State dependence of neural networks on past history and stimulus presentation in *Caenorhabditis elegans*

Zachary T. Cecere[1,2,3], Eviatar Yemini[4], Tatyana O. Sharpee[1,3], and Sreekanth H. Chalasani[1,2*]

[1]Neurosciences Graduate Program, University of California, San Diego, La Jolla, CA, USA.

[2]Molecular Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, USA.

[3]Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, USA.

[4]Department of Biological Sciences, Howard Hughes Medical Institute, Columbia University, New York, NY, USA.

*Address for correspondence: schalasani@salk.edu (S.H.C.)

Abstract

Neurons cooperate and interact with one another to generate behavior. While individual neurons and their ensembles are often probed in isolation, very little is known about their relationships on a global scale. Here, we define six groups of neurons with interactions that encode and process food stimulus responses in *Caenorhabditis elegans*. Using brain-wide calcium imaging in restrained animals, we categorized active neurons into either: sensory (On or Off) or motor (forward, backward, dorsal or ventral turn) neuron groups. Next, we show that stimulus onset rapidly shifts activity from backward to forward motor neuron groups, while prolonged stimulus removal is required to initiate converse activity. Finally, we use a gated multinomial logistic regression model to understand how stimulus processing depends on network state. This study shows how neural activity is coordinated across a network to generate global dynamics that drive behavior.

Introduction

Nervous systems integrate sensory information with changes in internal states to generate behaviors on multiple timescales (Dugatkin, 2020). Sensory neurons are specialized to detect and distinguish both spatial and temporal features of various stimuli (Dubs et al., 1981; Hallock and Di Lorenzo, 2006; Reidenbach and Koehl, 2011; Riffell et al., 2008; Yu and Goodrich, 2014) and play a crucial role in driving animal behavior. Identifying how sensory input modifies neural activity is a necessary step in understanding how the brain encodes and generates behaviors. Additionally, efforts to monitor neural activity patterns from populations of sensory neurons and downstream circuitry in the mammalian brain are limited by the need to average across trials, or to subsample from local brain regions (Blackwell and Geffen, 2017; Churchland et al., 2012; Williamson et al., 2019). This problem is alleviated by studying simpler animals like zebrafish, fly larvae and adult *C. elegans* where neural activity at single-cell resolution has been monitored across the entire brain (Ahrens et al., 2012; Ahrens et al., 2013; Lemon et al., 2015; Schrodel et al., 2013). However, the relationships between sensory input and whole-brain activity have not been defined.

With its well defined sensory neurons and the corresponding stimuli along with a fully mapped neuroanatomy and powerful genetic tools, the nematode *Caenorhabditis elegans* is ideally suited for this analysis (Antoshechkin and Sternberg, 2007; Bargmann, 2006; White et al., 1986). Functional imaging studies using genetically-encoded calcium indicators have shown that various *C. elegans* chemosensory neurons respond to stimuli with calcium transients that last many seconds (Chalasani et al., 2007; Hilliard et al., 2005; Ibsen et al., 2015; Suzuki et al., 2008; Zimmer et al., 2009). Moreover, previous studies have shown that *C. elegans* whole-brain

activity can be monitored both in trapped and freely moving animals (Nguyen et al., 2016; Schrodel et al., 2013; Venkatachalam et al., 2016). Additionally, studies have also shown that *C. elegans* whole brain activity mostly represents various motor states (Kato et al., 2015), is likely organized into hierarchical patterns (Kaplan et al., 2020) and is modified by sleep (Nichols et al., 2017), starvation (Skora et al., 2018) and attractive odor and repulsive salt responses (How et al., 2020; Yemini et al., 2021). While these studies demonstrate the utility of whole-brain imaging and provide insights into our understanding of the nematode nervous system, they have to-date been unable to go beyond a simple descriptive model of sensory-neuron activity responses to stimuli. Thus, a model of how sensory inputs and internal state combine to drive neural activity is lacking. A model that accounts for these real-world phenomena would provide a significant advance in explaining global brain dynamics that are shaped by an ever-changing world of external inputs and internal states.

Here, we use bacterial stimuli (food that these animals consume) to monitor activity across multiple neurons in the head of a trapped *C. elegans* in using two strains expressing a pan-neuronally nuclear-localized calcium indicator. One strain has multiple fluorophores that aid in neuron identification (NeuroPAL, (Yemini et al., 2021)) while the other does not (non-NeuroPAL, (Schrodel et al., 2013), thus permitting us to develop a generalized model that is strain agnostic. We show that active neurons can be divided into six groups and then use computational models to derive relationships between them. Specifically, we show that prior states and stimulus input are crucial to forward and reverse motor neuron activity, thus forging a model between external inputs and internal states.

Results

Active neurons fall into one of six groups

We designed an automated system for simultaneous calcium imaging and delivery of a rapidly fluctuating bacterial stimulus within a microfluidic device (Figure 1A). To monitor whole-brain activity, we trapped transgenic animals expressing a genetically-encoded calcium indicator in all of their neural nuclei. While recording neural activity, buffer is exchanged with bacterial stimulus around the animal's nose to mimic various stimulus sequence patterns. Previously, this device has been used to show that individual chemosensory neurons detect and respond to a bacterial stimulus (Calhoun et al., 2015; Chalasani et al., 2007). To identify general principles of information flow, we imaging whole-brain neural activity in two independent transgenic strains. One of these strains expresses the calcium indicator GCaMP6s (Chen et al., 2013) in all neural nuclei as well as a combination of other fluorophores, which allow for unambiguous identification of neural identity along with the cell's responses. Neurons are identified by a combination of fluorophores such that neighboring neurons have different colors (Figure 1B, NeuroPAL (Yemini et al., 2021)). The second transgenic expresses the calcium sensor, GCaMP5k in all neural nuclei (Akerboom et al., 2013), but does not include any information about neural identity. This strain has been previously used to obtain whole brain activity from trapped *C. elegans* (Kato et al., 2015; Schrodel et al., 2013). While GCaMP5K and GCaMP6s have different calcium affinities (189 ± 5 nM vs 144 ± 4nM) and hill coefficients (3.8 ± 0.3 vs 2.90 ± 0.17) *in vitro* (Akerboom et al., 2013; Chen et al., 2013), we hypothesized that using these two strains will allow us to uncover broad principles of information processing in the *C. elegans* nervous system.

We recorded changes in neural nuclear-GCaMP fluorescence using a Zeiss AiryScan LSM 880 at 1.5 volumes/second, and extracted neural activity data using a non-native matrix

factorization (see Strains and Calcium imaging in Methods for further details). We found that in response to bacterial stimuli, all active neurons could be divided into six groups: two sensory neurons and four motor neurons groups (Figure 1C). Sensory neurons consistently responded either to the addition (On) or removal (Off) of stimulus. Moreover, while the “On” cells are positively correlated with stimulus, “Off” cells are negatively correlated with stimulus (Figure 1D). In contrast, neural activity in the four motor neuron groups was not consistent with either stimulus addition or removal. However, these four groups could be further divided into two subgroups of anti-correlated pairs. One of these anti-correlated pair included neurons that were on the dorsal and ventral side and identified to be SMDD and SMDV. Consistently, SMDD neurons had similar calcium transients that were anti-correlated with SMDV neurons (Figure 1E). Their position along with their stereotyped activity patterns also allowed for these neurons to be identified in the non-NeuroPAL GCaMP5K transgenic animal. The second anti-correlated pair included either AVA or RME neurons along with others. AVA and RME were found to be anti-correlated in every animal examine. Most animals had a large number of cells that strongly correlated with these two neurons. Neurons whose transients matched AVA or RME activity at 85% or greater were added to the “AVE cell cluster” and “RME cell cluster” respectively (Figure 1F). Neural activity of the “RME” group was anti-correlated with those neurons in the “AVA” group, consistent with previous observations (Kato et al., 2015). We were unable to identify any other active neuron classes in either NeuroPAL or non-NeuroPAL transgenic animals whose activity patterns did not fall into one of the six above mentioned groups. Furthermore, we analyzed multiple animals in both GCaMP transgenics and found that they had similar numbers of neurons in each of these six groups (Figure 1G). Collectively, these data show that in the

context of bacterial stimuli and trapped animals, active neurons could be divided into one of six groups.

On and Off sensory neurons have different adaptation kinetics in response to bacterial stimulus

We identified sensory neurons that responded to addition ("On") and removal ("Off") of bacterial stimulus (Figure 2A,B). Specifically, we calculated every neuron's stimulus-induced fluorescence change (averaged across all stimulus pulses) and sorted across all neurons. The top five neurons that scored highest in this metric and also did not belong to one of the motor neuron groups were identified as On sensory neurons, while the lowest scoring five neurons were classified as Off sensory neurons.

Next, we sought a model that can explain different aspects of these responses including their initial rise and the subsequent return to baseline. A previous study used a cascade of first-order linear ordinary differential equations to model activity in sensory neurons in response to food-associated odors. This model included a process that reproduced the rise time of the neural response, followed by two parallel paths of opposite sign with either fast or slow dynamics and was able to accurately predict the sensory neural responses. The contribution of the slow component of the opposite sign also allows this model to predict response decay after peak activity, a feature that resembles adaptation of that sensory neurons (Kato et al., 2014). The core of this cascade model is built on the following differential equations:

$$dx_i/dt = tau * ([input] - x_i(t))$$

$$X(t) = \sum_i a_i . x_i(t)$$

tau is the filter time constant. It controls how fast the filter responds to stimulus [input]. $X(t)$ is the inferred calcium level of the cell. It is the sum of the solutions to the ordinary differential

equations. The original cascade model assumed perfect adaptation with two model filters by constraining $a1 = a0$ (Kato et al., 2014). When encountering a stimulus step change, a perfectly adapting filter will exhibit a brief change in activity before returning to its baseline. We utilize the following model, based on the cascade model, to test whether this perfect adaptation assumption is justified in *C. elegans* sensory cells. We also use the model to obtain estimates of the adaptation timescales of On and Off cells.

$$dx_i/dt = tau * ([input] - x_i(t))$$

$$X(t) = \sum_i a_i . x_i(t)$$

$$g(t) = (X(t)^p * f_g)(t)$$

$$\text{perfect adaptation: } \sum_i a_i = 0$$

$X(t)$ is the inferred calcium, f_g is the GCaMP filter (see GCaMP filters in Methods for details), and p is a positive value required for GCaMP transformation. The fitting procedure learns values for the time constant *tau* and the basis coefficients a_i. We use the solutions to these ordinary differential equations as basis functions. For a given model instantiation, each equation is assigned a value for *tau*. After the ordinary differential equations are solved analytically, the model can be linearized, allowing for robust estimates of the a_i coefficients (see Sensory neuron model in methods). A random search strategy is used to obtain good estimates of *tau*. This model formulation produced consistent and robust fits to worm sensory cells and allowed us to easily fit models with arbitrary numbers of cascade equations. This latter factor was thought necessary to model Off cell responses (Kato et al., 2014). Instead, we found the type of adaptation to be more important. Our model can learn coefficients (a_i) for an arbitrary number of ordinary differential equation basis functions and so can adapt on multiple timescales (see Sensory neuron model in methods). In addition, by toggling the coefficient constraint, we can test whether perfect adaptation is necessary to predict sensory neuron activity.

We used a 3-fold cross validation approach to assess the performance of various models. We found that adapting models performed significantly better than non-adapting ones for both On and Off sensory neurons. Moreover, we found that a perfect adaptation model with one adapting filter (two filters total) was best suited to explain adaptation in On, but not Off sensory neurons (Figure 2C). In contrast, an imperfect adaptation model with a single adaptation filter (two filters total) best described Off sensory neuron adaptation (Figure 2D). We also compared the two filters for both On and Off sensory neurons. We found that the slow filter of the Off sensory neuron rises faster than the corresponding, slow On sensory neuron filter, indicating that the Off neurons adapt faster than On neurons (Figure 2E,2F). In the perfect adaptation model, the coefficients for the positive and negative filters are guaranteed to have equal magnitude, which is not the case for the imperfect adaptation model. Here, we find that for Off sensory neurons, the positive secondary filter has a smaller magnitude than the negative primary filter, suggesting that these neurons do not completely adapt to a stimulus pulse on any timescale we tested. It should be noted that the pulses in our stimulus protocols last from a few seconds up to a minute. Collectively, these data suggest that adaptation to the bacterial stimulus in Off sensory neurons is distinct from that seen in On sensory neurons.

A Bacterial stimulus drives activity of neurons in the RME group

Given the central role that food plays in worm behavior and survival, we hypothesized that the bacterial stimulus would affect the neural activity of the various motor neuron groups. To test our hypothesis, we used a multinomial logistic regression decoding model to predict stimulus classes from neural activity in the RME and AVA motor neuron groups (see Decoding targets in

methods) (schematic in Figure 3A). Worm identity was also considered to capture cross-worm variability. We fit the model to NeuroPAL and non-NeuroPAL animals separately to test whether decoding features were preserved across strains. Multinomial logistic regression is perhaps the simplest possible classification model. Also, with class balancing, it has a very natural null model: prediction from worm identity only (see Decoding - class balancing in methods). Here, predictions are generated by linearly combining a set of weights with explanatory variables of a given observation:

$$P(S_i(t) \mid X(t)) = softmax(\sum_k \beta_{i,k} X_k(t))$$

Where $S_i(t)$ is the stimulus class i at time window t. $X(t)$ captures the activity of RME and AVA neurons at time window t along with worm identity information. $\beta_{i,k}$ is made up of the coefficients for stimulus class i, which are regularized using an L1 norm (Haynes, 2011; Kriegeskorte and Douglas, 2019; Meyers, 2013). In particular, we used AVA and RME neural activity in a time window of t-8 to t+8 seconds to predict four stimulus states: Stimulus ON, Stimulus OFF, Stimulus OFF-to-ON and ON-to-OFF transitions over the time windows of t-4 to t+4 seconds. Please see ‘Blockstrap’ and ‘Decoding targets’ in methods for more details and rationale. Inclusion of RME and AVA information improved cross-validation decoding model performance in both NeuroPAL and non-NeuroPAL animals (Figure 3B). We find that for both NeuroPAL and non-NeuroPAL transgenic animals, Stimulus OFF state is associated with decreased or low RME neural activity (Figure 3C, top). In contrast, Stimulus ON is associated with increasing RME and decreasing or low AVA neural activity (Figure 3C, second from top). These results suggest that the bacterial stimulus biases the network towards RME activation, while its removal drives AVA activity.

Next, we analyzed the linear filters of RME and AVA activity associated with Stimulus OFF-to-ON and ON-to-OFF transitions. The stimulus On-to-Off filters (Figure 3C, third from top) look like diminished versions of the stimulus On filters (Figure 3C, second from top). This suggests that stimulus presentation continues to drive RME activation for a period of time after its removal. Similarly, the Off-to-On filters (Figure 3C, bottom) are similar to the stimulus Off filters (Figure 3C, top). However, the AVA filters have a larger magnitude in Off-to-On transitions than in the stimulus Off condition. Combining this with the fact that these AVA filters look like second derivative filters suggests stimulus onset has an immediate effect on AVA. In both strains, it appears to immediately inhibit AVA, yielding a peak in AVA fluorescence and maximizing the value of the second derivative filter. These results suggest that the RME-to-AVA transition requires prolonged stimulus removal, while the AVA-to-RME transition begins at stimulus onset.

Network history and stimulus states can predict neural activity

We adapted our multinomial logistic regression (MLR) approach to predict RME and AVA cell cluster activity. We found that both RME and AVA cell clusters have a number of features in their activity that were difficult to describe with a linear model like MLR. For instance, members of the AVA cell cluster exhibit boundedness: their activity is limited to the range between their upper and lower stable states. Also, activity in these cells appears to have momentum: when one of these cells begins a transition between stable states, it will tend to complete that transition. A single, linear model is unable to describe both momentum and boundedness. It must learn the positive correlation between a future AVA rise and a past AVA increase to capture momentum. However, this positive correlation should weaken and become negative as AVA nears its upper

bound since AVA activity does not rise above a bounded limit. Thus, a gating model is required to capture the change in this positive correlation.

Here, we employ multiple MLR models to capture nonlinear dynamics, allowing us to capture features like momentum and boundedness. Initially, these models predict changes in RME and AVA activity. A gating model is used to divide the space of network histories into subspaces. Each of these different subspaces is associated with a different MLR model. This approach with two submodels can simultaneously model both momentum and boundedness in AVA. In one theoretical solution, one MLR submodel is only active when the AVA cell cluster is near its lower bound. Another is only active when AVA nears its upper bound. Thus, the first submodel learns a positive correlation between a past AVA increase in activity and a future AVA rise (momentum), while the second learns a weak positive or negative correlation (boundedness).

A gating model is used to divide the network history space into different subspaces. In our study, we used Soft Trees as the gating model (Frosst and Hinton, 2017). A Soft Tree is a form of oblique, decision tree that is differentiable from end-to-end (see Prediction: hybrid model in Methods for more details). Soft Trees are excellent for this task because they can model high-dimensional, nonlinear features in a completely interpretable fashion. We fed the network history of neural activity (AVA, RME, SMDV and SMDD cell clusters), just preceding the prediction window, into the Soft Tree gating models. These models output N probabilities where N is the number of associated MLR submodels. During fitting, these probabilities form the weights in a weighted MLR. Thus, a given Soft Tree learns how to allocate data to its different MLR submodels based on recent network activity, while the models themselves learn to predict AVA and RME fluorescence changes. Henceforth, the Soft Tree along with its associated MLR

submodels is referred to as the “hybrid model”. It should be noted that cross-validation predictions are averaged across multiple hybrid models to increase the model’s robustness (see Prediction: hybrid model in Methods for more details).

In this section, there are two relevant time windows. If t is the start of the prediction window, our goal is to predict changes in AVA and RME cell cluster fluorescence from time t to time t+4 seconds. These predictions are the sum of two different hybrid models: the “baseline hybrid model” and the “stimulus hybrid model”. In both these hybrid models, the gating model is fed the network history activity from t-16 seconds to time t (see Prediction task in methods). The models differ in what information is sent to their MLR submodels. The MLR submodels in the baseline model make their predictions from the same network history that is fed into the gating submodels. The MLRs in the stimulus hybrid model are fed On and Off Sensory neurons activity from time t-12 seconds to t+4 seconds. Schematics for these models are shown in Figure 4A and Figure 4B. In figures 4-6, N refers to the number of MLR submodels in the baseline hybrid model, while M refers to the number of MLR submodels in the stimulus hybrid model. In this model setup, we are probing how cellular communication and stimulus drive are gated by recent network activity.

A hybrid model with more states can model more complex dynamics. However, this comes at the cost of having less data per MLR submodel because the dataset is split across more subsets. Our initial goal is then to use cross-validation to decide how to parcel network history space in order to optimally predict AVA and RME activity. For both NeuroPAL and non-NeuroPAL animals, four submodels were found to be ideal for the baseline hybrid model. They significantly improve upon a single MLR model (Figure 4C). The inclusion of the stimulus was found to be useful as well. A stimulus hybrid model with four MLR submodels was found to be

ideal for NeuroPAL animals, while two submodels were found to be best for the non-NeuroPAL transgenics (Figure 4C). Why ideal model parameterization differed across these two strains is unclear. However, despite this difference in the number of states, the models are qualitatively very similar across strains. Also, for both strains, there is sufficient data to find multiple receptive fields for AVA and RME neurons.

As described above, each overall model is composed of a baseline hybrid model and a stimulus hybrid model. Moreover, the cross-validation predictions are averaged across three hybrid models. The 0th overall model's baseline hybrid model is visualized in Figure 5. Its stimulus hybrid model is shown in Figure 6. The three hybrid models used for AVA and RME predictions in the NeuroPAL set are shown in Supplementary Figures S2-4, while those used for the non-NeuroPAL animals are indicated in Supplementary Figures S5-7.

We first inspected the baseline hybrid models to understand how AVA and RME activity depends on recent network history. Figure 5 shows a hybrid model's MLR submodels alongside the network trajectory windows (orange vertical bars) in which these submodels are most active. The most obvious trend is that the dominant filter components almost always occur in the last three seconds before the prediction window. Interestingly, the signatures of momentum and boundedness can be seen in states 2 and 3 in Figure 5B. State 2 is active when AVA is rising from its lower baseline. In this state, AVA rise probability is positively correlated with recent increases in AVA activity (momentum). State 3 is active when AVA nears its upper limit and the aforementioned positive correlation disappears (boundedness). Thus, the hybrid model captures the nonlinear features we expected along with additional interesting features.

In all four network history states identified, increased AVA activity is correlated with a decrease in RME activity, while increased RME activity is associated with decreased AVA

activity, except for State 1 (Figure 5B). This is not surprising as State 1 specifically involves negligible AVA activity thereby eliminating the contribution of AVA responses to RME neural activity prediction (Figure 5B). Moreover, we found that increased RME activity is often correlated with increased SMDD or SMDV activity, while AVA neural activity is associated with SMDD and SMDV inactivity (Figure 5B). This relationship is consistent with previous study showing hierarchical control of activity in the RME-AVA and SMDD-SMDV motor neuron groups (Kaplan et al., 2020). Collectively, these data show that network history plays a crucial role in determining future activity of the RME and AVA motor neurons, thereby modifying animal behavior. RME and AVA motor neurons have been shown to drive forward and backward movement respectively (Chalfie et al., 1985; Kawano et al., 2011; Varshney et al., 2011).

In state 3 in Figure 5B, RME activity from 14 seconds before the prediction window to 5 seconds before the prediction window significantly decreases the probability of RME rising over the prediction window. This is particularly interesting as this state tends to be active when RME is at or near its lower baseline. This suggests that RME activity is rectifying: recent RME activity decreases the probability of more RME activity. This more nuanced trend is hard to capture in a single MLR model as it is relevant in only a small number of network activity windows.

In Figure 6, we show the stimulus hybrid model belonging to overall model 0. Activity in On sensory neurons drives RME neurons and inhibits AVA neurons consistently across all states (Figure 6B). In contrast, Off sensory neurons only drive AVA activity in states 1 and 3, where AVA neurons have low activity (Figure 6B). Some of this state-dependence is due to the boundedness of AVA as states 0 and 2 tend to occur when AVA nears its upper activity bound. However, in state 3, AVA also exhibits a diminished sensitivity to Off sensory neuron activity

and this state occurs when AVA is close to or at its lower activity bound. In fact, the gating models are particularly sensitive to SMDD activity (Supplementary Figures S2-4). When SMDD is high or rising, the influence of Off sensory neurons on AVA activity is diminished. A similar trend can also be seen in the non-NeuroPAL transgenics (Supplementary Figures S5-7). Along with earlier findings on the relationship between SMDD and AVA, this suggests that the animal must complete a dorsal turn before initiating backward locomotion (SMDD drives dorsal turns, while AVA drives backward movement) and that stimulus removal has a limited ability to interrupt this motor pattern. Broadly, SMDV and SMDD (motor neurons that drive dorsal and ventral turns (Kaplan et al., 2020; Varshney et al., 2011)) appear to set the time course for changes in RME and AVA activity. RME rises in concert with SMDV and SMDD activation (these cells can activate with stable RME) (Figure 5B). Similarly, SMDD gates the effect of stimulus on AVA (Figure 6B).

To increase the model robustness, we averaged cross-validation predictions across three hybrid models. This approach also allows us to check our results as these models typically parcel their states differently. Importantly, every result described above is found in multiple hybrid models and across both NeuroPAL and non-NeuroPAL transgenics. Conveniently, all of the NeuroPAL trees exhibit RME rectification in State 3 (Supplementary Figures S2-4). The 1st and 3rd non-NeuroPAL trees (Supplementary Figures S5 and S7) exhibit RME rectification in States 0 and 1. SMDD gating of Off sensory-AVA drive can also be observed in a number of different trees. Stimulus states 2 and 3 show this gating effect for all of the NeuroPAL trees (Supplementary Figures S2-4). The 1st and 3rd non-NeuroPAL trees (Supplementary Figures S5 and S7) also show this gating phenomenon in Stimulus State 0.

The hybrid model identified interesting, nonlinear features of RME and AVA activity. We wondered whether it could reveal insights into SMDD and SMDV activity. We used cross-validation performance on the hyperparameter set for our exploratory analysis (see Prediction Task in Methods for details). Stimulus inclusion was found to be useful for predicting RME and AVA activity in the hyperparameter set. This was also found to be useful in predicting SMDV activity in non-NeuroPAL animals. As a result, the Soft Tree-MLR hybrid model was fit to SMDV activity on the non-NeuroPAL cross-validation dataset. The best performing model used four network history states and two stimulus states (Supplementary Figure S8A). In every network history state, SMDD activity decreased the probability of SMDV rise and increased the probability of SMDV fall (Supplementary Figures S8-S10). Along with the observation that SMDD and SMDV never activate simultaneously, this supports the idea that these cells are mutually inhibitory (Kaplan et al., 2020). Interestingly, the effect of stimulus on SMDV activity is strongly gated by RME and AVA activity. When RME is low (and AVA is high), On Sensory neuron activity increased SMDV rise probability, while Off neuron activity decreased SMDV rise probability (this can also be seen in the full decoding model in supplemental figure 1). Thus, when RME is low, bacterial stimulus tends to drive RME and SMDV. However, when RME is activating, stimulus removal yields additional SMDV events (Supplementary Figures S8-S10). This valence agnosticism in SMDV differs substantially from RME and AVA neural activity.

Discussion

We monitor *C. elegans* whole brain neural activity while delivering random patterns of bacterial stimulus to the animal's nose, and identify six groups of active neurons: two sensory and four motor neuron groups. We use first order linear ordinary differential equations to model sensory

neuron activity and show that neurons responding to removal of stimulus likely adapt faster than neurons activated by stimulus addition. Next, we use multinomial logistic regression to show that the bacterial stimulus drives a transition from the backward to forward motor neuron groups. We also combine this method with Soft Trees and identify a role for both stimulus and network state in predicting neural activity. These studies leverage statistical and mathematical methods to link stimulus to whole-brain activity and animal behavior.

Previous studies have used principal components analysis to show that *C. elegans* whole brain activity is primarily driven by motor neurons and is organized into hierarchical patterns of network states (Kaplan et al., 2020; Kato et al., 2015). These studies trapped animals in a two-layered microfluidic device and monitored neural activity while delivering different levels of oxygen stimuli (Zimmer et al., 2009). In contrast, we trapped animals in a single-layered device and identified both sensory and motor neuron activity while delivering patterns of bacterial stimuli. Our findings are consistent with previous studies showing that while ASI chemosensory neurons respond to the addition a bacterial stimulus, AWC and ASK are activated by its removal (Calhoun et al., 2015; Chalasani et al., 2007). Sensory neuronal responses to the addition of bacterial stimuli ("On") and its removal ("Off") can be modeled by two linear ordinary differential equations using one fast and one slow filter. Linearity in sensory neurons is observed in both vertebrates and invertebrate photoreceptors (Pahlberg and Sampath, 2011), rat trigeminal neurons (Jones et al., 2004), primate vestibular neurons (Massot et al., 2012), with more complex features emerging in higher-order neurons (Lochmann and Deneve, 2011).

We also show that "On" sensory neurons perfectly adapt to the bacterial stimulus, consistent with previous studies (Kato et al., 2014), while "Off" neurons adapt imperfectly. Similar differences in adaptation are also observed in single-cell electrophysiological recordings

from vertebrate photoreceptors and olfactory sensory neurons (De Palo et al., 2013). Altogether, these results provide some hints about the encoding strategies of these two sensory-neuron classes. We suggest that the “On” and “Off” neurons encode stimuli over a large and limited range of stimulus concentrations, respectively. Consistently, we have previously shown that AWA sensory neurons (“On”) have a larger dynamic range compared to AWC sensory neurons (“Off”) in detecting Benzaldehyde (Leinwand et al., 2015). Moreover, studies in the vertebrate retina have shown that the dynamic range of the ON pathway is much greater than that of the OFF one, likely due to a selective effect of pre-synaptic inhibition on the ON, but not OFF bipolar cells (Sagdullaev et al., 2006). Collectively, these results imply that “On” sensory neurons might detect a larger dynamic range of stimulus concentrations compared to “Off” sensory neurons.

We also identify state dependence in motor neurons (AVA and RME) by using Soft Trees to section the neural network history into distinct subspaces. These Soft Trees then learn the relevant and interpretable space divisions that vastly improve model performance. Our Soft Tree-multinomial logistic regression model accounts for the boundedness and stereotypy in AVA and RME neural activity. While Soft Trees have been previously used in classification tasks (Frosst and Hinton, 2017), combining this method with a multinomial logistic regression is novel. Our model suggests that addition of a bacterial stimulus biases the network towards increased RME activity, while its prolonged removal drives AVA activity. This is consistent with previous studies showing that addition of food-associated odors drives forward movement, while its removal triggers reversal and pirouettes (Albrecht and Bargmann, 2011). In addition, we find that activity in SMDD and SMDV neurons influences and likely sets the time course for RME and AVA neural activity. A central role for SMD neurons was also previously identified from

whole-brain imaging studies, as part of a hierarchical model of neural activity (Kaplan et al., 2020).

Our study identifies six groups of active neurons in animals experiencing bacterial stimulus and uses computational and statistical methods to identify relationships between them. Specifically, we show that activity in AVA and RME neurons is modified by both neural network history and sensory input. We suggest that our approach of combining soft trees with multinomial logistic regression can be broadly used to identify relationships, not only in neural networks, but also in cellular signaling pathways, transcription factor networks, and other complex biological and physical entities.

Materials and Methods

Strains and calcium imaging

We used two transgenic animals expressing GCaMP5k (ZIM294) and GCaMP6s (OH15500) in all neural nuclei and monitored changes in GCaMP fluorescence using a Zeiss Airyscan880. Day 1 adult animals were trapped in a microfluidic device (Chronis et al., 2007). We delivered precise patterns of a 1:1 resuspension of a bacterial culture (OD_{600} = 0.4) in LB media as previously described (Calhoun et al., 2015), using a custom designed Arduino device. We then used piece-wise rigid registration to remove motion artefacts (Pnevmatikakis and Giovannucci, 2017) and non-negative matrix factorization to isolate individual neurons and extract their fluorescence values (Pnevmatikakis et al., 2016).

GCaMP filters

The GCaMP filter *g(x)* is modeled as a difference of exponentials with parameters matching those of Chen et al (Chen et al., 2013). This procedure is complicated by the volumetric nature of the imaging data. Consider two sensory neurons with identical calcium dynamics; they respond to stimulus with the same timescale. Neuron A is in imaging slice 0, while neuron B is in slice 7. These two neurons will have the same calcium timings relative to the stimulus. However, Neuron B will appear to have faster response kinetics since it is acquired over half a second later (relative to stimulus on/offset) compared to neuron A. Thus, the slice in which the neuron appears needs to be considered in the creation of its GCaMP filter:

$$g(t) = \sum_{t_1}^{t_2} g(x)\, dx)$$

$$t_1 = t + z/n_z$$

$$t_2 = t + (z+1)/n_z$$

where *g(x)* refers to the difference of exponentials and n_z refers to the number of z slices. The normalized *g(t)* filter is applied via a linear convolution to transition between calcium and GCaMP dynamics.

Sensory neuron model

The core primary sensory model consists of three parts: (1) a set of cascade basis functions, (2) coefficients for the basis functions, and (3) a GCaMP transformation.

$$dx_i/dt = tau * ([input] - x_i(t))$$

$$X(t) = \sum_i a_i . x_i(t)$$

$$g(t) = (X(t)^p * f_g)(t)$$

A random search strategy was used to find the correct set of basis functions. The strategy chooses *N* basis time constants for an *N* cascade model and an initial estimate of *p* power. The

cascades are solved analytically for a given stimulus pattern. An initial estimate for the coefficients are found by fitting to an approximation of the neuron's calcium trace. In this approximation, the raw sensory neuron fluorescence trace is deconvolved using the Richardson-Lucy method (Fish et al., 1995) and taken to the *(1/p)* power. The initial estimate for the coefficients is the solution to the resulting linear regression equation. Finally, the full model (free variables consist of basis coefficients and *p* power) is fit to the raw fluorescence trace using gradient descent. This process is repeated for a large number of random searches to define the basis functions.

Blockstrap (used for both decoding and predictions)

For both the decoding and prediction analyses, the data is split into contiguous blocks of ~10 second (16 volume) duration with sub windows. For instance, for an N volume prediction window, there are 16-N legal, overlapping sub windows within each block. Moreover, there is no overlap between prediction windows of adjacent blocks. Test/train sampling is done at the block level, guaranteeing no train/test prediction overlap. This system captures sharp transitions while measuring targets over multiple time bins, while limiting the noise in the targets. The length of 16 volumes was chosen as it captures the entirety of motor neuron event initiation (for all classes). This causes bootstrap sampling to be performed at the event level, stopping a small, handful of events from dominating model outcomes. Other blockstrap sizes were not tested.

Gaussian basis functions

Gaussian basis functions were used for both decoding and prediction tasks. Network history and stimulus data were filtered using these basis functions before being fed into decoding and

prediction models. Specific gaussians (parametrized by mean and variance) were chosen by hand-tuning model performance on the hyperparameter set.

Decoding targets – stimulus on, off, on-off, or off-on transitions

Given the relative simplicity of the binary stimulus patterns, a small number of simple features can capture most of the stimulus variance. Five decoding targets were chosen for ~5 second (8 volume) time windows: stimulus on for duration, stimulus off for duration, stimulus on transition, stimulus off transition, and multiple transitions. Thus, there are 16-8=8 prediction windows within each time block [see Blockstrap above for details]. Unlike the prediction task, the decoding task centers the prediction windows relative to the input data. For instance, in one decoding task, RME/AVA cell cluster data from t+4-8 to t+4+8 volumes is used to predict whether the stimulus is on/off/altered from t to t+8 volumes. The unit of 8 volumes was chosen because many of the pulse protocols use 10 volume pulses as a base unit. Thus, using 8 volumes guarantees enough samples of the on stimulus class. Other volume lengths were not tested.

Decoding – class balancing

In the decoding analyses, five classes of stimulus patterns are predicted from network activity (Paninski et al., 2007). Prediction from worm identity alone serves as the null model for all decoding analyses. In order to ensure best performance for the null model, classes are balanced within each worm. For each worm, if there are N occurrences of class A in the training set, there are N occurrences of class A in the test set. This is done by randomly removing prediction windows.

Decoding/Prediction: one hyperparameter training set and 1 test set

For both the prediction and decoding analyses, the data blocks (see Blockstrap for details] were split into four sets: OH15500 hyperparameter set, OH15500 cross-validation set, ZIM294 hyperparameter set, and ZIM294 cross-validation set. For each of these analyses, a leave-N-out strategy was used to choose hyperparameters (using only the hyperparameter set). The different models were then compared via out-of-bootstrap cross-validation on the cross-validation set (Tsamardinos et al., 2018). There is overlap between the hyperparameter set and the cross-validation training sets.

L1-norm regularization is used for both decoding and prediction analyses. The prediction analysis has additional hyperparameters. It includes different L1 norms for each of the multinomial logistic regression input classes: 1. AVA/RME/SMDV/SMDD terms, 2. On/Off cell terms, 3. worm identity terms. The worm identity terms essentially make the prediction analyses random-intercept models (Bates et al., 2014). All of the L1 regularization coefficients are chosen on the hyperparameter set.

Prediction task

In the prediction task, a change in GCaMP fluorescence from t to t+T1 is predicted using network history from t-T2 to t and sensory neuron activity from t-T2+T1 to t+T1. A length of twenty-four volumes are used for T2 for all prediction analyses. The length of T1 is treated as an additional hyperparameter. As in the decoding analysis, these prediction windows are contained inside larger data blocks. Train/test sampling is done at the block level (See Blockstrap for more details). Sixteen seconds (24 volumes) were used for T1 in all models as 16 seconds is sufficient to capture the majority of command neuron events.

RME and AVA activation and inactivation is clearly probabilistic. We used a multinomial logistic regression (MLR) as the base model for motor-neuron activity prediction. In order to use MLR, GCaMP fluorescence activity is discretized in every prediction window. This is done by subtracting the average GCaMP fluorescence over the prediction window by the GCaMP fluorescence level at the beginning of that window. This average difference quantity is then binned and treated as additional hyperparameters. This approach was particularly useful to predict the relationship between stimulus and SMD neurons. On the hyperparameter set, the stimulus was found to provide the greatest improvement in model performance for only three bins and a T2 of 8 volumes. Using more bins led to massive overfitting and using shorter T2 lengths introduces too much target noise. Additionally, incorporating stimulus was found to increase predictability of RME and AVA for all T2 and binning choices. Consequently, eight bins and a T2 of 4 volumes were used for RME and AVA in the cross-validation set.

Prediction: hybrid model

We use a Soft Tree - Multinomial Logistic Regression (MLR) hybrid method to divide the network trajectory space into different subspaces. GCaMP fluorescence changes in each of these subspaces is assumed to be well-described by an MLR. A Soft Tree is a form of oblique decision tree that is end-to-end differentiable (Frosst and Hinton, 2017). Each branch of the Soft Tree is a different logistic regression model on the same input vector, which output a left vs right probability. These left and right probabilities are multiplied by the predicted class probabilities of the corresponding left and right subtrees. In this manner, Soft Trees are essentially hierarchical filters that can be learned through gradient descent. The outputs of the Soft Tree weigh the different MLR models. For instance, a Soft Tree with depth 2 and width 2 will have 4 output

states. Each of these output states are associated with a different MLR. All of these MLRs are trained against the entire dataset. However, the data points are weighted by the particular Soft Tree. Hence, different MLR models will focus on different subsets of the data.

We found that averaging predictions across hybrid models improved cross-validation performance. This averaging is done at two levels. First, within a hybrid model, the predictions of each MLR are weighted by the Soft Tree and averaged. Second, these averaged predictions are further averaged across several hybrid models. Here, the full hybrid models are fit separately to the hyperparameter set. The best three hybrid models were chosen for analysis on the cross-validation set. For training on the cross-validation set, the Soft Trees are frozen; only the MLRs are trained. Hence, training is convex (weighted multinomial logistic regression) on the cross-validation set. Reported log-likelihoods are calculated using the averaged predictions of the three hybrid models.

Prediction: Boosting analysis

There is a performance tradeoff involved in the hybrid Soft Tree-MLR system. Deeper trees with more states can learn more complex features by sorting the network trajectory data into smaller units. However, this involves learning more models, requiring more data for good performance. This problem is solved with a boosting strategy across the hybrid models. The first hybrid model outputs predictions based on network state. The second hybrid model outputs predictions based on the stimulus (network state into tree; stimulus into associated MLR). These outer products form the weights of the weighted multinomial logistic regression, which are then then averaged across states and models. We limited our analysis to double boosting in this study. Every network state and stimulus MLR model combination is fit to the whole dataset. The weights for

these model combinations are given by the outer product of the network state and stimulus Soft Trees. Similarly, predictions are averaged across states and three models for cross-validation.

Acknowledgements

We thank Manuel Zimmer, Oliver Hobert and CGC (Caenorhabditis Genetics Center) for worm strains and Uri Manor, Tong Zhang and the Waitt Advanced Biophotonics Center for advice and assistance with our imaging experiments. We also thank Saket Navlakha, Ryan Rowekamp, Javier How, Jessica Haley, Michael Rieger, and Eric Edsinger for discussions and comments on the manuscript. This work was supposed by grants from the National Institutes of Health NIH 5R01MH096881 (S.H.C); 5T32DK7328-37, 5T32DK007328-35, 5T32MH015174-38, and 5T32MH015174-37 (E.Y). Z.T.C. was supported by a fellowship from the Dan and Marina Lewis Foundation. CGC is funded by the NIH Office of Research Infrastructure Programs (P40 OD010440).

Competing interests

The authors declare that no competing interests exist.

References:

Ahrens, M.B., Li, J.M., Orger, M.B., Robson, D.N., Schier, A.F., Engert, F., and Portugues, R. (2012). Brain-wide neuronal dynamics during motor adaptation in zebrafish. Nature *485*, 471-477.

Ahrens, M.B., Orger, M.B., Robson, D.N., Li, J.M., and Keller, P.J. (2013). Whole-brain functional imaging at cellular resolution using light-sheet microscopy. Nat Methods *10*, 413-420.

Akerboom, J., Carreras Calderon, N., Tian, L., Wabnig, S., Prigge, M., Tolo, J., Gordus, A., Orger, M.B., Severi, K.E., Macklin, J.J., *et al.* (2013). Genetically encoded calcium indicators for multi-color neural activity imaging and combination with optogenetics. Front Mol Neurosci *6*, 2.

Albrecht, D.R., and Bargmann, C.I. (2011). High-content behavioral analysis of Caenorhabditis elegans in precise spatiotemporal chemical environments. Nat Methods *8*, 599-605.
Antoshechkin, I., and Sternberg, P.W. (2007). The versatile worm: genetic and genomic resources for Caenorhabditis elegans research. Nat Rev Genet *8*, 518-532.
Bargmann, C.I. (2006). Chemosensation in C. elegans. WormBook, 1-29.
Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:14065823.
Blackwell, J.M., and Geffen, M.N. (2017). Progress and challenges for understanding the function of cortical microcircuits in auditory processing. Nature communications *8*, 2165.
Calhoun, A.J., Tong, A., Pokala, N., Fitzpatrick, J.A., Sharpee, T.O., and Chalasani, S.H. (2015). Neural Mechanisms for Evaluating Environmental Variability in Caenorhabditis elegans. Neuron *86*, 428-441.
Chalasani, S.H., Chronis, N., Tsunozaki, M., Gray, J.M., Ramot, D., Goodman, M.B., and Bargmann, C.I. (2007). Dissecting a circuit for olfactory behaviour in Caenorhabditis elegans. Nature *450*, 63-70.
Chalfie, M., Sulston, J.E., White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. (1985). The neural circuit for touch sensitivity in Caenorhabditis elegans. J Neurosci *5*, 956-964.
Chen, T.W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., *et al.* (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature *499*, 295-300.
Chronis, N., Zimmer, M., and Bargmann, C.I. (2007). Microfluidics for in vivo imaging of neuronal and behavioral activity in Caenorhabditis elegans. Nat Methods *4*, 727-731.
Churchland, M.M., Cunningham, J.P., Kaufman, M.T., Foster, J.D., Nuyujukian, P., Ryu, S.I., and Shenoy, K.V. (2012). Neural population dynamics during reaching. Nature *487*, 51-56.
De Palo, G., Facchetti, G., Mazzolini, M., Menini, A., Torre, V., and Altafini, C. (2013). Common dynamical features of sensory adaptation in photoreceptors and olfactory sensory neurons. Scientific reports *3*, 1-8.
Dubs, A., Laughlin, S.B., and Srinivasan, M.V. (1981). Single photon signals in fly photoreceptors and first order interneurones at behavioral threshold. J Physiol *317*, 317-334.
Dugatkin, L.A. (2020). Principles of animal behavior (University of Chicago Press).
Fish, D., Brinicombe, A., Pike, E., and Walker, J. (1995). Blind deconvolution by means of the Richardson–Lucy algorithm. JOSA A *12*, 58-65.
Frosst, N., and Hinton, G. (2017). Distilling a neural network into a soft decision tree. arXiv preprint arXiv:171109784.
Hallock, R.M., and Di Lorenzo, P.M. (2006). Temporal coding in the gustatory system. Neurosci Biobehav Rev *30*, 1145-1160.
Hilliard, M.A., Apicella, A.J., Kerr, R., Suzuki, H., Bazzicalupo, P., and Schafer, W.R. (2005). In vivo imaging of C. elegans ASH neurons: cellular response and adaptation to chemical repellents. Embo J *24*, 63-72.
How, J.J., Navlakha, S., and Chalasani, S.H. (2020). Neural network features distinguish chemosensory stimuli in <em>Caenorhabditis elegans</em&gt. bioRxiv, 2020.2002.2018.955245.
Ibsen, S., Tong, A., Schutt, C., Esener, S., and Chalasani, S.H. (2015). Sonogenetics is a non-invasive approach to activating neurons in Caenorhabditis elegans. Nature communications *6*, 8264.

Jones, L.M., Depireux, D.A., Simons, D.J., and Keller, A. (2004). Robust temporal coding in the trigeminal system. Science *304*, 1986-1989.
Kaplan, H.S., Salazar Thula, O., Khoss, N., and Zimmer, M. (2020). Nested Neuronal Dynamics Orchestrate a Behavioral Hierarchy across Timescales. Neuron *105*, 562-576 e569.
Kato, S., Kaplan, H.S., Schrodel, T., Skora, S., Lindsay, T.H., Yemini, E., Lockery, S., and Zimmer, M. (2015). Global brain dynamics embed the motor command sequence of Caenorhabditis elegans. Cell *163*, 656-669.
Kato, S., Xu, Y., Cho, C.E., Abbott, L.F., and Bargmann, C.I. (2014). Temporal responses of C. elegans chemosensory neurons are preserved in behavioral dynamics. Neuron *81*, 616-628.
Kawano, T., Po, M.D., Gao, S., Leung, G., Ryu, W.S., and Zhen, M. (2011). An imbalancing act: gap junctions reduce the backward motor circuit activity to bias C. elegans for forward locomotion. Neuron *72*, 572-586.
Leinwand, S.G., Yang, C.J., Bazopoulou, D., Chronis, N., Srinivasan, J., and Chalasani, S.H. (2015). Circuit mechanisms encoding odors and driving aging-associated behavioral declines in Caenorhabditis elegans. eLife *4*, e10181.
Lemon, W.C., Pulver, S.R., Hockendorf, B., McDole, K., Branson, K., Freeman, J., and Keller, P.J. (2015). Whole-central nervous system functional imaging in larval Drosophila. Nature communications *6*, 7924.
Lochmann, T., and Deneve, S. (2011). Neural processing as causal inference. Curr Opin Neurobiol *21*, 774-781.
Massot, C., Schneider, A.D., Chacron, M.J., and Cullen, K.E. (2012). The vestibular system implements a linear-nonlinear transformation in order to encode self-motion. PLoS Biol *10*, e1001365.
Nguyen, J.P., Shipley, F.B., Linder, A.N., Plummer, G.S., Liu, M., Setru, S.U., Shaevitz, J.W., and Leifer, A.M. (2016). Whole-brain calcium imaging with cellular resolution in freely behaving Caenorhabditis elegans. Proceedings of the National Academy of Sciences *113*, E1074-E1081.
Nichols, A.L.A., Eichler, T., Latham, R., and Zimmer, M. (2017). A global brain state underlies C. elegans sleep behavior. Science *356*.
Pahlberg, J., and Sampath, A.P. (2011). Visual threshold is set by linear and nonlinear mechanisms in the retina that mitigate noise: how neural circuits in the retina improve the signal-to-noise ratio of the single-photon response. Bioessays *33*, 438-447.
Paninski, L., Pillow, J., and Lewi, J. (2007). Statistical models for neural encoding, decoding, and optimal stimulus design. Progress in brain research *165*, 493-507.
Pnevmatikakis, E.A., and Giovannucci, A. (2017). NoRMCorre: An online algorithm for piecewise rigid motion correction of calcium imaging data. J Neurosci Methods *291*, 83-94.
Pnevmatikakis, E.A., Soudry, D., Gao, Y., Machado, T.A., Merel, J., Pfau, D., Reardon, T., Mu, Y., Lacefield, C., Yang, W., *et al.* (2016). Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. Neuron *89*, 285-299.
Reidenbach, M.A., and Koehl, M.A. (2011). The spatial and temporal patterns of odors sampled by lobsters and crabs in a turbulent plume. J Exp Biol *214*, 3138-3153.
Riffell, J.A., Abrell, L., and Hildebrand, J.G. (2008). Physical processes and real-time chemical measurement of the insect olfactory environment. J Chem Ecol *34*, 837-853.
Sagdullaev, B.T., McCall, M.A., and Lukasiewicz, P.D. (2006). Presynaptic inhibition modulates spillover, creating distinct dynamic response ranges of sensory output. Neuron *50*, 923-935.

Schrodel, T., Prevedel, R., Aumayr, K., Zimmer, M., and Vaziri, A. (2013). Brain-wide 3D imaging of neuronal activity in Caenorhabditis elegans with sculpted light. Nat Methods *10*, 1013-1020.
Skora, S., Mende, F., and Zimmer, M. (2018). Energy Scarcity Promotes a Brain-wide Sleep State Modulated by Insulin Signaling in C. elegans. Cell reports *22*, 953-966.
Suzuki, H., Thiele, T.R., Faumont, S., Ezcurra, M., Lockery, S.R., and Schafer, W.R. (2008). Functional asymmetry in Caenorhabditis elegans taste neurons and its computational role in chemotaxis. Nature *454*, 114-117.
Tsamardinos, I., Greasidou, E., and Borboudakis, G. (2018). Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation. Machine Learning *107*, 1895-1922.
Varshney, L.R., Chen, B.L., Paniagua, E., Hall, D.H., and Chklovskii, D.B. (2011). Structural properties of the Caenorhabditis elegans neuronal network. PLoS Comput Biol *7*, e1001066.
Venkatachalam, V., Ji, N., Wang, X., Clark, C., Mitchell, J.K., Klein, M., Tabone, C.J., Florman, J., Ji, H., Greenwood, J., *et al.* (2016). Pan-neuronal imaging in roaming Caenorhabditis elegans. Proc Natl Acad Sci U S A *113*, E1082-1088.
White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. Philos Trans R Soc Lond B Biol Sci *314*, 1-340.
Williamson, R.C., Doiron, B., Smith, M.A., and Yu, B.M. (2019). Bridging large-scale neuronal recordings and large-scale network models using dimensionality reduction. Curr Opin Neurobiol *55*, 40-47.
Yemini, E., Lin, A., Nejatbakhsh, A., Varol, E., Sun, R., Mena, G.E., Samuel, A.D.T., Paninski, L., Venkatachalam, V., and Hobert, O. (2021). NeuroPAL: A Multicolor Atlas for Whole-Brain Neuronal Identification in C. elegans. Cell *184*, 272-288 e211.
Yu, W.M., and Goodrich, L.V. (2014). Morphological and physiological development of auditory synapses. Hear Res *311*, 3-16.
Zimmer, M., Gray, J.M., Pokala, N., Chang, A.J., Karow, D.S., Marletta, M.A., Hudson, M.L., Morton, D.B., Chronis, N., and Bargmann, C.I. (2009). Neurons detect increases and decreases in oxygen levels using distinct guanylate cyclases. Neuron *61*, 865-879.

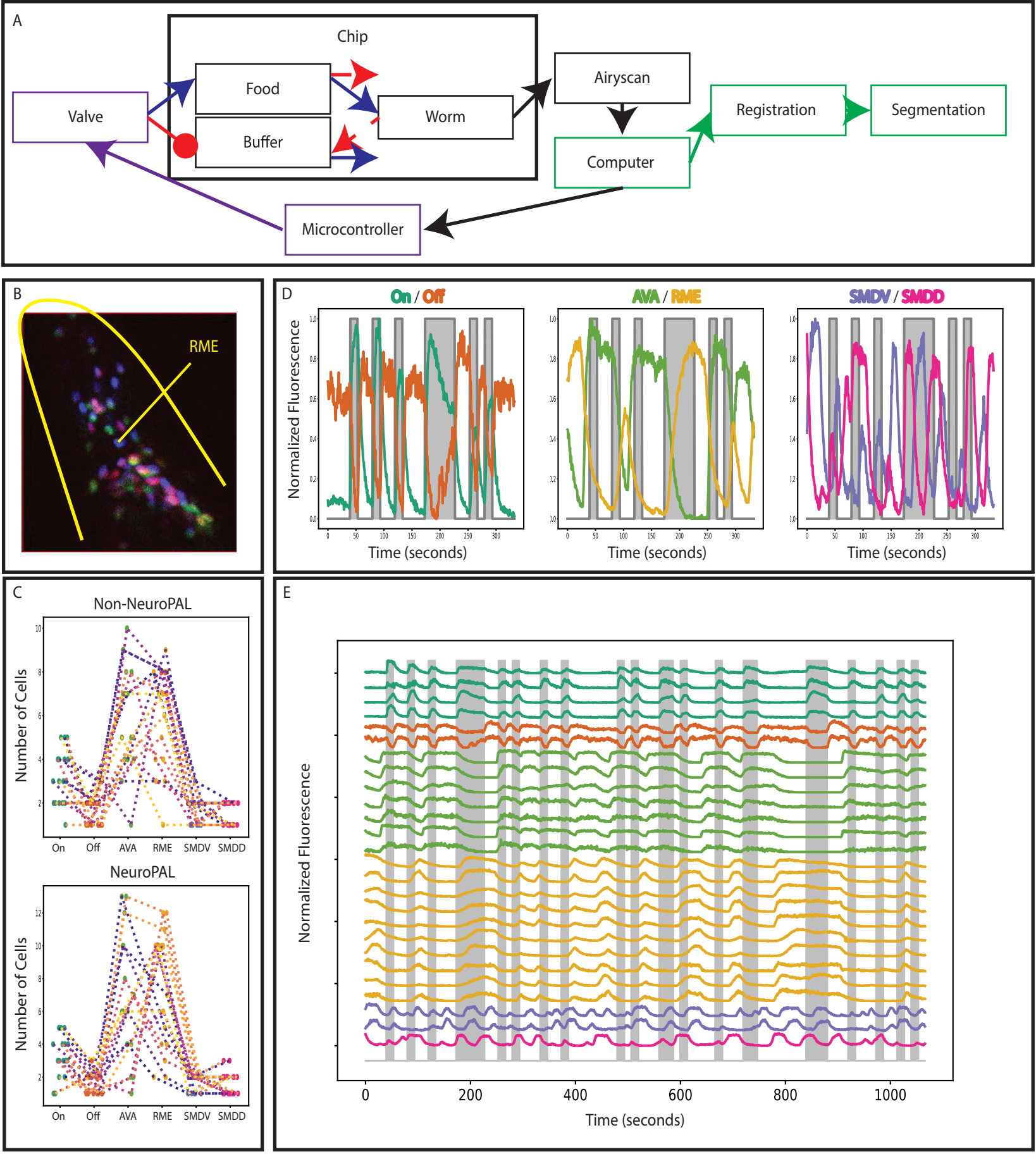

Figure 1. C. elegans full-brain imaging reveals six cell clusters. (A) Imaging setup. Worms are restrained in a microfluidic chip. A microcontroller determines whether odor or buffer is flowing to the head of the worm via a valve. The microcontroller synchronizes valve control with volumetric image acquisition from a Zeiss Airyscan microscope. Image stacks are run through registration and segmentation processing to obtain calcium traces for individual cells. (B) Single z-slice of a NeuroPAL worm. The outline of the worm is drawn and an RME cell is identified. (C) Number of cells in each cell cluster across NeuroPAL and Non-NeuroPAL animals. Each line is a different worm. There are 17 worms in each set. (D) Averaged normalized calcium traces for each of the 6 cell clusters for a single worm. (E) Normalized calcium traces for the same worm as in (D). The trace colors indicate cell identity throughout the paper.

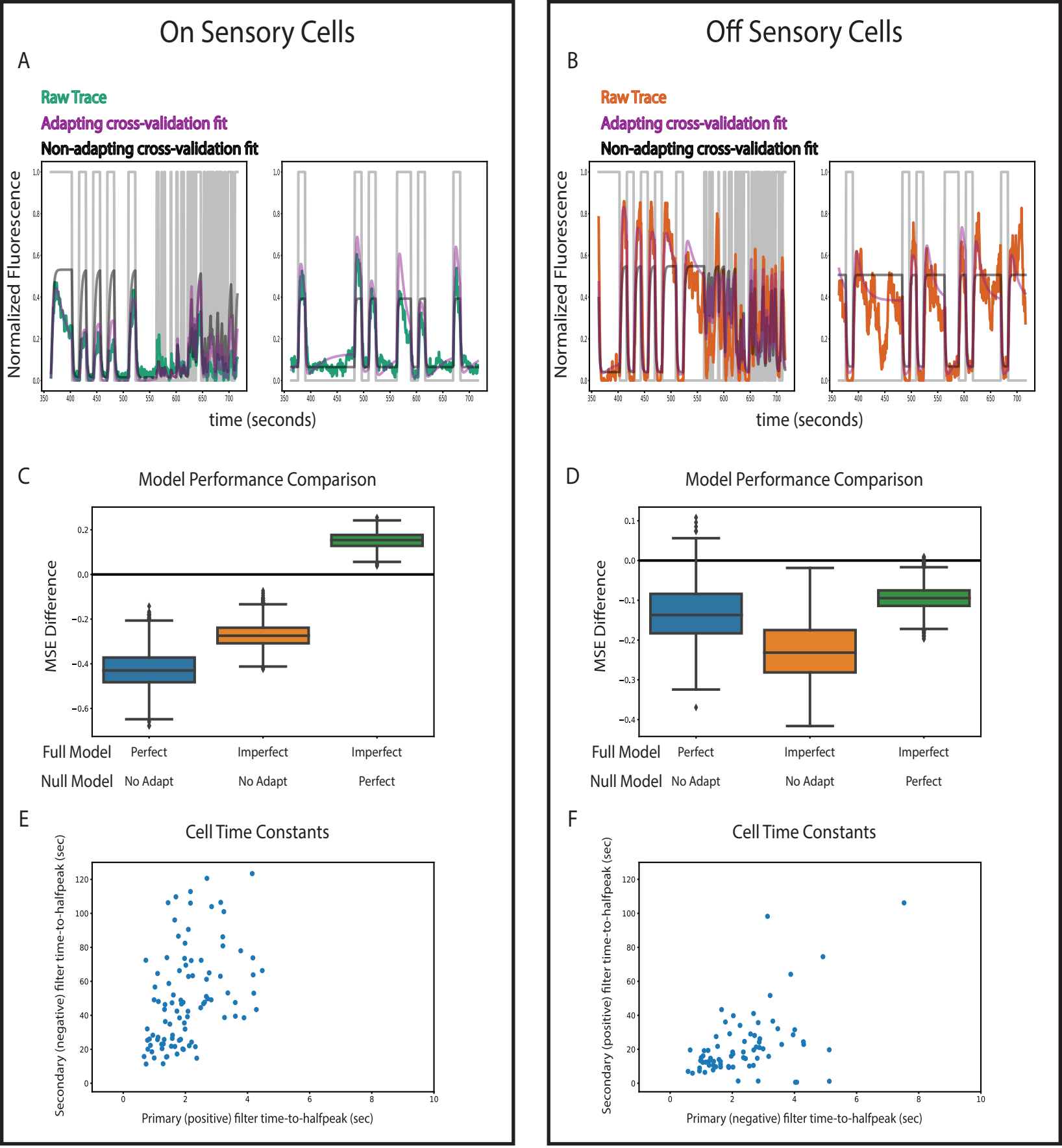

Figure 2. On and Off sensory cells are described by different adaptation models. (A,B) Adaptation models capture sensory cell dynamics. Calcium traces for On and Off sensory cells are shown for two different worms. Cross-validation model predictions are plotted on top of the observed calcium traces. The black traces are the non-adapting model predictions. In (A), the red trace is the two-state, perfect adaptation model prediction. In (B), the red trace is the two-state, imperfect adaptation model prediction. (C,D) Model performance comparison for On (C) and Off (D) cell types. MSE difference refers to the change in mean-squared-error from the null model to the full model. The boxplots indicate the distribution of MSE differences across hierarchical bootstraps. If the boxplot is below 0, the full model has a smaller MSE for all hierarchical bootstraps. The perfect adaptation model is best for On cells (C). The imperfect adaptation model is best for Off cells (D). (E,F) Time-to-half-peak distributions for the best performing On and Off cell models. The two-state adapting model describes calcium traces as the sum of two filters (ODE model solutions). In On cells, the negative filter is slower than the positive filter, producing rapid response and habituation (E). The secondary (adapting) filter (y-axis) is generally slower in On cells than Off cells.

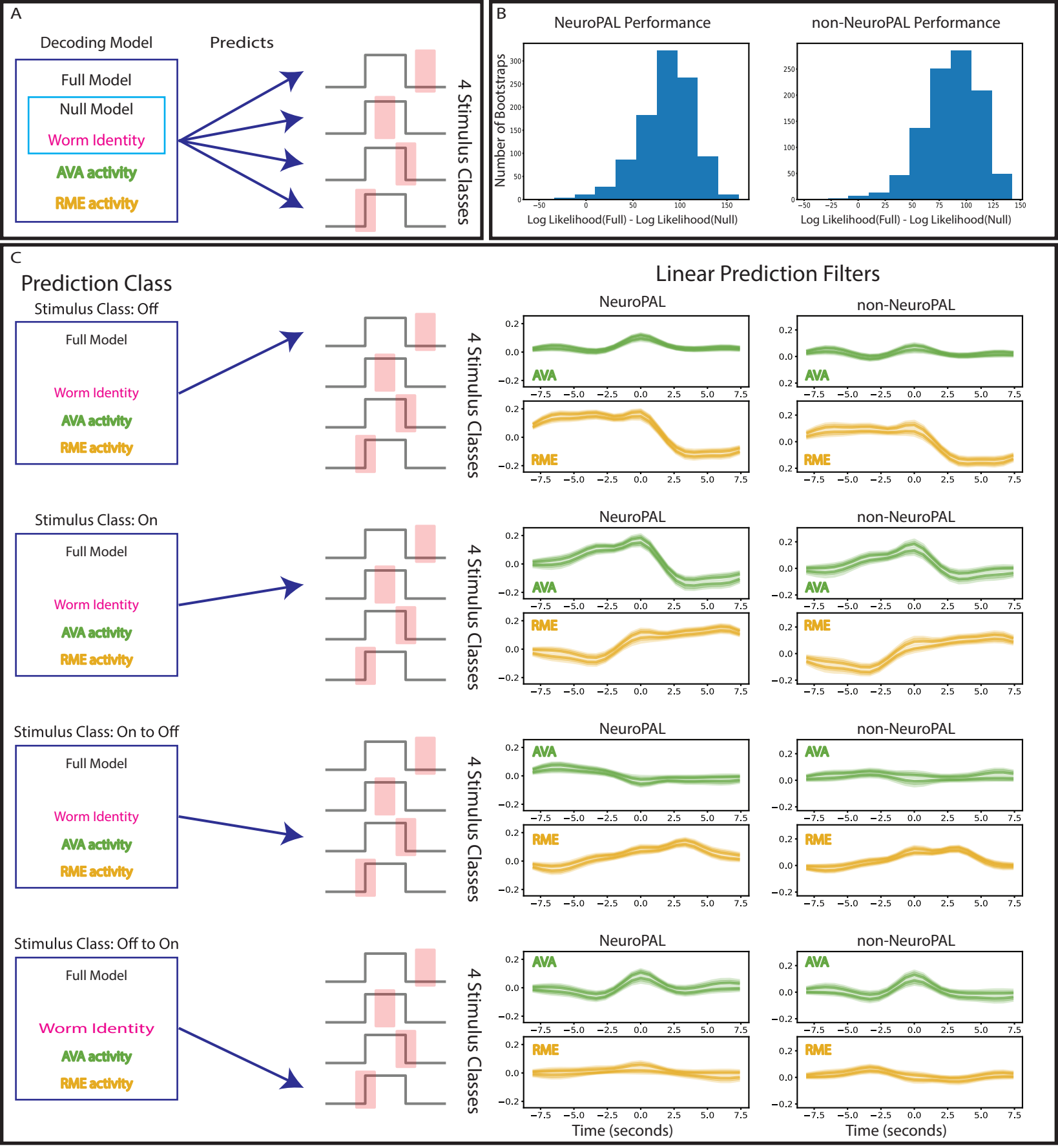

Figure 3. Linear decoder successfully predicts stimulus state from RME and AVA cell cluster activities. (A) Decoding model schematic. AVA and RME activity is used along with worm identity to predict 4 stimulus states: stimulus off, stimulus on, stimulus off transition, and stimulus on transition. In the null model, only worm identity is used. In the full model, AVA and RME activity is used in addition to worm identity. (B) Out-of-bootstrap cross-validation performance for NeuroPAL (left) and non-NeuroPAL (right) worms. Log-likelihood is nearly always increased with inclusion of AVA and RME activity. (C) Visualization of the linear AVA/RME filters for the 4 stimulus classes: stimulus off (top), stimulus on, stimulus off transition, stimulus on transition (bottom). Each filter has two sets of lines. The thicker set of lines contains the middle 50% of bootstraps. The finer set of lines contains the middle 90% of bootstraps.

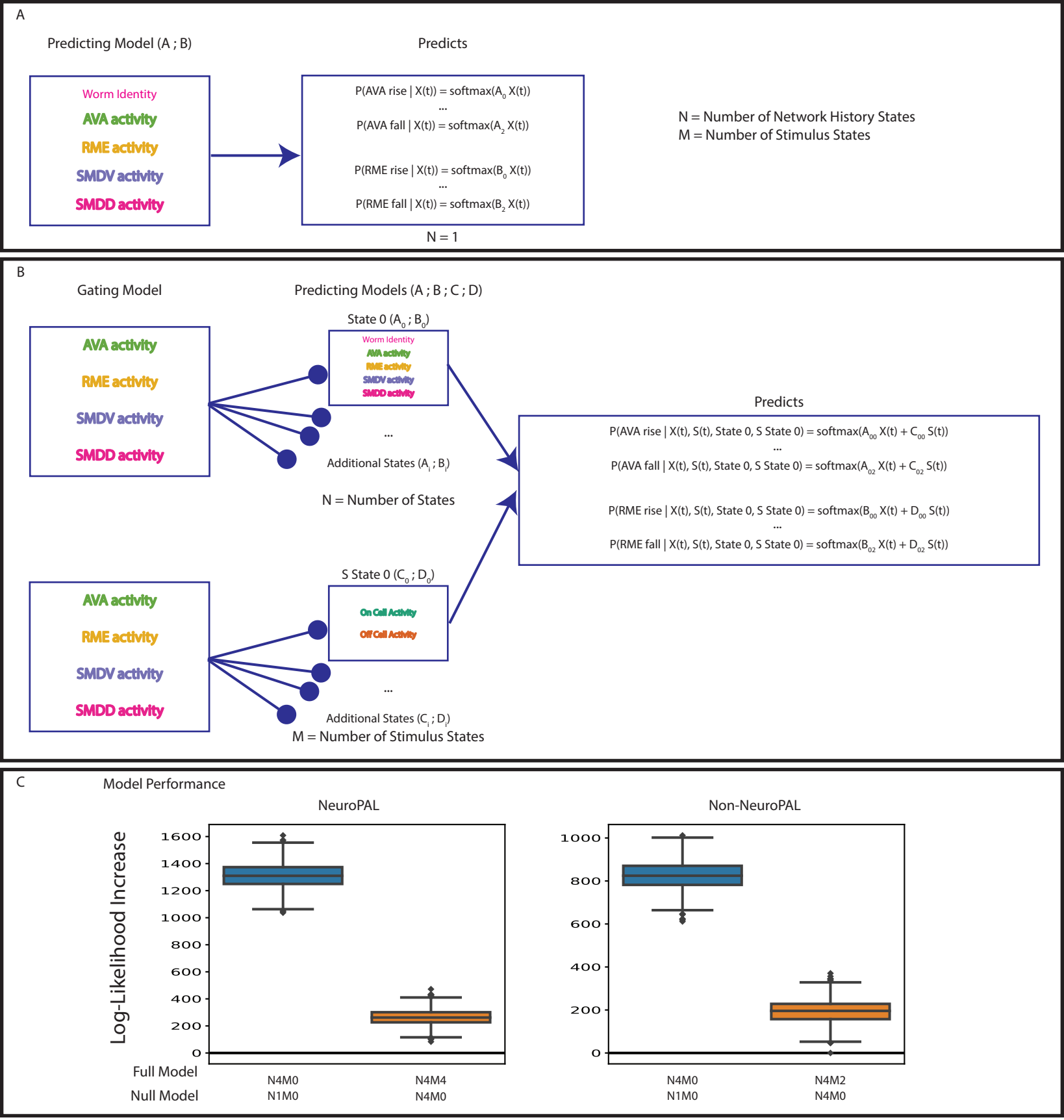

Figure 4. Model schematics and cross-validation performances for predicting models. (A) Model schematic for multinomial, linear prediction model. Worm identity and network history (X(t)) are used to predict the probability of changes in RME and AVA cell cluster fluorescence. (B) Model schematic for the Soft-Tree - MLR hybrid model. Network history is fed into a Soft-Tree gating model. The gating model outputs probabilities for a number of states. Each state is associated with a different multinomial logistic regression model (MLR). MLRs are fit using weighted multinomial logistic regression where the weights are given by the gating model output probabilities. Separate models are used for network history prediction and stimulus prediction. N refers to the number of MLR submodels (equal to number of states) that are fed network history. M is the number of submodels that are fed On and Off sensory cell activity. When M,N > 0, all M*N MLR combinations are fit to the data. The weights for these fits are given by the outer product of the two gating trees. The equations for one of these combinations is shown. When M=0, stimulus information is not used for prediction. When M=0 and N=1, the tree-MLR hybrid model simplifies to the model in (A). (C) Out-of-bootstrap cross-validation performance for several models. As an example, N4M2 means that 4 states are used for network history prediction and 2 states are used for stimulus prediction. N4 models were found to be the best models for network history only prediction. N4M4 was found to be the best model for NeuroPAL animals, while N4M2 was found to be the best model for non-NeuroPAL animals.

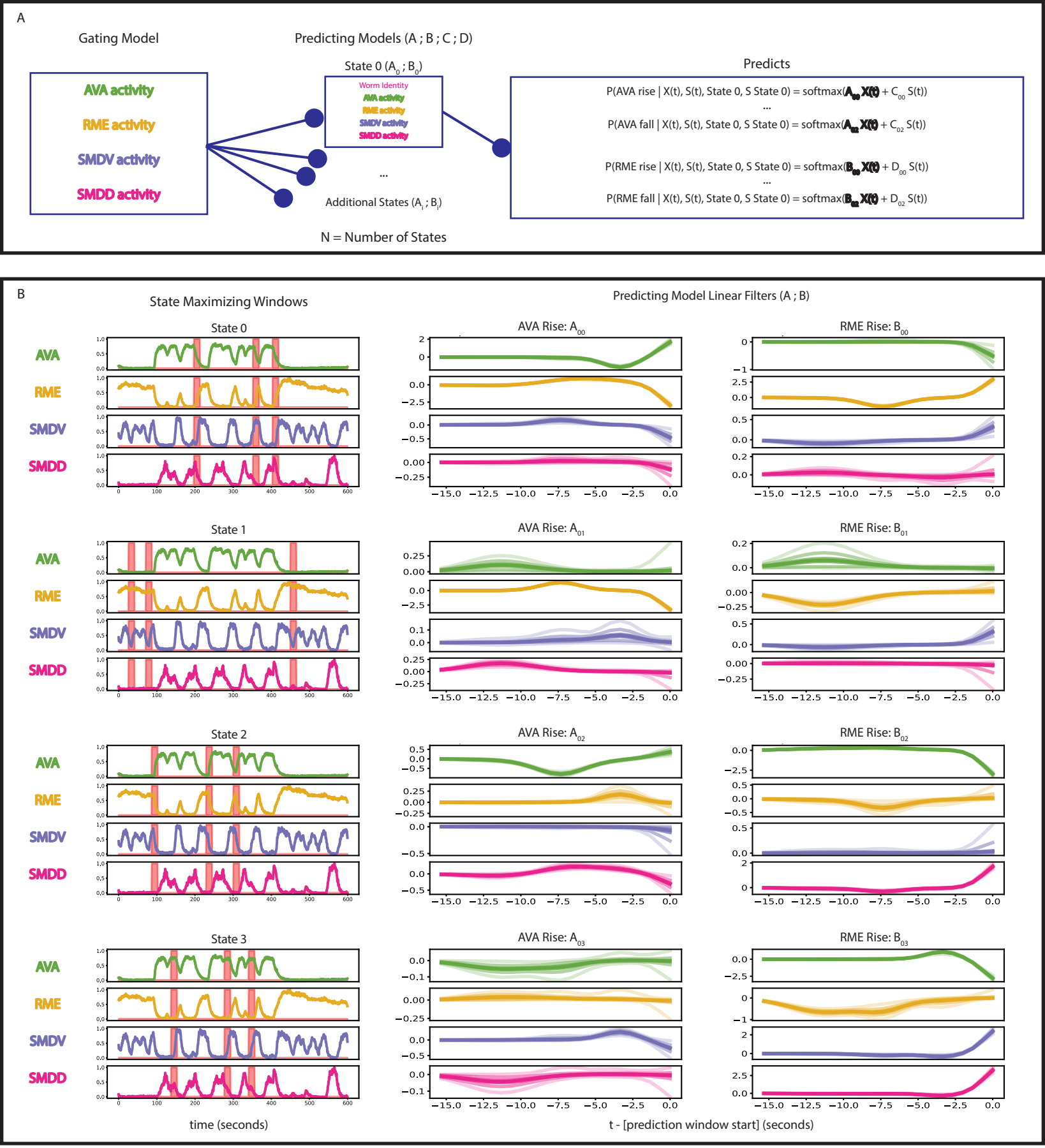

Figure 5. The top performing tree-MLR (multinomial logistic regression) hybrid model on the NeuroPAL data is visualized. The top performing model for this data has four network history states and four stimulus states. In this figure, only the network history filters are plotted. (A) Model schematic for top performing model. (B) Visualization of each state's MLR. The left set of figures identifies the windows for which a given state's probability is maximized. For instance, state 0 is maximally active when AVA and SMDD are falling and RME is rising. The middle column is made up of the linear filters predicting AVA rise. In general, AVA increase and RME decrease predict AVA rise. The right column is made of the linear filters predicting RME rise. RME fall and SMDV/D increase tend to predict RME rise. Each filter plot has five lines. The thickest line is the median filter across bootstraps. The next thickest lines contain the middle 50% of bootstraps at each time offset. The thinnest lines contain 90% of bootstraps.

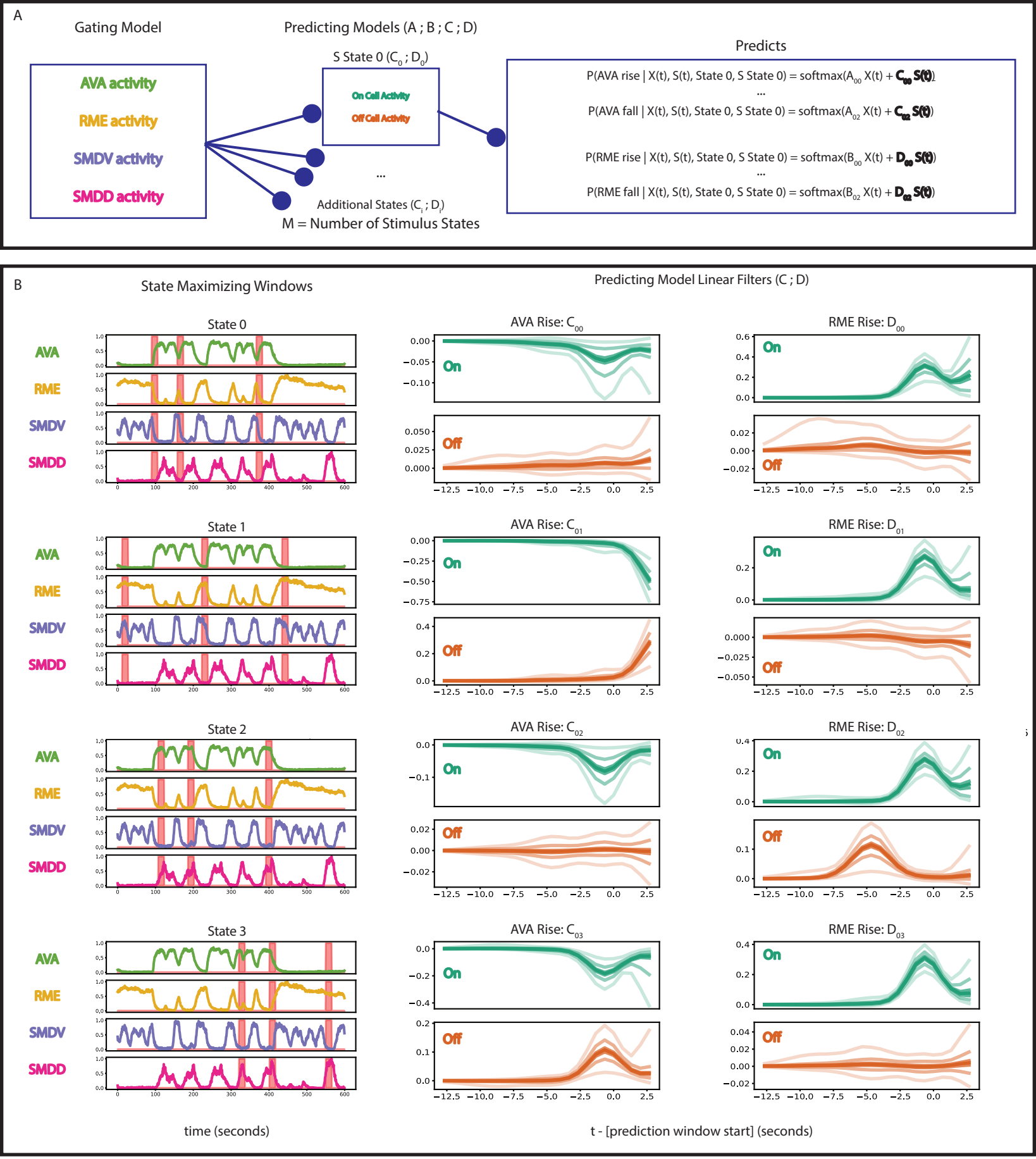

Figure 6. The top performing tree-MLR (multinomial logistic regression) hybrid model on the NeuroPAL data is visualized. The top performing model for this data has four network history states and four stimulus states. In this figure, only the On and Off cell filters are visualized. (B) Visualization of each state's MLR. The left set of figures identifies the windows for which a given state's probability is maximized. For instance, state 0 is maximally active when AVA is rising, RME is falling, and SMDD is low. The middle column is made up of the linear filters predicting AVA rise. The right column is made up of the linear filters predicting RME rise. In general, On cell activity dives RME rise and suppresses AVA rise. Off cell activity tends to drive AVA rise. Each filter plot has five lines. The thickest line is the median filter across bootstraps. The next thickest lines contain the middle 50% of bootstraps at each time offset. The thinnest lines contain 90% of bootstraps.