series A (General)* 135:370–384.
- 1061 Nocedal J, Wright S (2006) Numerical optimization. Springer Science & Business Media.
- 1062 Norheim ES, Wyller J, Nordlie E, Einevoll GT (2012) A minimal mechanistic model for temporal signal
1063 processing in the lateral geniculate nucleus. *Cognitive Neurodynamics* 6:259–281.
- 1064 Paninski L (2004) Maximum likelihood estimation of cascade point-process neural encoding models. *Network:*
1065 *Computation in Neural Systems* 15:243–262.
- 1066 Paninski L, Pillow J, Lewi J (2007) Statistical models for neural encoding, decoding, and optimal stimulus
1067 design. *Progress in Brain Research* 165:493–507.
- 1068 Phipson B, Smyth GK (2010) Permutation P-values Should Never Be Zero: Calculating Exact P-values
1069 When Permutations Are Randomly Drawn. *Statistical Applications in Genetics and Molecular Biology*
1070 9.
- 1071 Pillow JW, Paninski L, Uzzell VJ, Simoncelli EP, Chichilnisky EJ (2005) Prediction and Decoding of Retinal
1072 Ganglion Cell Responses with a Probabilistic Spiking Model. *Journal of Neuroscience* 25:11003–11013.
- 1073 Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP (2008) Spatio-temporal

- 1074 correlations and visual signalling in a complete neuronal population. *Nature* 454:995–999.
- 1075 Rathbun DL, Alitto HJ, Warland DK, Usrey WM (2016) Stimulus Contrast and Retinogeniculate Signal
1076 Processing. *Frontiers in Neural Circuits* 10:8.
- 1077 Rathbun DL, Warland DK, Usrey WM (2010) Spike timing and information transmission at retinogeniculate
1078 synapses. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 30:13558–
1079 13566.
- 1080 Reid RC, Victor JD, Shapley RM (1997) The use of m-sequences in the analysis of visual neurons: Linear
1081 receptive field properties. *Visual Neuroscience* 14:1015–1027.
- 1082 Shapley R, Enroth-Cugell C (1984) Visual adaptation and retinal gain controls. *Progress in Retinal Research*
1083 3:263–346.
- 1084 Sincich LC, Adams DL, Economides JR, Horton JC (2007) Transmission of Spike Trains at the Retinogenic-
1085 ulate Synapse. *Journal of Neuroscience* 27:2683–2692.
- 1086 Sincich LC, Horton JC, Sharpee TO (2009) Preserving information in neural transmission. *The Journal of*
1087 *Neuroscience: The Official Journal of the Society for Neuroscience* 29:6207–6216.
- 1088 Sutter EE (1987) A practical non-stochastic approach to nonlinear time-domain analysis In: *Advanced*
1089 *Methods of Physiological Systems Modeling (Marmarelis VZ ed)*, Los Angeles, California: Biomedical
1090 *Simulations Resource*.
- 1091 Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN (2005) A Point Process Framework for Relating
1092 Neural Spiking Activity to Spiking History, Neural Ensemble, and Extrinsic Covariate Effects. *Journal*
1093 *of Neurophysiology* 93:1074–1089.
- 1094 Uglesich R, Casti A, Hayot F, Kaplan E (2009) Stimulus size dependence of information transfer from retina
1095 to thalamus. *Frontiers in Systems Neuroscience* 3.
- 1096 Usrey WM, Reppas JB, Reid RC (1999) Specificity and Strength of Retinogeniculate Connections. *Journal*
1097 *of Neurophysiology* 82:3527–3540.
- 1098 Usrey WM, Reppas JB, Reid RC (1998) Paired-spike interactions and synaptic efficacy of retinal inputs to
1099 the thalamus. *Nature* 395:384–387.
- 1100 Wang X, Hirsch JA, Sommer FT (2010) Recoding of Sensory Information across the Retinothalamic Synapse.
1101 *Journal of Neuroscience* 30:13567–13577.
- 1102 Weyand TG (2007) Retinogeniculate Transmission in Wakefulness. *Journal of Neurophysiology* 98:769–785.
- 1103 Weyand TG, Gafka AC (1998) Activity of neurons in area 6 of the cat during fixation and eye movements.
1104 *Visual Neuroscience* 15:123–140.
- 1105 Williamson RS, Sahani M, Pillow JW (2015) The Equivalence of Information-Theoretic and Likelihood-Based
1106 Methods for Neural Dimensionality Reduction. *PLOS Computational Biology* 11:e1004141.

1107 **Figure legends**

1108

1109 **Figure 1**

1110

1111 Data from an example pair (pair 214). **A and B:** Receptive field (RF) maps derived from reverse correlation
1112 between recorded spike trains and binary white noise stimulus. Red (blue) denotes regions of the RF that
1113 were excited by brighter (darker) pixels. White (black) circle in **A (B)** is the 1 SD contour of a circular
1114 Gaussian fit to RF of the LGN cell (RGC) overlaid on the RGC (LGN) RF to illustrate the high degree
1115 of spatial overlap. **C and D:** Cross-correlation between RGC and LGN spike trains for binary white noise
1116 (**C**) and drifting sinewave grating (**D**) stimuli. The inset text indicates the number of spikes recorded from
1117 each of the two neurons (**C**: 14,675 retinal spikes, 5,706 LGN spikes; **D**: 29,305 retinal spikes, 18,236 LGN
1118 spikes). The red line in **D** shows the correlation due to the stimulus that is attained if the spike train of the
1119 RGC is shifted in time by one stimulus cycle.

1120

1121 **Figure 2**

1122

1123 Comparison of ISI efficacy (left column) and retinal history (RH) models (right column). **A:** Relationship
1124 between retinal ISI and retinal efficacy for binary white noise (blue) and drifting grating data (red) for pair
1125 208. **B:** Retinal filters learned by the RH model fit to binary white noise (blue) and drifting grating (red) data
1126 from pair 208. Shading indicates ± 1 standard error of the optimization (see Methods). The time base for
1127 GLM filters is always relative to the retinal spike about which a prediction (relayed or non-relayed) is being
1128 made (i.e., the “target spike”). **C:** Normalized ISI-efficacy relation averaged across the population. Efficacies
1129 for each pair were normalized to the mean efficacy across all ISIs for that pair before averaging. Shading
1130 represents the 95% confidence interval (CI) across pairs from 5000 bootstrap resamples (see *Statistics*). **D:**
1131 Same as **B** but showing the average filters across pairs. Filters fit to the data from each pair were scaled to
1132 have a unit norm prior to averaging. Shading represents the 95% CI across pairs. **E:** Normalized ISI-efficacy
1133 relations for all eight pairs from the awake data set (thin grey lines) and the population average (thick gold
1134 line). Normalization was performed as in **C**. **F:** Retinal filters learned by RH models fit to data from each
1135 pair in the awake data set (thin grey lines) and the population average (thick gold line). Filters were scaled
1136 to have unit norm (as in **D**) to aid visualization.

1137 **Figure 3**

1138
 1139 Summary of filters learned by the two-component, combined history (CH) model. The left column shows
 1140 the retinal filters, and the right column shows the LGN filters for example pairs and the population for each
 1141 data set. **A:** Retinal filters learned by the CH model for binary white noise (blue) and drifting grating (red)
 1142 data from pair 208. **B:** Same as **A** but showing the LGN filters learned by the CH model. The time base for
 1143 retinal and LGN filters is the same (0 is the time of the “target” retinal spike), but LGN filters operate on
 1144 the prior activity of the LGN cell. **C:** Same as **A** but for the population. Filters fit to the data from each
 1145 pair were scaled to have a unit norm prior to averaging. Shading represents 95% CI across pairs. **D:** Same
 1146 as **C** but for LGN filters. **E:** Same as **C** but showing retinal filters learned from the awake data set (thin
 1147 grey lines show filters from each pair, the thick gold line shows the mean across pairs). **F:** same as **E** but
 1148 showing LGN filters.

1149
1150 **Figure 4**

1151
 1152 Retinal (**A & C**) and LGN (**B & D**) filters from the CH model fit to data where non-cardinal burst spikes
 1153 were first removed. The first row (**A & B**) use the classic burst spike definition of Lu et al. (1992): a
 1154 quiescent period ≥ 100 ms followed by two or spikes with ISIs ≤ 4 ms. The second row (**C & D**) use a more
 1155 relaxed criteria: a quiescent period ≥ 50 ms followed by two or more spikes with ISIs ≤ 6 ms.

1156
1157 **Figure 5**

1158
 1159 Qualitative comparison of model performance. **A left:** The predicted efficacies from each model were used
 1160 to group retinal spikes into bins, and the observed efficacy for each group (median across pairs) is plotted
 1161 against the corresponding bin label (error bars represent the MAD across pairs). Both predicted and observed
 1162 efficacies from each pair were normalized by the mean efficacy of that pair prior to calculating the median
 1163 and MAD. **A right:** The performance ($J_{\text{Bernoulli}}$) of the GLMs relative to the ISI-efficacy model is shown as
 1164 a function of ISI. Lines show the median performance difference across pairs; shading represents the MAD.
 1165 **B & C:** Same as **A** but for the binary white noise (**B**) and awake (**C**) data sets.

1166
1167 **Figure 6**

1168
 1169 Performance comparison of all models for binary white noise data. **A upper:** Comparison of ISI efficacy
 1170 and RH models. Each dot indicates the mean $J_{\text{Bernoulli}}$ for a given pair and model; lines connect data

1171 belonging to the same pair across models (thus the slope of the lines depicts the change in $J_{\text{Bernoulli}}$). The
1172 height of the vertical, colored bars indicates the median absolute deviation (MAD) of $J_{\text{Bernoulli}}$ across pairs
1173 for a given model, with the filled circle indicating the median value. **A lower:** Estimated paired median
1174 difference $J_{\text{Bernoulli}}$ between ISI efficacy and RH models. The black dot indicates the observed paired median
1175 difference and the vertical black line indicates the 95% confidence interval (CI) of the bootstrap distribution
1176 (5000 samples) shown in blue. **B:** Same as **A** but comparing ISI efficacy and CH model performance. **C:**
1177 Same as **A** but comparing RH and CH model performance.

1178

1179 **Figure 7**

1180

1181 Performance comparison of all models for drifting grating data. All conventions exactly follow those from
1182 Figure 6. Correlates of model performance are shown in Extended Data Figure 7-1. Model performance for
1183 the awake data set is shown in Extended Data Figure 7-2.

1184

1185 **Figure 8**

1186

1187 Comparison of RH models fit separately to subsets (quartiles) of the data grouped by LGN activity level. **A:**
1188 Average retinal filters from RH models fit to each quartile of the binary white noise data set from low (Q1,
1189 green) to high (Q4, purple) based on the activity level of the LGN neuron within a 100 ms period directly
1190 preceding the target retinal spike at $t = 0$. Shading represents 95% CI across $N=38$ pairs. **B Upper:**
1191 Comparison of model performance ($J_{\text{Bernoulli}}$) across all activity subsets. Each dot represents the model
1192 performance for a single pair (the spread along the x-axis is to aid visualization). **B Lower:** Bootstrap
1193 estimation of median model performance for each subset. Black dots indicate the median across pairs and
1194 black vertical lines indicate the 95% CI of the bootstrap distribution (shown in color, 5000 samples). **C, D**
1195 Same as **A, B** but for the drifting gratings data set ($N=33$). Results from a control analysis wherein relay
1196 status was simulated via GLMs is shown in Extended Data Figure 8-1 (see main text for details). Results
1197 of changing the spike quartile classification window are shown in Extended Data Figure 8-2.

1198

1199 **Figure 9**

1200

1201 Quantification of differences between filters learned from highest (Q4) and lowest (Q1) activity data sets.
1202 **A:** Population distributions (filled bars) and kernel density estimates (thick lines) of absolute differences
1203 between Q4 and Q1 filters for binary white noise (blue) and drifting grating (red) data. Filled triangles
1204 denote the median of each distribution. The gold triangle indicates the median difference for the awake data

1205 set for reference (where “high” and “low” were defined by a median split due to fewer spikes in that data
 1206 set). **B**: estimation of population medians from **A**. Filled black dots indicate the median and black vertical
 1207 lines indicate the 95% CI of the bootstrap distributions of population medians shown in blue (red) for binary
 1208 white noise (drifting grating) data.

1209

1210 **Extended data figure legends**

1211

1212 **Extended Data Figure 7-1**

1213

1214 Correlates of model performance. **A**, left: residual Spearman correlation between $J_{\text{Bernoulli}}$ from RH models
 1215 and retinal contribution where the effect of retinal efficacy on each variable has been removed prior to the
 1216 analysis. Right, estimation of correlation coefficient using 5000 bootstrap resamples. Black dots denote
 1217 point estimates, vertical black lines denote 95% CI, and filled distributions summarize the results of the
 1218 resampling. **B**, left: Spearman correlation between model performance improvement ($\Delta J_{\text{Bernoulli}}$) between
 1219 CH and RH models and retinal contribution. As retinal efficacy is not correlated with $\Delta J_{\text{Bernoulli}}$ regular
 1220 Spearman correlation was used. Right, estimation analysis for correlation show at left. **C**, left: Spearman
 1221 correlation between $\Delta J_{\text{Bernoulli}}$ and the percent of LGN spikes that were part of identified bursts (using the
 1222 traditional criteria of Lu et al. (1992)) excluding the cardinal spike of each burst. Right, estimation analysis
 1223 for correlation show at left.

1224

1225 **Extended Data Figure 7-2**

1226

1227 Model comparison and activity level analysis for awake data. **A**: Model performance (Mean $J_{\text{Bernoulli}}$ across
 1228 folds) for each model and pair (grey points) where grey lines connect points that correspond to the same
 1229 pair. Large, solid color circles indicate the median, and solid-color vertical lines show the MAD, across pairs
 1230 for a given model (Black: ISI model, green: retinal (RH) model, purple: combined (CH) model). **B**: Retinal
 1231 filters learned by RH models from low (green) and high (purple) activity data sets (similar to Figure 6 but
 1232 utilizing a median split to assign each retinal spike to a data set). Filters from individual pairs are show
 1233 in less saturated, thin lines while thick saturated lines indicate the mean across pairs (all filters are scaled
 1234 to have unit norm to aid visualization). Inset axis highlights the boxed region corresponding to the 30 ms
 1235 immediately preceding each “target spike” (at $t=0$). The red arrows indicate the filters learned from pair
 1236 200001250, which is the only pair of the awake data set that was stimulated with gratings during recording.

1237 **Extended Data Figure 8-1**

1238

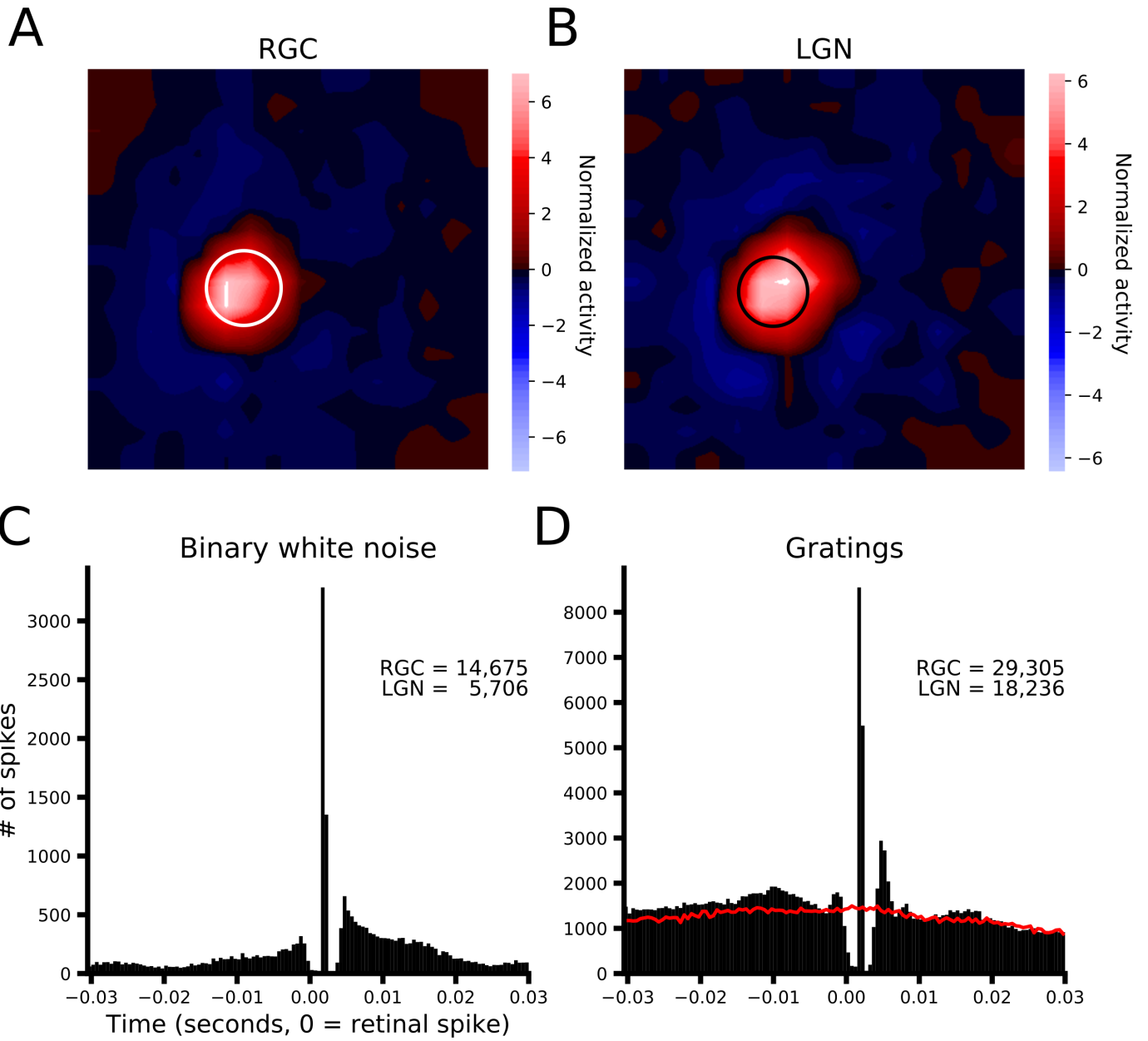
1239 Comparison of RH models fit separately to subsets (quartiles) of simulated data grouped by LGN activity
1240 level. The relay status of each retinal spike was determined by simulating a RH GLM with a fixed retinal
1241 filter (i.e. the filter did not change with activity level). **A:** Average retinal filters from RH models fit to each
1242 quartile of the binary white noise data set from low (Q1, green) to high (Q4, purple) based on the activity
1243 level of the LGN neuron within a 100 ms period directly preceding the target retinal spike at $t = 0$. Shading
1244 represents 95% CI across $N=38$ pairs. **B Upper:** Comparison of model performance ($J_{\text{Bernoulli}}$) across all
1245 activity subsets. Each dot represents the model performance for a single pair (the spread along the x-axis
1246 is to aid visualization). **B Lower:** Bootstrap estimation of median model performance for each subset.
1247 Black dots indicate the median across pairs and black vertical lines indicate the 95% CI of the bootstrap
1248 distribution (shown in color, 5000 samples). **C, D** Same as **A, B** but for the drifting gratings data set
1249 ($N=33$).

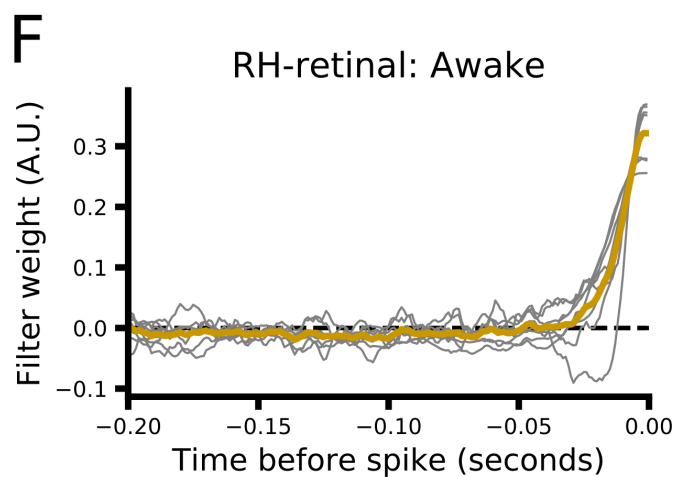
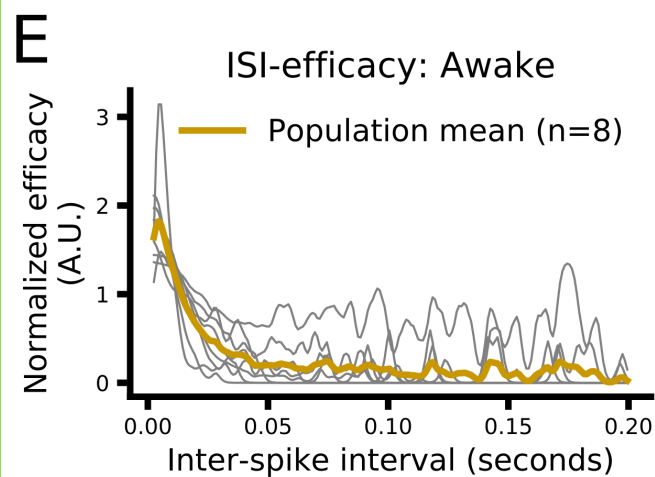
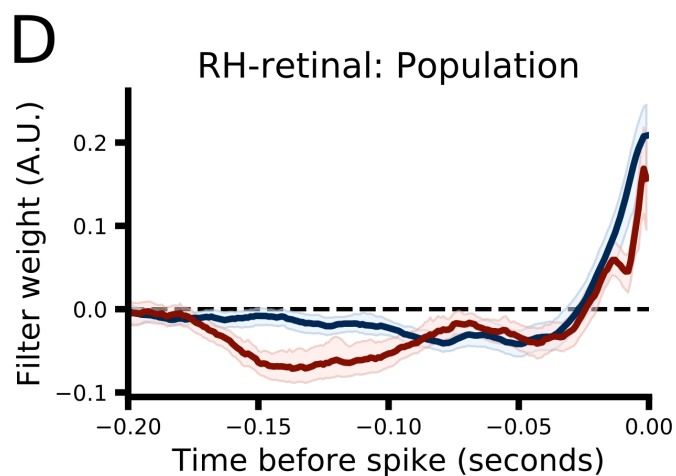
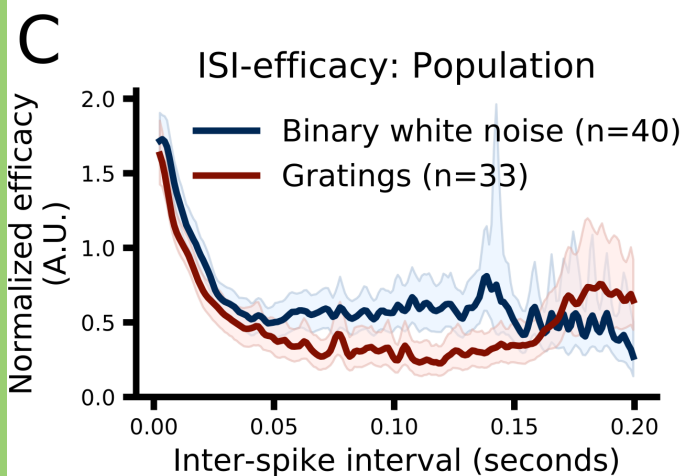
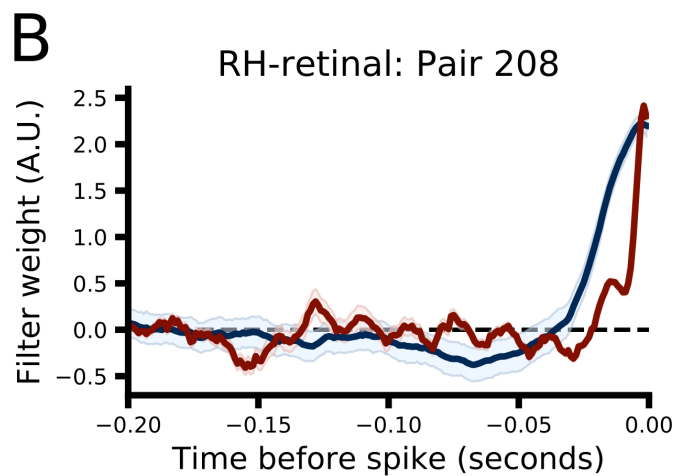
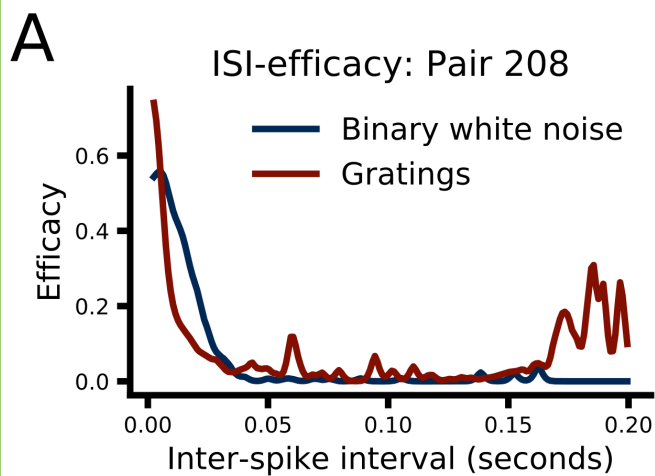
1250

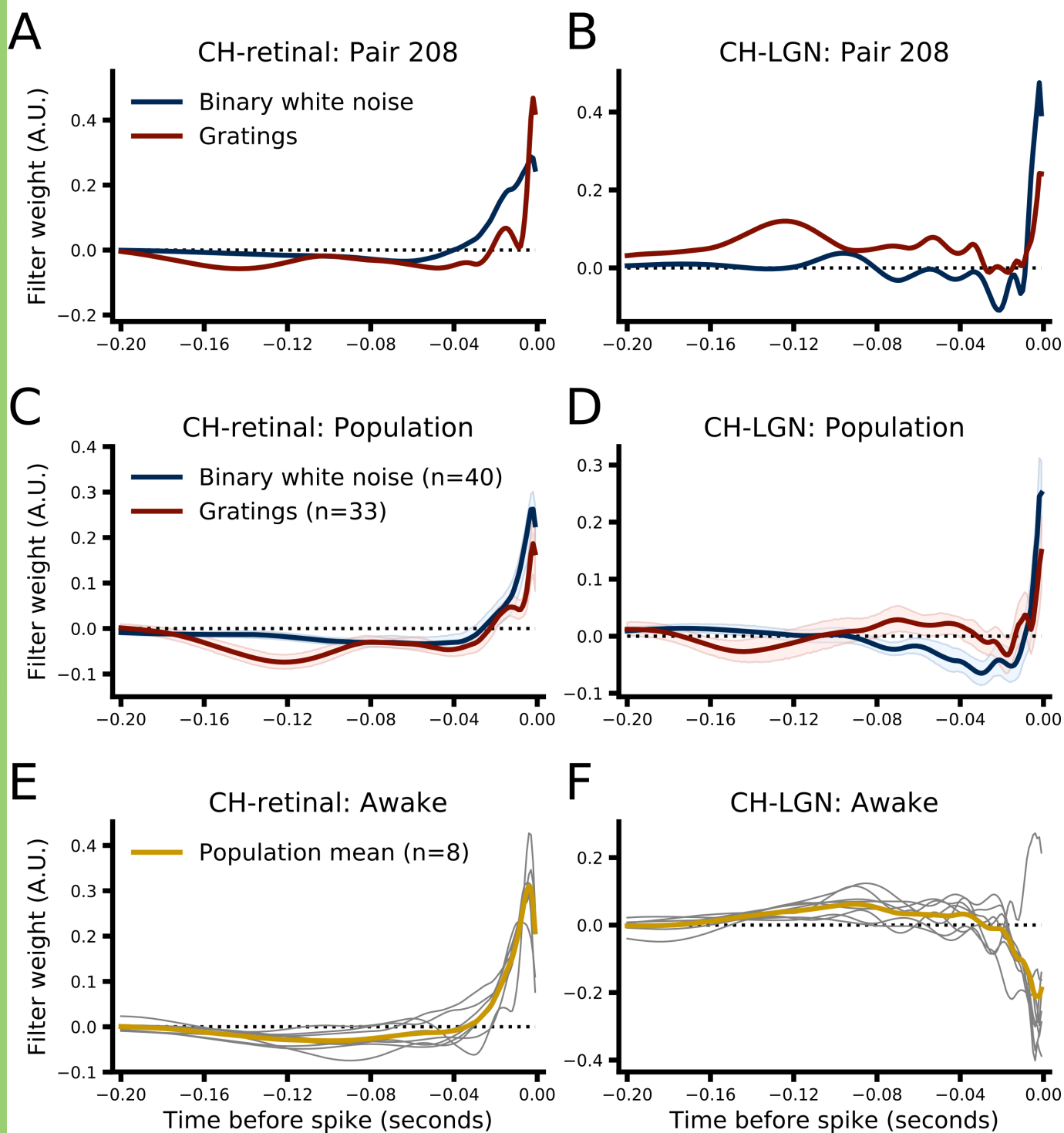
1251 **Extended Data Figure 8-2**

1252

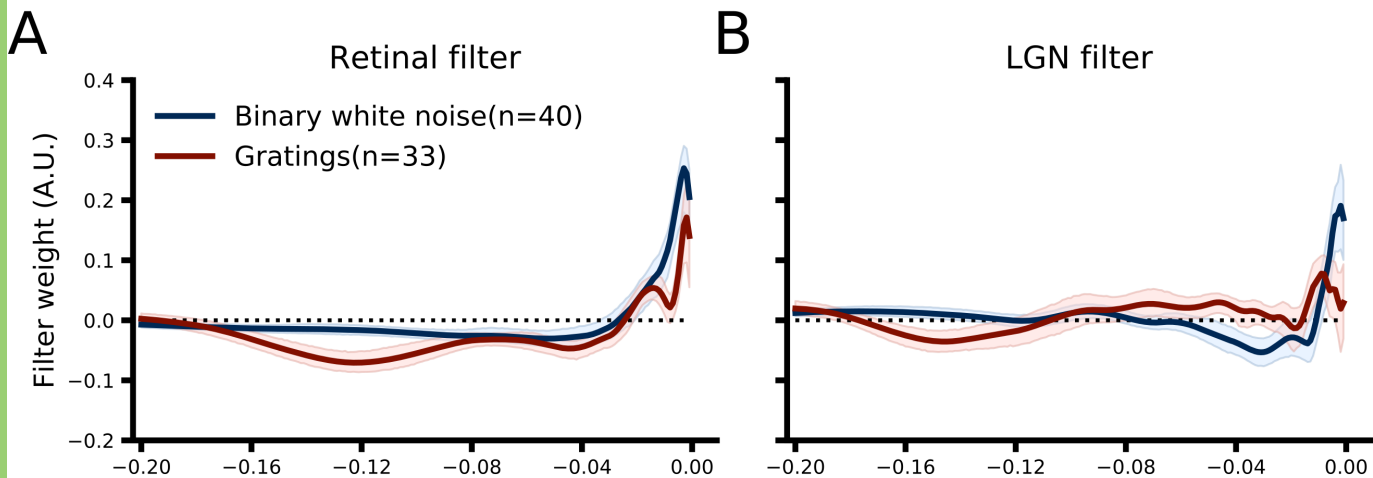
1253 Activity level analysis utilizing different time windows for partitioning retinal spikes. **A:** RH model filters
1254 learned from lowest (Q1) to highest (Q4) activity level subsets for binary white noise data where retinal
1255 spike assignment is based on a quartile partitioning of LGN spike count within a 250ms window preceding
1256 each retinal spike. **B:** Same as **A** but for drifting grating data. **C & D:** Same as **A & B** but using a 125ms
1257 window for partitioning.



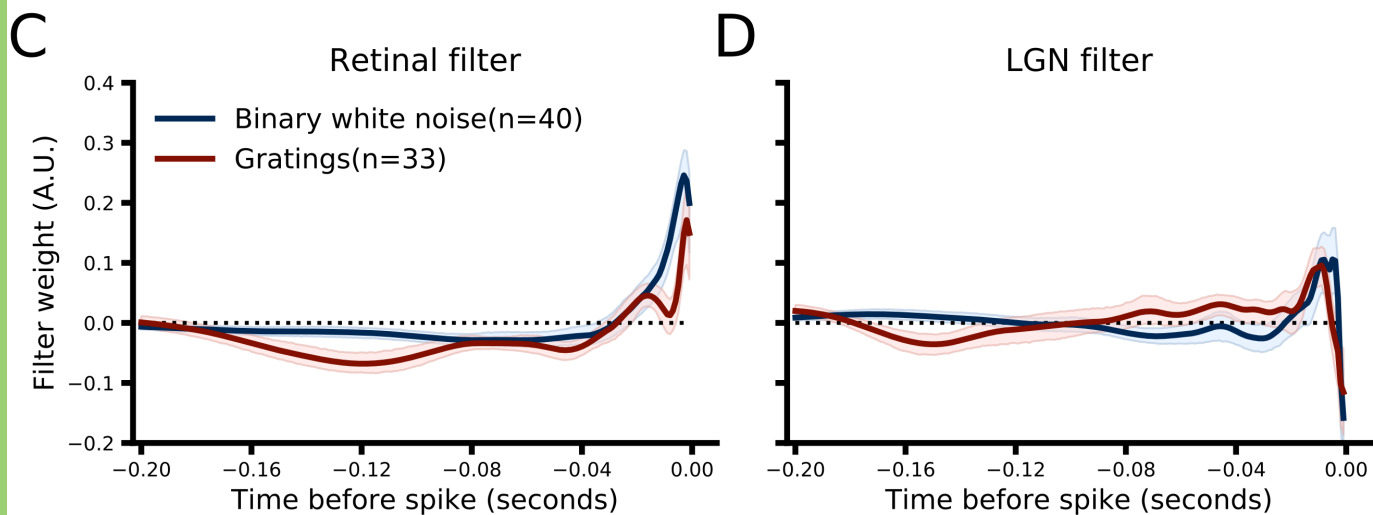


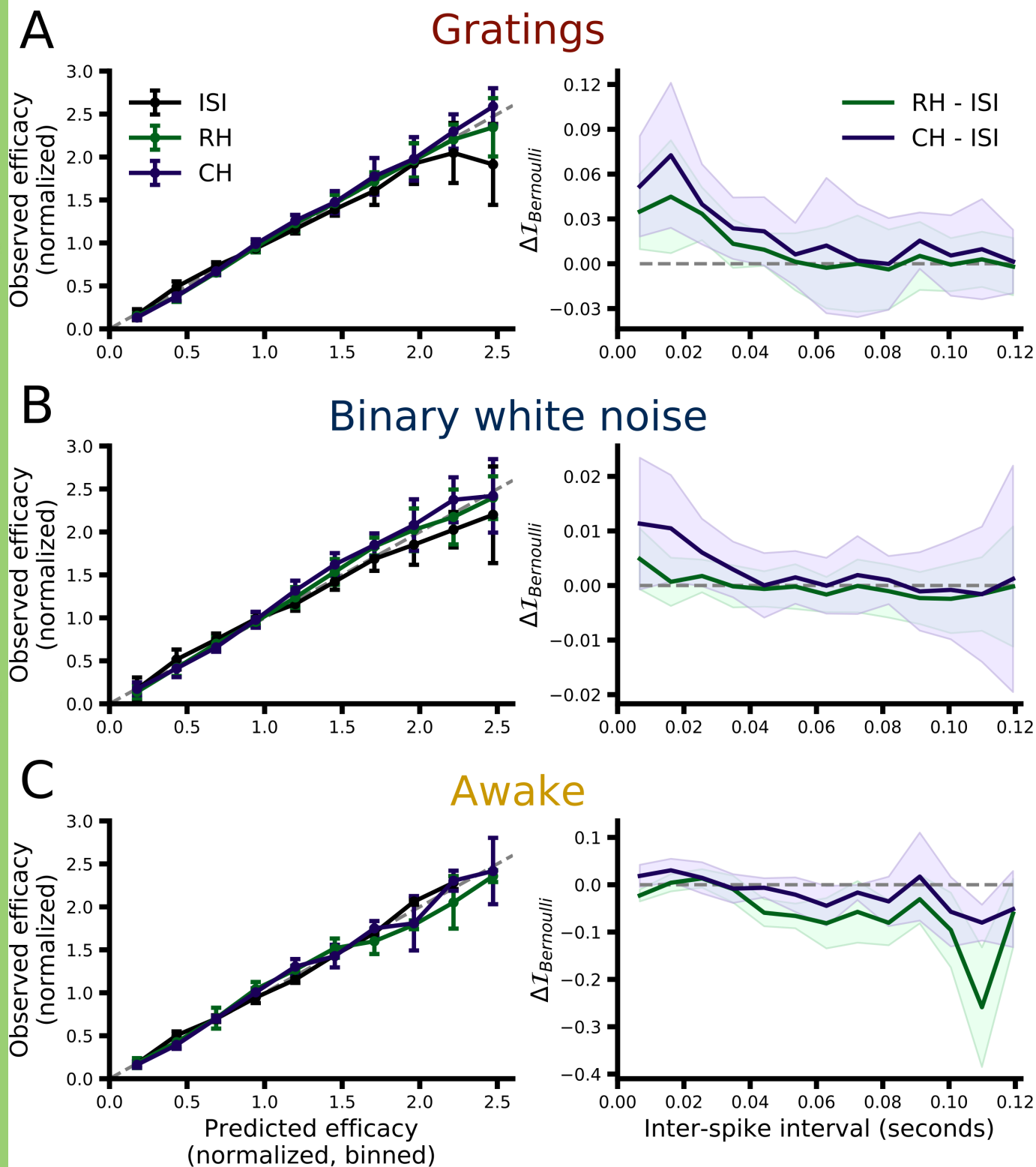


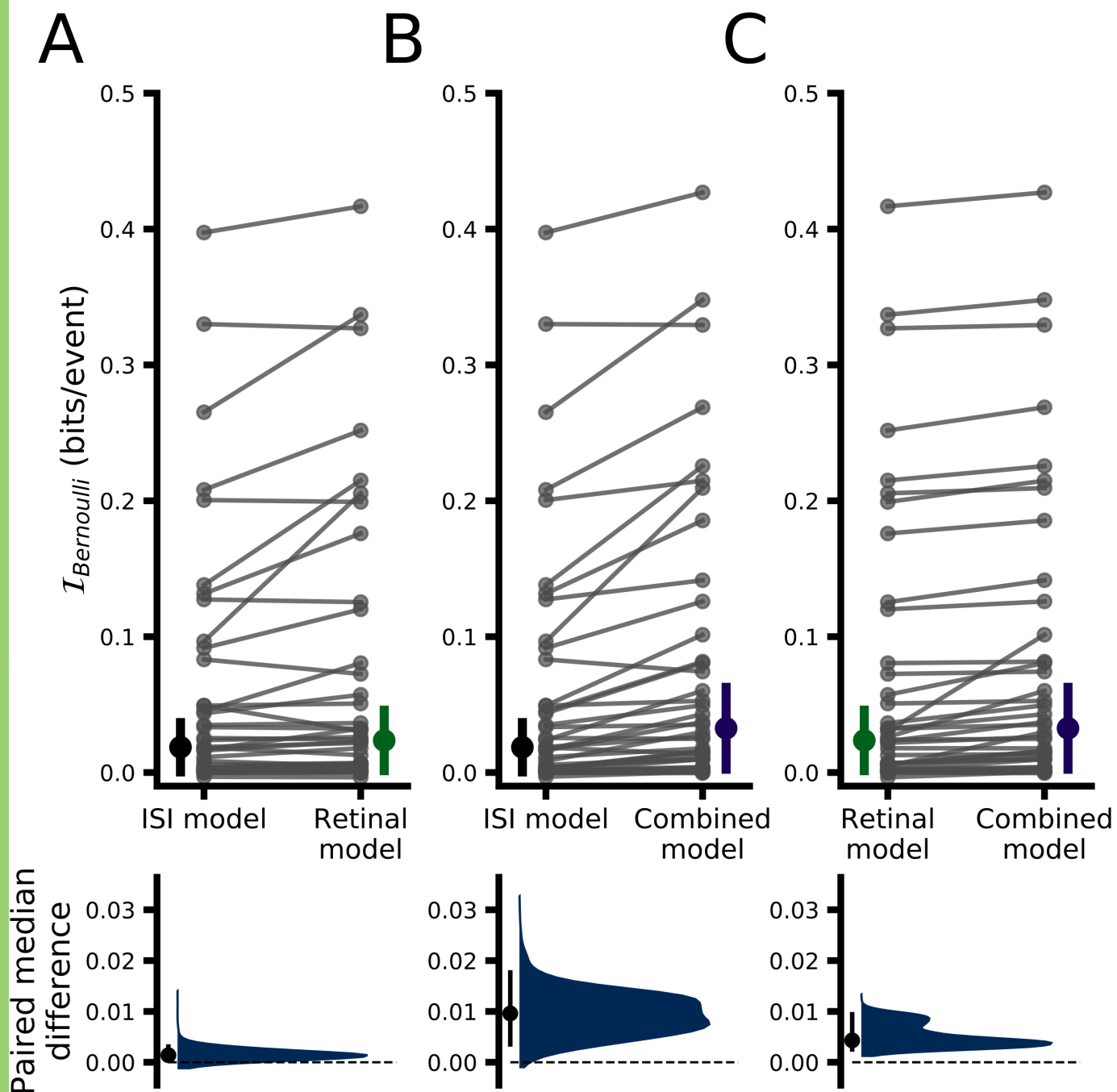
Deadtime ≥ 100 ms, ISI ≤ 4 ms

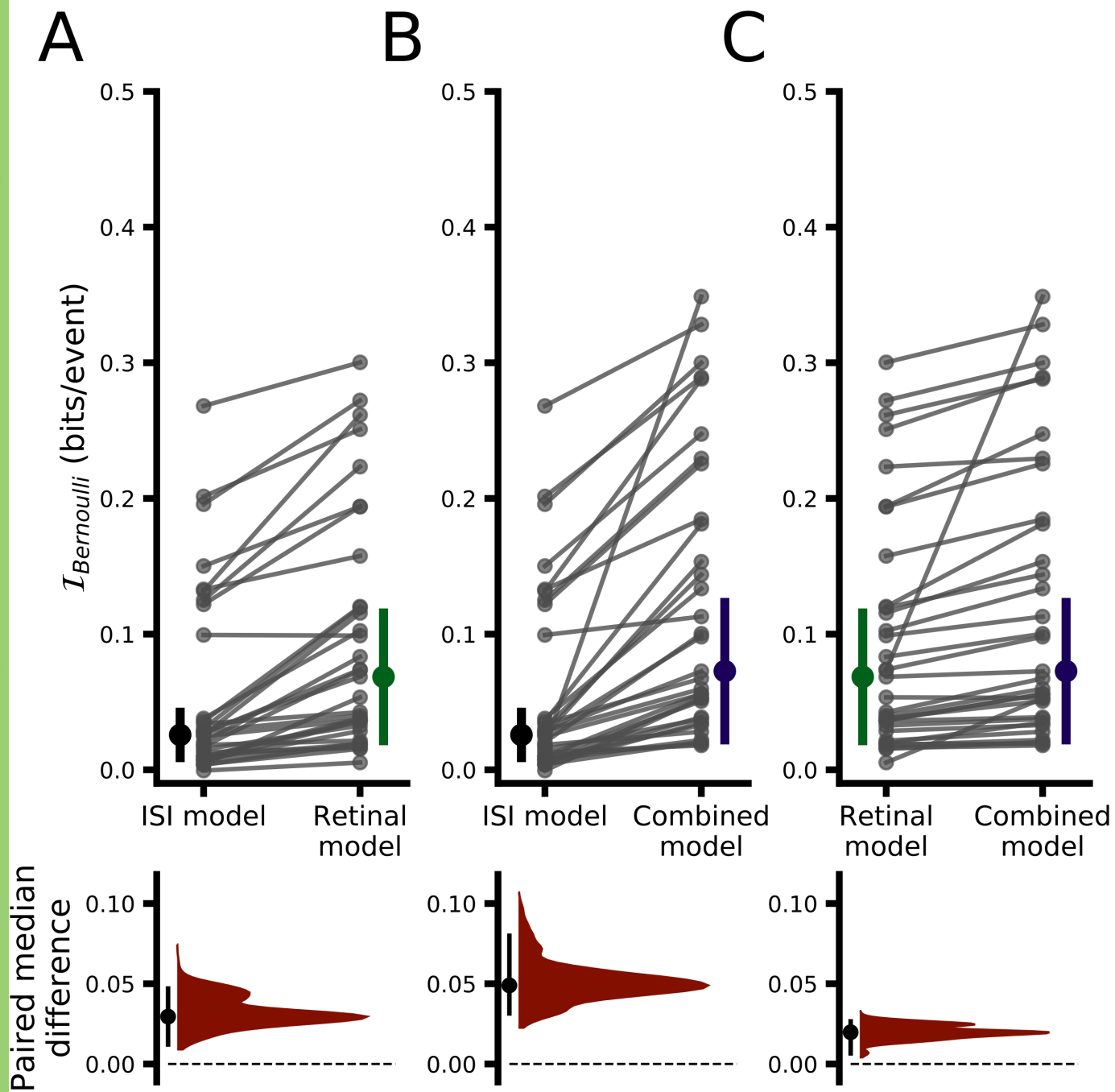


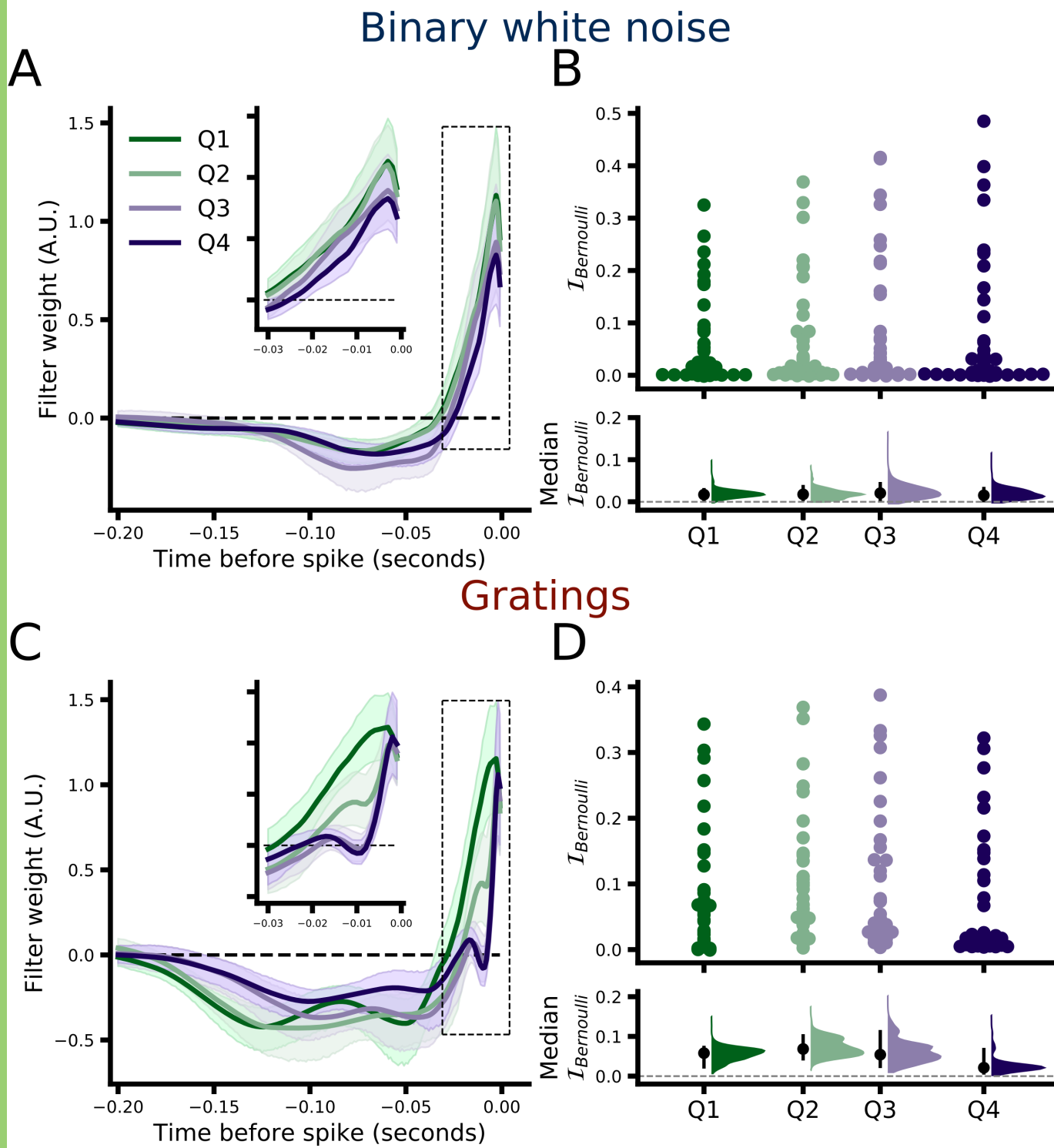
Deadtime ≥ 50 ms, ISI ≤ 6 ms



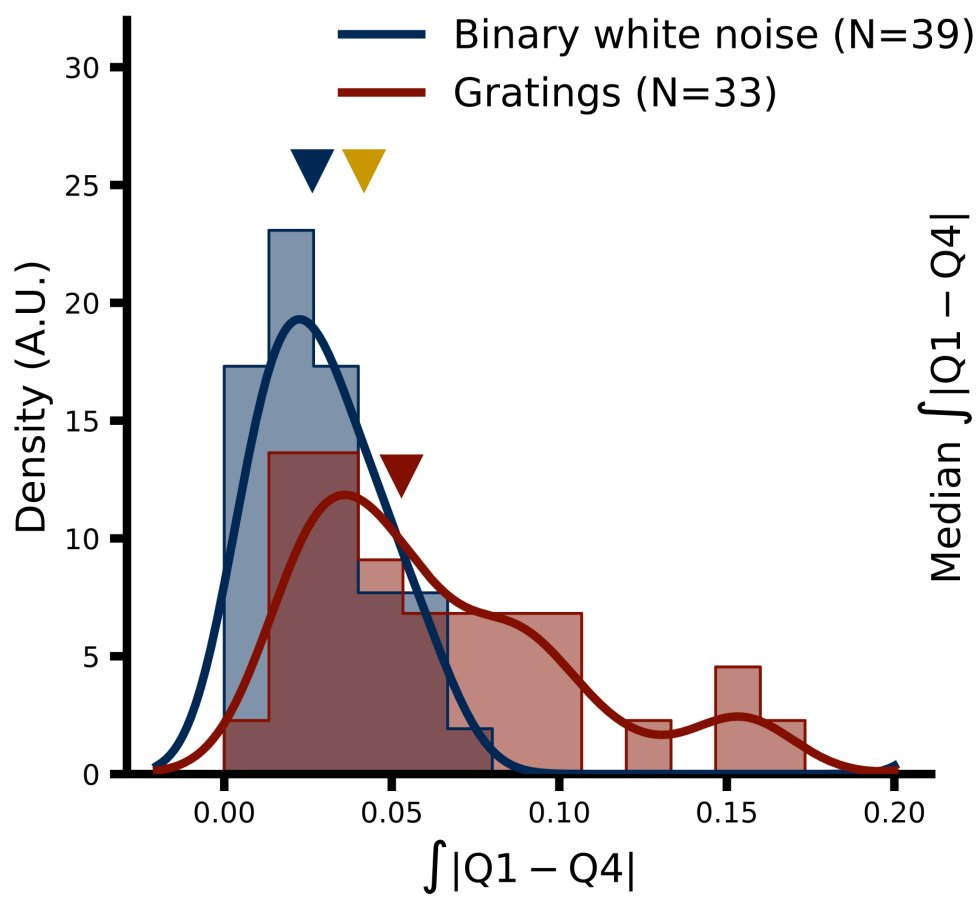




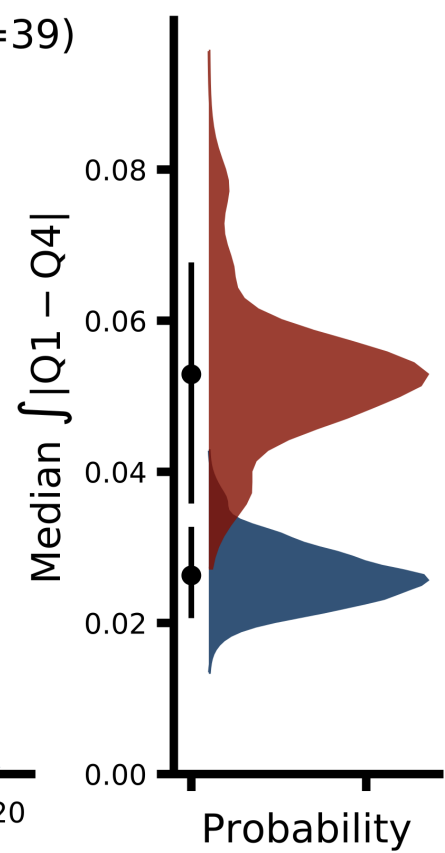




A



B



	Data set	Metric	Conditions	Paired median difference	MAD	95% CI	<i>p</i> -value
Figure 6							
a	Binary Noise (N=40)	$\mathcal{I}_{\text{Bernoulli}}$	RH - ISI	0.002 bits/spike	0.003	[0.000, 0.003]	0.0092
b	Binary Noise (N=40)	$\mathcal{I}_{\text{Bernoulli}}$	CH - RH	0.009 bits/spike	0.008	[0.004, 0.015]	0.0002
c	Binary Noise (N=40)	$\mathcal{I}_{\text{Bernoulli}}$	CH - ISI	0.004 bits/spike	0.004	[0.003, 0.009]	0.0002
Figure 7							
d	Gratings (N=33)	$\mathcal{I}_{\text{Bernoulli}}$	RH - ISI	0.030 bits/spike	0.020	[0.012, 0.047]	0.0002
e	Gratings (N=33)	$\mathcal{I}_{\text{Bernoulli}}$	CH - ISI	0.049 bits/spike	0.033	[0.032, 0.080]	0.0002
f	Gratings (N=33)	$\mathcal{I}_{\text{Bernoulli}}$	CH - RH	0.020 bits/spike	0.014	[0.006, 0.027]	0.0002
Figure 8							
g	Gratings (N=33)	$\mathcal{I}_{\text{Bernoulli}}$	Q4 - Q1	-0.005 bits/spike	0.041	[-0.047, 0.003]	0.353
h	Binary Noise (N=39)	$\mathcal{I}_{\text{Bernoulli}}$	Q4 - Q1	0.001 bits/spike	0.007	[-0.001, 0.004]	0.396
Figure 9							
i	Anesthetized (N=27)	Absolute difference	Gratings - Noise 100ms	0.031	0.016	[0.018, 0.039]	0.0002
j	Anesthetized (N=27)	Absolute difference	Gratings - Noise 30ms	0.008	0.007	[0.002, 0.012]	0.0004

Table 1: Statistical table of results. Confidence intervals are derived from 5,000 bootstrap resamples and are bias corrected and accelerated. *p*-values are derived from paired-permutation tests with 5,000 permutations. See Methods for details.