Published in final edited form as:
Neuron. 2013 September 4; 79(5): 987–1000. doi:10.1016/j.neuron.2013.06.041.

A modeling framework for deriving the structural and functional architecture of a short-term memory microcircuit

Dimitry Fisher[1], **Itsaso Olasagasti**[2], **David W. Tank**[3], **Emre R.F. Aksay**[4,*], and **Mark S. Goldman**[1,5,*]

[1]Center for Neuroscience, University of California, Davis, CA 95618, USA [2]Department of Neurology, Zurich University Hospital, Zurich, Switzerland [3]Princeton Neuroscience Institute and Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA [4]Department of Physiology and Biophysics, Weill Medical College of Cornell University, New York, New York 10021, USA [5]Departments of Neurobiology, Physiology, and Behavior; and Ophthalmology and Visual Sciences, University of California, Davis, CA 95618, USA

Summary

Although many studies have identified neural correlates of memory, relatively little is known about the circuit properties connecting single-neuron physiology to behavior. Here we developed a modeling framework to bridge this gap and identify circuit interactions capable of maintaining short-term memory. Unlike typical studies that construct a phenomenological model and test whether it reproduces select aspects of neuronal data, we directly fit the synaptic connectivity of an oculomotor memory circuit to a broad range of anatomical, electrophysiological, and behavioral data. Simultaneous fits to all data, combined with sensitivity analyses, revealed complementary roles of synaptic and neuronal-recruitment thresholds in providing the nonlinear interactions required to generate the observed circuit behavior. This work introduces a new methodology for identifying the cellular and synaptic mechanisms underlying short-term memory, and demonstrates how the anatomical structure of a circuit may belie its functional organization.

Introduction

Understanding the mechanisms underlying complex behaviors requires bridging the gap between cellular properties and circuit-level interactions that drive system function. This problem is particularly acute in short-term memory systems, where the identified kinetics of synaptic and intrinsic cellular processes operate on a much shorter time scale (typically 1–100's of ms) than the observed behavior. A neural correlate of short-term memory over the seconds to tens of seconds time scale has been identified in the persistent firing of neuronal populations during memory periods following the offset of a stimulus. Such activity has been recorded across a wide range of brain regions and tasks, and been shown to maintain representations of both discrete and graded stimuli (for review, see Brody et al., 2003; Durstewitz et al., 2000; Major and Tank, 2004; Wang, 2001).

*Correspondence: ema2004@med.cornell.edu or msgoldman@ucdavis.edu.

Many explanations have been proposed for how persistent neural activity is generated. Various studies have hypothesized roles for intrinsic neuronal properties (Egorov et al., 2002; Fall and Rinzel, 2006; Koulakov et al., 2002; Lisman et al., 1998); synaptic mechanisms (Mongillo et al., 2008; Shen, 1989; Wang et al., 2006); or specialized anatomical architectures (for review, see Brody et al., 2003; Goldman, 2009; Wang, 2001). More likely, however, the generation of memory-storing neural activity reflects a combination of cellular, synaptic, and network properties (Major and Tank, 2004). Thus, fully understanding the mechanisms underlying memory-guided behaviors will require methods that combine data from experiments probing neural circuits at each of these levels in order to relate neuronal responses to behavior.

Computational modeling has been used to bridge the gap between cellular physiology, circuit interactions and memory function. However, modeling the responses of neurons in recurrent circuits is highly challenging because each neuron's activity influences, and is influenced by, potentially every other neuron in the circuit. Furthermore, due to features such as activation thresholds and saturation, both the neuronal and synaptic (or dendritic) responses may be highly nonlinear. Thus, for a circuit consisting of N neurons, there may be of order N^2 nonlinear synaptic interactions.

This modeling challenge has traditionally been tackled by two highly disparate approaches. Conceptual models use strong simplifying assumptions on the forms of synaptic connectivity and neuronal responses to provide tractability in modeling complex neural circuits (Figure 1). While such studies provide qualitative insight, the chosen assumptions limit the set of possible mechanisms explored and make close comparison to experiment difficult. Alternatively, to make close contact with experiment, other studies have used brute-force explorations of the large parameter space defined by multiple intrinsic and synaptic variables (Goldman et al., 2001; Prinz, 2007; Prinz et al., 2004). These studies have successfully demonstrated how circuit function can be highly sensitive to changes in certain combinations of parameters, but insensitive to changes in others. However, the combinatoric explosion of parameter combinations has limited such studies to exploration of only ~10 parameters or less at a time, a minute fraction of the total parameter space needed to fully describe a circuit.

Here we describe a modeling framework in which a wide range of experimental data from cellular, network, and behavioral investigations are directly incorporated into a single coherent model, while predictions for difficult-to-measure quantities like synaptic connection strengths and synaptic nonlinearities are generated by directly fitting the model to these data. This approach is applied to data from a well-characterized circuit exhibiting persistent neural activity, the oculomotor neural integrator of the eye movement system (Robinson, 1989). This circuit receives transient inputs that encode the desired velocity of the eyes, and stores the running total of these inputs – the desired eye position – as a pattern of persistent neuronal firing across a population of cells. Such maintenance of a running total represents the defining feature of temporal integrators or accumulators, which are widely found in neural systems (Gold and Shadlen, 2007; Goldman et al., 2009; Major and Tank, 2004). Previous studies of the goldfish oculomotor integrator have gathered data at each of the levels of analysis typical of studies of memory systems: intrinsic cellular properties (Aksay et al., 2001), anatomy (Aksay et al., 2000), behavior (Aksay et al., 2000), and functional circuit interactions (Aksay et al., 2003; Aksay et al., 2007). Thus, this system provides an ideal setting in which to illustrate how data at each of these levels can be coherently combined to gain a fuller understanding of memory-guided behavior.

The results described below comprise the following principal findings. First, we construct a spiking network model that reproduces the tuning curve response of every neuron in an

experimentally recorded database, while simultaneously fitting the single-neuron intrinsic response properties, circuit anatomy, and network response to inactivations. Second, sensitivity analyses rigorously identify the features of synaptic connections most critical to persistent neural firing. Third, the functional connectivities of the well-fit models are shown to differ markedly from their anatomical connectivities. Fourth, concrete experimental predictions are generated to differentiate between models based upon different forms of threshold mechanisms predicted to be present in the oculomotor integrator circuit.

Results

In the following, we describe a framework for constructing models of memory-storing circuits by simultaneously fitting experiments conducted in the oculomotor neural integrator that probed intrinsic cellular response properties, interactions between integrator neurons, and neuronal responses during behavior. Three features must be ascertained to fully describe circuit function: the spike-generating process of the individual neurons, the connectivity between the neurons, and the functional response properties of the synaptic connections and dendrites onto which they project. Accordingly, our overarching strategy to determine these features is to: first, construct a model of the spike-generating process of individual neurons by fitting the responses of oculomotor integrator neurons to current injections that slowly drive neuronal firing across the full range of observed firing rates; second, incorporate these model neurons into an anatomically-constrained circuit model and fit the circuit connection strengths and synapto-dendritic response properties to neuronal recordings obtained during normal behavior and pharmacological inactivation; and third, perform sensitivity analyses to reveal which features of the best-fit connectivity are essential to circuit function.

Data incorporated into the model

Below we summarize the key experiments used to fit the computational model. The goldfish oculomotor neural integrator is a bilateral circuit located in the caudal hindbrain. As animals make a sequence of fixations from left to right, neurons on the right side of the midline become activated above their respective firing thresholds and maintain persistent firing rates that linearly increase with the eyes' position (Figure 2A). Conversely, neurons on the left side exhibit a linear decrease in persistent firing with more rightward fixations. Neuronal tuning curves are therefore well-characterized by two parameters (Aksay et al., 2000): the threshold eye position at which they become active E_{th}, and the rate of increase of firing rate with increasing eye position k. Equivalently, one of these parameters can be replaced by the y-intercept r_0, or "primary rate", which gives the firing rate when the eyes are at the central eye position $E = 0$ degrees. This push-pull arrangement, in which neurons on the left increase in firing rate when those on the right decrease in firing rate (Figure 2A, bottom), is thought to reflect an anatomical arrangement in which excitatory neurons on each side of the midline project ipsilaterally, whereas inhibitory neurons project contralaterally (Figure 2B; see Aksay et al., 2000; Aksay et al., 2003; Aksay et al., 2001). No apparent association has been found between tuning curve shapes and whether a neuron is excitatory and inhibitory (Aksay et al., 2003).

Further constraints on the properties of synaptic connections were obtained by pharmacologically inactivating a portion of the integrator circuit while spared neurons were recorded (Aksay et al., 2007). These experiments produced a distinct pattern of deficits in persistent neuronal activity (Figure 2C). When the spared neuron was located contralateral to the side of the inactivation (Figure 2C, blue trace), it exhibited upward firing rate drift when its firing rate was close to or less than the primary rate r_0. For higher rates, stable persistent firing was maintained. A complementary pattern of firing rate drift was exhibited for neurons located ipsilateral to the inactivation: downward firing rate drift when rates were close to or higher than r_0, and stable persistent firing at lower rates (red trace).

Finally, the spike-generating properties of model cells were based upon responses of oculomotor integrator neurons to current injection *in vivo* during fixation. Somatic injections of slowly ramping up and down currents lead to a nearly linear firing rate response, with a narrow region of higher slope at the onset of firing (Figure 2D). Negligible hysteresis was found between the responses to increasing versus decreasing currents.

Fitting a single-cell model to intracellular current injection data

We first constructed a spiking single-neuron model that reproduced the spike-generating properties illustrated in Figure 2D. The model included leak, Na^+, delayed rectifier K^+, and transient K^+ conductances. These conductances, which corresponded to a Hodgkin-Huxley spiking mechanism plus a weak adaptation current mediated by the transient K^+ conductance, comprised a minimal set for describing the observed single-neuron dynamics in the sense that notably worse fits occurred if any individual conductance was removed. Maximal conductance parameters, as well as the total membrane capacitance and time constant of inactivation of the transient K^+ conductance, were fit using a cost function that minimized the difference between the current injection data of Figure 2D and the model responses to the same pattern of current injection. The fit was performed by re-plotting the spiking responses as a cumulative sum over time to produce a nearly smooth curve (Figure 3B, blue line) that enabled the model parameters to be fit using a standard nonlinear optimization routine (Experimental Procedures). The resulting single-compartment model matched the experimental response to the slowly varying current injection on a nearly perfect spike-by-spike basis (Figure 3A) and also had a membrane time constant consistent with integrator neuron recordings (Supplemental Methods).

Fitting a circuit model to anatomical, cellular-behavioral, and inactivation studies

We next sought to determine both the functional form of the synaptic interactions between integrator neurons, and the patterns of connections throughout the integrator memory network. The primary challenge in constructing recurrent network models of graded persistent activity is to tune the synaptic inputs so that the circuit can maintain persistent firing across a continuous range of firing rates. If the net synaptic current provided to a neuron is too weak, neuronal firing during memory periods will drift downward due to the intrinsic leakiness of the neuronal membrane. If the net input current is too strong, neuronal firing will drift upwards. In the context of the oculomotor integrator, this tuning requirement implies that, at each stably maintained eye position, there is a precise level of current required to sustain each neuron's firing rate at its experimentally observed value. We therefore asked what possible sets of connection strengths and synaptic nonlinearities could enable the circuit to simultaneously reproduce all of the experiments illustrated in Figure 2.

The details of the model-fitting procedure are given in the Experimental Procedures and Supplemental Methods. In brief, the model contained a total of 100 neurons, the estimated number in the goldfish integrator circuit, divided into excitatory and inhibitory populations on each side of the midline as suggested by experiment (Figure 2A,B). Synaptic inputs were modeled as a sum of recurrent excitatory, recurrent inhibitory, and tonic background currents (Figure 3F). Each recurrent synaptic input was modeled as the product of a "synaptic strength" parameter W_{ij} representing the maximal possible somatic current provided from neuron j to neuron i, and a "synaptic" (and/or dendritic) activation $s(r_j)$ representing the fraction of this maximal current provided when presynaptic neuron j fires at rate r_j (Figure 3E,F).

The best-fit connection strengths onto any given neuron were found by minimizing a cost function (Equation (4) of Experimental Procedures) whose individual terms enforced that each neuron maintain persistent firing at its experimentally observed firing rate $r(E)$ for

every stable eye position (Figure 3D). This was done by penalizing, for each neuron, any differences between the current required to generate the experimentally observed firing rate at each eye position (Figures 3F, S1F, dashed black line, obtained from combining the single-neuron response curve, Figure 3C, with the neuron's tuning curve, Figure 3D) and the summed excitatory (red), inhibitory (blue), and tonic background current (orange) for a given set of synaptic weights W_{ij} and synaptic nonlinearities $s(r_j)$. For control animals, the circuit was required to maintain persistent activity at all eye positions (Figure 3F). For the inactivation studies, we only enforced that persistent activity be maintained at those firing rates experimentally observed to remain stable (Figure 2C). For example, we only enforced that firing rates above a value close to r_0 be stably maintained following the removal of recurrent inhibition through total unilateral inactivation (Figure 3G, colored portions). Finally, a regularization term (Hastie et al., 2009) was added to the cost function to penalize exceptionally large connection strengths that lead to synaptic response magnitudes inconsistent with intracellular measurements (Aksay et al., 2001).

This procedure succeeded in generating circuits that simultaneously reproduced all of the experimental data of Figure 2 (Figures 4,5). The circuits temporally integrated arbitrary patterns of saccadic inputs (Figure 4E,F, left, two example circuits) and precisely reproduced the tuning curves of every experimentally recorded neuron in our database (Figure 4E,F, right, four example neurons). Furthermore, inactivations of these well-fit circuits reproduced the characteristic pattern of drifts following both contralateral and ipsilateral inactivations (Figure 5). Thus, the model recapitulated both the gross and neuron-specific properties of an entire vertebrate neuronal circuit.

Sensitivity of circuit fits to synaptic activation parameters

Given that the complete circuit connectivity is defined by 5100 synaptic weight parameters, as well as the unknown form of the synaptic activations $s(r)$, we expected that many different parameter value combinations could provide optimal or near-optimal fits to the available experimental data. To explore this large parameter space, we implemented a formal two-stage sensitivity analysis, first characterizing the dependence of the model fits on the form of synaptic activations, and then, for a given form of synaptic activations, the dependence on the pattern of connection strengths.

The sensitivity of the model fits to the form of excitatory and inhibitory synaptic activation was explored by systematically varying the two parameters describing the activation function: θ, which controlled the width, and R_f, which controlled the point of inflection (Figure 4A). This allowed us to consider models in which the transformation at synapses was linear, saturating (e.g. due to synaptic depression or saturation of driving forces), or sigmoidal (e.g. due to synaptic facilitation or voltage-activated dendritic currents). Excitatory and inhibitory recurrent synapses were allowed to have different forms of nonlinearity.

This analysis showed that the integrator network can utilize only a restricted set of synaptic activation functions to generate persistent firing. Figure 4B shows the space of synaptic activations permitted (blue) and prohibited (red) by the experimental constraints when inhibitory and excitatory synapses have identical (left) or different (middle, right) forms. Circuits using predominantly saturating activations (Figure 4A, lower left corner) failed to meet the conditions imposed by the experimental constraints. The poor performance of circuits with saturating synapses was true for strongly saturating excitation or inhibition (Figure 4B, middle, L-shaped poorly fitting region), and even for mildly saturating excitation alone (right panel, bottom region). The mechanistic reason for this poor performance is that neurons with saturating synapses transmit a large fraction of their maximal currents when they fire at low rates, so that silencing such neurons greatly disrupts

the balance of currents required to maintain stable persistent activity even when these neurons fire at low rates. This violates the constraint imposed by the inactivation experiments, which found that stable persistent firing was maintained at times when the inactivated population would have been firing at low rates (Figure 2C).

In contrast, we found that circuits utilizing sigmoidal (Figure 4B, point 1; Figure 4C) or more linear (Figure 4B, point 2; Figure 4D) synaptic activations were able to match all experimental constraints. Neurons in well-fit models received little or no current from cells firing at rates much lower than their primary firing rates r_0 (Figure S1), thus satisfying the constraints imposed by the inactivation experiments. In models with strongly sigmoidal activation functions, characterized by a large inflection point R_f and narrow width θ so that the synaptic response was strongly superlinear at low presynaptic firing rates (Figure 4A, 4C left, large R_f and low θ values), low firing rates drove little synaptic current into the postsynaptic cell due to the (soft) threshold occurring at the synapse. We refer to this as a *synaptic threshold mechanism* (Figures 4C,E, S5A). Models with more linear synaptic activations instead depended critically on input from high eye-position threshold neurons (Figures 4D,F, S5B). In these circuits, the constraint imposed by the inactivation experiments is met because the high eye-position threshold neurons transmit a large portion of the total current received by each neuron; thus, there is very little current transmitted over the portion of the oculomotor range (negative eye positions in Figures 2A and S1F) observed to be minimally affected by unilateral inactivations. While for excitation some input from low recruitment-threshold neurons was tolerated, for inhibition connection weights from such neurons had to be nearly zero (Figure 4D, right; Figure S2). We refer to this as a *neuronal recruitment threshold mechanism*.

More generally, we found that well-fit models could utilize combinations of the above two mechanisms. As seen in Figure 4B, the well-fit (blue) models occupied a connected region of parameter space within which the circuits illustrated in Figures 4C and 4D represented relatively pure examples of each mechanism. Traversing the region between these examples by progressively decreasing the synaptic threshold, we found a compensatory increase in the reliance on higher recruitment threshold neurons and decrease in reliance upon lower recruitment threshold neurons, especially for inhibition (Figure 4G; Figure S5). This path through parameter space represents an insensitive direction of movement along the model cost-function surface, with a tradeoff between the use of synaptic and recruitment thresholds.

Sensitivity of circuit fits to synaptic weight parameters

We next asked which features of the circuit connectivity were necessary and which could be changed with minimal degradation of model performance. To address this question, we performed a sensitivity analysis on the connections weights for circuits based on both the synaptic threshold and neuronal recruitment-threshold mechanisms.

For a given form of the synaptic activation function, we first determined the best-fit connectivity pattern from the minimum of our fit cost function. We then asked how the cost function changed when individual synaptic connections were altered from their best-fit values, and which concerted *patterns* of synaptic connection changes caused the greatest changes in the fit performance. These quantities were found by calculating, for each neuron, how rapidly the cost function curved away from its minimum value when the presynaptic weights onto the neuron were varied around their best-fit values. Mathematically, this curvature is defined by the sensitivity (or Hessian) matrix $H_{ij}^{(k)}$ whose $(i, j)^{th}$ element contains the second derivative of the cost function with respect to changes in the weights of the i^{th} and j^{th} presynaptic inputs onto neuron k (Figure 6A). Sensitivity to changes in a

single presynaptic input weight are given by the diagonal elements of the matrix. Sensitivity to concerted patterns of weight changes are found from the eigenvector decomposition of the matrix. Eigenvectors corresponding to the largest eigenvalues give the patterns of weight changes along which the cost function curves most sharply, and hence identify the most sensitive directions of the circuit to perturbations. Eigenvectors corresponding to small eigenvalues define patterns of weight changes to which the cost function is insensitive.

Figure 6A shows the sensitivity matrix for a neuron from the synaptic threshold model of Figure 4C. The sensitivity matrix separates into diagonal blocks, indicating that changes in the cost function due to perturbations in excitatory (inputs 1–25) and inhibitory (26–50) weights were nearly independent of one another. Within these blocks, the precise grid-like pattern of sensitivities was dependent upon the exact choice of tuning curves used in any given simulation and was removed by averaging the sensitivity matrices of 100 circuit simulations with different random draws of tuning curves (Figure 6B).

From the diagonal elements of these mean sensitivity matrices, we computed the tolerance (Figure 6D, red bars) of each model to changes in any individual synaptic weight away from its best-fit value (green). This type of analysis corresponds to a traditional “vary one parameter at a time” sensitivity analysis, and is useful in predicting the effect of perturbing a single or small set of connection weights. For the synaptic-threshold mechanism circuits (Figure 6D, left), only the connections from the low-threshold inhibitory neurons were sensitively different from zero. By contrast, for the neuronal recruitment-threshold mechanism circuits (Figure 6D, right), only connections from high-threshold inhibitory neurons were sensitively different from zero. These results suggest that experimental manipulations that remove individual high- or low-threshold inhibitory neurons will have opposite effects in circuits based upon the different threshold mechanisms (see *Model predictions*).

The above analysis describes the effect of varying single weights onto a neuron. However, it does not address the question of whether a particular weight onto a neuron must be held close to its best-fit value. This is because studying the effects of changing one weight at a time does not consider whether such changes could be offset by compensatory changes in weights arriving from other, correlated inputs.

To address this latter question, we calculated the eigenvectors of the sensitivity matrix to determine which concerted *patterns* of connection weights most sensitively affect the tuning of the circuit. Figure 6E–G shows the leading eigenvectors for a neuron from the synaptic threshold mechanism circuit of Figure 4C. The most sensitive perturbation corresponds to making all weights more excitatory (Figure 6E, eigenvector 1) or, equivalently, making all weights more inhibitory since eigenvectors are only defined up to a sign change. Changes along this direction lead to a unidirectional “bias” in the inputs to this neuron that also was observed in the first eigenvectors of the other neurons in this circuit (Figure 6F, Figure S6F). As a result, perturbing the first eigenvectors of all neurons lead to dramatic unidirectional drift in neuronal firing (Figure 6G).

The second through fourth columns of Figure 6 show the next most sensitive patterns of connection weights for this circuit. The second eigenvector defined a “leak-instability axis” defined by increasing or decreasing the magnitude of all excitatory or inhibitory inputs. Perturbing this pattern of weights changed the amplitude of both the excitatory and disinhibitory feedback loops in the network, leading to strong exponential decay (leak) or instability of firing rates around a single fixed value (Figure 6G, 2nd column). The third and fourth eigenvectors corresponded to making inputs from low-threshold neurons more excitatory or less inhibitory while making those from high-threshold neurons more

inhibitory or less excitatory and therefore defined “high- vs. low-threshold neuron” axes. Changes along eigenvectors 3 and 4 lead to drift towards or away from two or three fixed points, respectively (Figure 6G, 3rd and 4th columns).

For circuits based on neuronal recruitment-thresholds, the eigenvector analysis revealed a similar pattern of most sensitive directions (Figure S6B,C). Thus, even though circuits based on the different threshold mechanisms had very different best-fit connectivities, the patterns of perturbations to which they were most sensitive was highly similar. This reflected that, in all circuits, the prime determinant of circuit architecture was providing the appropriate balance of currents to maintain persistence across the entire firing rate range, and this balance was determined by three dominant factors: first, providing the correct average level of input to each neuron; second, providing the correct balance of inhibition and excitation; and third, providing an appropriate balance of input from high-and low-threshold neurons.

For all circuits, the least sensitive directions corresponded to oppositely directed changes in just a few inputs (Figure 6E, eigenvector 50) or offsetting and often noisy-appearing changes in larger groups of inputs (Figure 6E–G, eigenvectors 10 and 30). Changes along these insensitive directions could yield other circuits that looked quite different in their connectivity structure but had nearly identical model performance. For example, changing the neuronal recruitment-threshold circuit of Figure 4D along an insensitive direction showed that excitation for this set of synaptic activations could either use (Figure 4D) or not use (Figure S7A,B) low-threshold excitatory neurons. This insensitive change would not be identified by the traditional individual connection-weight analysis of Figure 6D (right), which shows that the circuit performance is sensitive to changing individual low-threshold excitatory weights. This might seem to contradict the observation that input from low-threshold excitatory neurons is not required. However, the individual connection weight analysis only identifies the effects of changing individual weights *when no other compensations are made in other weights*. The eigenvector analysis resolves this seeming discrepancy by showing that low-threshold excitatory connections are not *necessary* because they can be compensated for by making *offsetting* changes in broader patterns of weights.

Model predictions differentiating synaptic and neuronal-recruitment threshold mechanisms

Circuits based on the synaptic and neuronal-recruitment threshold mechanisms make distinctly different predictions for targeted neuronal ablation experiments. As was seen in the sensitivity analyses (Figure 6D), circuits based upon these mechanisms rely in opposite manners upon high-versus low-threshold inhibitory neurons. Circuits using the synaptic threshold mechanism are most sensitive to removal of low eye-position-threshold inhibitory neurons (Figure 6D left). This predicts that ablating low-threshold inhibitory neurons in such circuits would have a much larger effect on the drift patterns than ablating high-threshold inhibitory neurons, and this was seen in our simulations (compare Figures 7A, middle and bottom). By contrast, circuits based upon neuronal thresholds are insensitive to loss of inputs from low-threshold inhibitory neurons but highly sensitive to loss of high-threshold inhibitory neurons. Thus, ablating high-threshold inhibitory neurons in such circuits would have a much larger effect on the drift patterns than ablating low-threshold inhibitory neurons (Figure 7B). For detailed analysis of the specific patterns of drift seen in Figure 7, we refer the reader to the simplified analytic model of the Supplemental Methods and Figure S2.

A second prediction arises from analyzing the time constants of drift following inactivation. Both in the well and poorly fit models, the rate of drift following inactivation scaled approximately linearly with the inverse of the recurrent excitatory synaptic time constant. To reproduce quantitatively the drift rates observed experimentally following inactivation, a recurrent excitatory synaptic time constant of approximately 1 second was required. This

finding predicts a role for a slow cellular component of persistence at excitatory synapses or dendrites (see Discussion).

Disparity between structural and functional connectivity

The results above show that there are multiple circuit structures, understandable by the tradeoff between two thresholding mechanisms, that could reproduce the experimental data. As shown next, however, these structural differences masked strong similarities in functional connectivity that were revealed only when the combined effects of the structural connectivity W_{ij}, the synaptic nonlinearities $s(r_j)$, and the threshold nonlinearity of the tuning curves were considered.

To generate the functional connectivity (also known as “effective connectivity”; Sporns et al., 2004) between neurons at different eye positions, we calculated the amount of current provided by any given neuron to its postsynaptic targets at different eye positions. These currents then were normalized by the presynaptic firing rate to obtain a functional connectivity measure, current per presynaptic spike, that did not simply reflect the strength of presynaptic firing. Below-threshold neurons were assigned a functional connectivity strength of zero.

The resulting functional connectivities for all circuits exhibited a striking pattern not evident in the anatomical structure: when the eyes were directed leftward, the left-side inhibitory neurons projected strong functional connections. However, the functional weights of inhibitory right-side neurons were almost zero (Figure 8D–F). When the eyes were directed rightward, the opposite pattern emerged, with the right side inhibitory neurons dominating and those on the left side contributing little (Figure 8G–I). This result illustrates a paradox: although inhibitory connections appear from the anatomy to create a disinhibitory feedback loop between the two sides of the integrator, the functional connectivity demonstrates that inhibition acts in a feedforward manner.

Discussion

Persistent neural activity has been identified in a wide range of memory circuits, but previous models of such activity have been primarily conceptual in nature and not easy to compare directly to experimental recordings of individual neurons. Here, we developed a regression-based fitting routine that directly incorporates anatomical constraints on connectivity, intracellular current injection recordings, neuronal tuning curves recorded during behavior, and neuronal drift patterns following pharmacological inactivation. This approach enables biophysically detailed predictions to be made regarding both the properties of synaptic signal transformation and the patterns of connectivity between constitutive neurons. Furthermore, sensitivity analyses enabled us to make strong statements about which features of the model were, and were not, essential. Our analysis revealed two circuit mechanisms, one based on synaptic thresholds and one on neuronal recruitment thresholds, that were required of all well-fit networks. Despite very different anatomical connectivity, the functional connectivity of circuits utilizing these two mechanisms was similar, revealing a striking dichotomy that is likely to be present in many other circuits and discoverable utilizing the modeling framework developed here.

Relationship of approach to previous models

The model presented here provides, to our knowledge, the first example of a memory network in which such a wide range of experimental data are directly incorporated, while difficult-to-measure quantities like network connection strengths and synaptic nonlinearities are simultaneously fit to these data. We further have been able to identify sensitive and

insensitive combinations of synaptic parameters that change or leave unaffected circuit performance, respectively. Previous circuit studies utilizing a purely brute force approach have also performed sensitivity analyses (Prinz, 2007), but have been limited to the study of small networks and small numbers of parameters due to the explosion of possible parameter combinations. We instead used a brute force approach to study sensitivity to the small number of synaptic activation parameters, but implemented an eigenvector-based approach for analyzing the large number of synaptic connections. This procedure revealed a relatively small number of patterns of connection weights onto each neuron that must be sensitively maintained to have good model performance.

More generally, our use of a cost function to enforce different biological constraints permits the incorporation of results from additional experiments. For example, topographic organization consistent with recent optical recordings in the larval zebrafish integrator (Miri et al., 2011) could be incorporated by adding a term to the cost function that penalizes long-distance connections. In Figures S7C and S7D we show that such topographic connectivity is consistent with the experimental findings modeled here, and can be obtained through insensitive changes from the best-fit circuits. Such a pathway through the parameter space of network connectivity could be utilized during development, with the integrator network beginning in a more topographically organized form and moving to a more distributed connectivity pattern in the mature state, where the functional signatures of topography seem to be weaker (as discussed in Miri et al., 2011).

In addition, our approach can be extended to allow greater heterogeneity in synaptic parameters or to model circuits with non-monotonic tuning curves (D. Fisher, unpublished observations). We have considered a single shape of synaptic activation function for all excitatory neurons, and a separate single shape for all inhibitory neurons, regardless of threshold. Relaxing this constraint might identify circuit architectures in which there are gradients in synaptic activation parameters as a function of neuronal threshold.

Model predictions and implications for biophysical mechanisms of persistent activity

Our work makes several predictions about the mechanisms of integration in the oculomotor integrator and possibly other short-term memory circuits. First, in contrast to the previous spiking model of the oculomotor integrator based upon purely saturating synapses (Seung et al., 2000, Figure S4D), which modeled a single unilateral population and was generated before the inactivation experiments had been performed, our sensitivity analysis suggests that both inhibition and excitation are likely to be mediated by approximately linear or sigmoidal synaptic activation functions. Second, our quantitative fits to the drift rates following inactivation suggest that the observed long integration time constants are not solely due to network mechanisms, and instead suggest the presence of an intrinsic cellular or synaptic process with a time constant of order 1 second. Third, we suggest that integration depends critically upon the presence of a threshold mechanism. This could either take the form of a synaptic (or dendritic) threshold, as suggested by Aksay et al. (2007), or result from the circuit's recurrent connectivity depending critically upon neurons with high eye-position thresholds, particularly for inhibition.

Potential "synaptic" mechanisms consistent with a sigmoidal dependence upon presynaptic firing rate and an approximately 1 second time constant are presynaptic facilitation (Wang et al., 2006) or, postsynaptically, localized dendritic plateau potentials (Major et al., 2008; Wei et al., 2001). The long time constants associated with these mechanisms could provide robustness against disruptions of circuit connectivity (Camperi and Wang, 1998; Goldman et al., 2003; Koulakov et al., 2002; Mongillo et al., 2008). The high thresholds could be useful in filtering out low firing rates (Chichilnisky and Rieke, 2005), which are noisier in the oculomotor integrator than higher firing rates (Aksay et al., 2003).

Having recurrent inputs dominated by high eye-position-threshold neurons would suggest a functional difference between low- and high-threshold neurons. High-threshold neurons would be used for maintenance of persistent firing rates within the integrator, whereas low-threshold neurons might be used as readout neurons. Experimental tests of this threshold organization should be possible through targeted silencing of specific subsets of neurons, for example using halorhodopsin in the optically transparent larval zebrafish preparation (Schoonheim et al., 2010).

Functional versus structural connectivity, and relation to connectomics

One of the most striking features of these models is the difference between the *functional* and *structural* connectivities (Figure 8). As shown in Figure 2, the two sides of the circuit are connected by mutual inhibition, anatomically suggesting the presence of a “double negative” (disinhibitory) positive feedback loop. In most models with inhibition between two populations, such positive feedback loops generate persistent activity (Cannon et al., 1983; Machens et al., 2005; Sklavos and Moschovakis, 2002; Song and Wang, 2005). By contrast, our results suggest that the anatomical mutual inhibitory loop is functionally broken so that there is no disinhibitory feedback loop to sustain persistent activity. Rather, as suggested previously (Aksay et al., 2007; Debowy and Baker, 2011), recurrent excitation generates persistent activity at high firing rates, and low firing rates are held stable primarily by feedforward inhibition that is driven by the stable high rates of the opposing side.

The dichotomy between functional and anatomical connectivity demonstrated here suggests how a deeper understanding of the link between cellular properties and behavior can be facilitated by combining modeling work with large-scale anatomical studies. Serial-section electron microscopy (Briggman and Denk, 2006; Micheva and Smith, 2007) and automated image processing (Chklovskii et al., 2010; Jain et al., 2010) promise unprecedented opportunities for defining the anatomical connectivity of a circuit. However, much in the way the human genome project was successful in identifying genes but not directly informative of their functional roles, connectomics will provide only an identification of anatomical connections. An understanding of the functional connectome therefore will rely on a hybrid approach where data on neuronal responses is combined with high-resolution structural information. Importantly, we note that not all structural information is equally informative, as we showed that integrator function was highly dependent on the proper balance of interactions between high and low-threshold neurons, but insensitive to random changes in the connections between cells with similar thresholds. Thus, biophysically realistic circuit models can help guide anatomists in determining which aspects of the connectivity are most important to measure. In turn, the relevant structural data will be invaluable to refining model predictions, providing additional fitting constraints that help to further limit the space of possible functional connections.

Experimental Procedures

Data Analysis

Experimental data were previously obtained in the horizontal velocity-to-position neural integrator of the awake, behaving adult goldfish (Aksay et al., 2000; Aksay et al., 2003; Aksay et al., 2001; Aksay et al., 2007). Briefly, neuronal tuning curves were determined from extracellular recordings of integrator neuron activity. They were well-approximated by a threshold-linear relationship between firing rate r_i and eye position E during stable fixations,

$$r_i = \max\{k_i(E - E_{th,i}), 0\} = \max\{(k_i E + r_{0,i}), 0\}, \quad (1)$$

described for a given cell i by a sensitivity k_i and either eye-position threshold $E_{th,i}$ or intercept $r_{0,i}$ (Figure 2A). Neuronal excitability was determined from intracellular recordings of the response to current injection (Figure 2D). Circuit interactions were assessed by extracellular recording of single-unit activity immediately following localized pharmacological silencing of neighboring cells using lidocaine or muscimol.

Neuronal drift patterns characterizing the effects of pharmacological inactivation were obtained by comparing firing rate drifts before and after inactivation (Supplemental Methods). Drift was plotted as a function of firing rate rather than eye position in order to eliminate potential confounds that could occur if the inactivations affected the eye position readout from the circuit by altering the relationship between firing rates and eye position. To pool across cells recorded in different preparations, neuronal activity was normalized using the eye-position sensitivities and intercepts given by the steady-state (control) tuning curve relationships (equation (1)). Firing rates for cell i were normalized by first subtracting its primary rate $r_{0,i}$ and then dividing by its position sensitivity k_i, resulting in normalized rates in units of eye position. Firing rate drifts were normalized by the position sensitivity k_i. An identical analysis was performed on the model firing rate data, permitting a direct comparison between experiment and theory.

Computational Model

The model circuit contained 100 conductance-based neurons: 25 excitatory and 25 inhibitory neurons on each side of the midline. Tuning curves $r_i(E)$ for 37 of the neurons were taken directly from the experimental measurements, with the other 63 generated by varying the slopes k and thresholds E_{th} of the experimental ones by uniformly distributed factors between 0.9 and 1.1, and −1° and 1°, respectively. Tuning curves of excitatory and inhibitory neurons were drawn from the same distribution.

Neuronal description—The membrane potential $V(t)$ of each neuron i is determined by

$$C\frac{dV_i(t)}{dt} = -I_{leak}(V_i) - I_{Na}(V_i,t) - I_K(V_i,t) - I_{Kt}(V_i,t) + T_i + \sum_j W_{ij}s_j(u_j,t) + B_i(t) + I_i^{noise}(t) \quad (2)$$

The intrinsic leak I_{leak}, voltage-dependent transient sodium I_{Na}, delayed rectifier potassium I_K, and transient potassium I_{Kt} currents were modeled in the Hodgkin-Huxley formalism (Supplemental Methods). T_i represents tonic input currents of vestibular origin, $B_i(t)$ represents saccadic burst command inputs, and I_{noise} is a noise current. $W_{ij}s_j$ gives the recurrent input from neuron j to i, where W_{ij} is the connection strength and $s_j(u_j, t)$ is the synaptic activation.

The synaptic activation functions $s_j(u_j, t)$ are governed by a two time-constant approach (Supplemental Methods) to steady-state activation functions $s_{\infty,j}(r_j)$. $s_{\infty,j}(r)$ were chosen from a two-parameter family of functions that increase from 0 at $r = 0$ to 1 at large r:

$$s_{\infty,j}(r) = b_{\infty,j}\left\{\frac{1}{1+\exp\{(R_{f,j}-r)/\Theta_j\}} - a_{\infty,j}\right\}, \quad \text{where } a_{\infty,j} = \frac{1}{1+\exp\{R_{f,j}/\Theta_j\}}, \quad b_{\infty,j} = \frac{1}{1-a_{\infty,j}}. \quad (3)$$

$R_{f,j}$ gives the inflection point: $s_{\infty,j}(r)$ is superlinear for $r < R_{f,j}$ and sublinear for $r > R_{f,j}$. Θ_j scales the slope of the curves: $s_{\infty,j}(r)$ increases sharply over a narrow range of r for small Θ_j, and increases gently for large Θ_j. This family allowed us to generate a wide range of sigmoidal, saturating, and approximately linear curves within the relevant range of r.

Synaptic activation curves $s_{\infty,j}(r_j)$ were chosen to be different for excitatory and inhibitory synapses, but the same within each synapse type.

Model fitting procedure—The model fitting procedure was conducted in two steps. First, we fit a conductance-based model neuron that reproduced the current injection experiments of Figure 2D. Second, we incorporated this conductance-based neuron into a circuit model of the goldfish oculomotor integrator and used a constrained regression procedure to fit the connectivity parameters W_{ij} and T_i of the circuit model for different choices of the steady-state synaptic activation functions $s_\infty(r)$.

Single-neuron model calibration: Parameters of the intrinsic ionic conductances were calibrated to accurately match the current injection experiments illustrated in Figure 2D. In the experiments, slow up-and-down ramps of injected current drove the recorded neuron across the firing-rate range observed during fixations. The model neuron's parameters were optimized to reduce the least-squares distance between the experimental and simulated cumulative sum of the spike train as a function of time (Figure 3B). Parameter optimization was performed using the Nelder-Mead downhill simplex algorithm.

To obtain the steady-state response curve $r = f(I)$ (Figure 3C), the single-neuron model was injected for 60 seconds with constant currents of various, finely discretized strengths, and the firing rate r was found from the inverse interspike intervals (ISI), discarding the first 5 sec to assure convergence to steady-state. A noise current $I_i^{noise}(t)$ was included to approximately match the coefficient of variation (CV) of ISI's observed experimentally (Aksay et al., 2003; Supplemental Methods).

Fitting the recurrent connectivity: The circuit connectivity was fit using a constrained linear regression procedure that simultaneously incorporated anatomical constraints (Figure 2B), tuning curve data (Figure 2A), and firing rate drift patterns following inactivations (Figure 2C).

To enforce the anatomical constraint that excitatory neurons project ipsilaterally and inhibitory neurons project contralaterally, contralateral excitatory and ipsilateral inhibitory weights were set to zero. Dale's law was enforced by imposing hard constraints $W_{ij}^{exc} \geq 0$ on the weights W_{ij}^{exc} from excitatory neurons and $W_{ij}^{inh} \leq 0$ on the weights W_{ij}^{inh} from inhibitory neurons. Connection weights W_{ij} and T_i onto each neuron i then were fit simultaneously to the tuning curve data and firing rate drift data using the following cost function:

$$\varepsilon_i = \sum_m \left(f^{-1}(r_{0,i} + k_i E_m) - \sum_{j \text{ connect to } i} W_{ij} s_{\infty,j}(r_{0,j} + k_j E_m) - T_i \right)^2 + \rho_{inh}^2 \left(\sum_{\substack{\text{inhibitory } j \\ \text{connect to } i}} W_{ij}^{inh} \left\langle s_{\infty,j}^{inh} \right\rangle_{no-drift} \right)^2 + \rho_{exc}^2 \left(\sum_{\substack{\text{excitarory } j \\ \text{connect to } i}} W_{ij}^{exc} \left\langle s_{\infty,j}^{exc} \right\rangle_{no-drift} \right)^2 + \lambda^2 \sum_{j \text{ connect to } i} W_{ij}^2. \quad (4)$$

The first term above penalizes the sum, over a finely discretized set of eye positions m, of the squared differences between the current $f^{-1}(r_{0,i} + k_i E_m)$ required to drive neuron i at the firing rate $r_i(E_m) = r_{0,i} + k_i E_m$ given by its tuning curve, and the current it receives when all other neurons are firing at the rates given by their tuning curves (Figure 3F).

The second term enforces the observation that, following total contralateral inactivation, no drift is observed to occur for normalized firing rates greater than approximately 5° into the half of the oculomotor range ipsilateral to the recording (Figure 5D, blue points). Since loss of current due to the inactivation disrupts the balance of currents that maintain persistent activity, we penalized the squared sum over all inputs to neuron i of the mean inhibitory current $W_{ij}^{inh}\left\langle s_{\infty,j}^{inh}\right\rangle_{no-drift}$ received over the non-drifting range of firing rates.

The third (gray) term similarly penalized the squared sum of the total mean excitatory current over the range of firing rates that did not drift following the partial ipsilateral experiments. This term guaranteed that neurons ipsilateral to a partial inactivation could maintain persistent low firing rates by assuring that minimal recurrent excitatory current was present at such low firing rates. However, this condition is overly restrictive because these low rates might be held stable over at least a portion of their firing rate range by persistent synaptic drive arriving from the stably firing neurons of the unlesioned half of the integrator. Thus, this third term was not used to strictly rule out circuits as incompatible with experiment; instead, it was used in a subset of simulations to generate a lower bound on the number of well-fit networks. In Figure 4B, the error grids report the across-neuron averages of the maximum of the root-mean current mismatches for the first two terms of the cost function for model fits in which only these two conditions were enforced. The yellow-and-black boundary line reports the location of approximately 5 pA error in the same averages for the first three terms of the cost function for fits in which all three conditions were enforced. These reports thus represent upper and lower bounds on the set of well-fit activation parameters.

The fourth term is a regularization term that penalizes excessively strong weights. Similar goodness-of-fits and circuit connectivities were obtained when, instead of the soft constraint described by the fourth term, we applied a fixed maximum weight W_{max}=0.1 nA.

The cost function described above consists of a sum of quadratic terms, which allowed the weights onto each neuron to be fit with a constrained linear regression algorithm. Because each neuron could be fit separately from every other, the overall fitting procedure represented a sequence of 100 constrained linear regressions for 101 coefficients W_{ij} and T_i (of which 50 are constrained to be zero, see Figure 2B).

Coefficients of the different regression terms (ρ_{inh}, ρ_{exc}, λ) were chosen to maximize the number of circuits that provided good fits to both the tuning curve data and the inactivation experiments (Supplemental Methods). However, the region of well-fit activation curves and basic themes of circuit organization were not observed to change significantly over a broad range of coefficient values around the optimum.

Connectivity sensitivity analysis—The sensitivity of the circuit to changing patterns of synaptic connectivity were calculated from the Hessian matrix $H_{ij}^{(k)}=\frac{\partial^2\varepsilon_k}{\partial W_{ki}dW_{kj}}$ described in the Results. For the individual connection weight analyses, the Hessian matrix for a given neuron (e.g. the k^{th} lowest-threshold neuron in each circuit) was averaged across 100 circuits generated by randomly drawing tuning curves from the experimental distribution of Figure 2A (inset). Tolerance bars were generated for each connection weight onto neuron k by determining from the Hessian the amount this weight would be required to change in order to produce a noticeable (5 pA) change in the cost function. These bars then were overlayed upon the weighted average of the optimal connection weights for the 100 circuit simulations, where each model's connection strengths were weighted by their sensitivities.

Eigenvectors and eigenvalues were found for each of the 100 randomly generated circuits. To identify salient features present across circuits, we then generated the average 1st, 2nd, 3rd, etc. eigenvectors across all 100 circuits (Figure 6E, green lines). Perturbations in Figures 6F and 6G corresponded to changing weights by a fixed vector length along all of the shown eigenvectors; thus, differences between sensitive and insensitive perturbations reflected summing (for sensitive) or cancelling (for insensitive) effects of individual weight changes, and not different sizes of weight perturbations. To produce the n^{th} column of Figure 6G, each neuron was perturbed along its n^{th} eigenvector. Further description of the connectivity analysis and perturbations along insensitive directions (Figures S6, S7) are given in the Supplemental Methods.

The above analysis implicitly assumes that the minimum of the cost function over the allowed range of weights corresponds to a local minimum, so that the first derivative is zero and the second derivatives characterize deviations from the minimum. However, because Dale's law constrains the weights to be strictly non-negative or non-positive, the best-fit parameters can occur on the boundary of the permitted set of weights. In such cases, we also computed the gradient of the cost function to determine the direction of greatest sensitivity to infinitesimal changes in weights. However, for changes in weights large enough to lead to noticeable mistuning, the increase in the cost function due to linear changes along the gradient direction were much smaller than the quadratic changes determined by the sensitivity matrix (Figure S6G). In addition, because the gradient vector reflected weights that were prevented by Dale's law from changing signs, its direction corresponded to increasing magnitudes of all zero-valued weights and therefore overlapped with eigenvector 2. Thus, for the circuits analyzed here, the gradient provided little additional information beyond that provided by the sensitivity matrix.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by NSF grant IIS-1208218-0 (MG, EA), NIH grant R01 MH069726 (MG), a Sloan Foundation Research Fellowship (MG), a Burroughs Wellcome Collaborative Research Travel Grant (MG), a UC Davis Ophthalmology Research to Prevent Blindness grant (MG), a Wellesley College Brachmann-Hoffman Fellowship (MG), a Burroughs Wellcome Career Award at the Scientific Interface (EA), and the Searle Scholars program (EA). We thank Guy Major, Jennifer Raymond, Sukbin Lim, Andrew Miri, Brian Mulloney, Michael Wright, and Jochen Ditterich for helpful comments on this work.

References

Aksay E, Baker R, Seung HS, Tank DW. Anatomy and discharge properties of pre-motor neurons in the goldfish medulla that have eye-position signals during fixations. J Neurophysiol. 2000; 84:1035–1049. [PubMed: 10938326]

Aksay E, Baker R, Seung HS, Tank DW. Correlated discharge among cell pairs within the oculomotor horizontal velocity-to-position integrator. J Neurosci. 2003; 23:10852–10858. [PubMed: 14645478]

Aksay E, Gamkrelidze G, Seung HS, Baker R, Tank DW. In vivo intracellular recording and perturbation of persistent activity in a neural integrator. Nat Neurosci. 2001; 4:184–193. [PubMed: 11175880]

Aksay E, Olasagasti I, Mensh BD, Baker R, Goldman MS, Tank DW. Functional dissection of circuitry in a neural integrator. Nat Neurosci. 2007; 10:494–504. [PubMed: 17369822]

Briggman KL, Denk W. Towards neural circuit reconstruction with volume electron microscopy techniques. Curr Opin Neurobiol. 2006; 16:562–570. [PubMed: 16962767]

Brody CD, Romo R, Kepecs A. Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. Curr Opin Neurobiol. 2003; 13:204–211. [PubMed: 12744975]

Camperi M, Wang XJ. A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability. J Comput Neurosci. 1998; 5:383–405. [PubMed: 9877021]

Cannon SC, Robinson DA, Shamma S. A proposed neural network for the integrator of the oculomotor system. Biol Cybern. 1983; 49:127–136. [PubMed: 6661444]

Chichilnisky EJ, Rieke F. Detection Sensitivity and Temporal Resolution of Visual Signals near Absolute Threshold in the Salamander Retina. J Neurosci. 2005; 25:318–330. [PubMed: 15647475]

Chklovskii DB, Vitaladevuni S, Scheffer SK. Semi-automated reconstruction of neural circuits using electron microscopy. Curr Opin Neurobiol. 2010; 20:667–675. [PubMed: 20833533]

Debowy O, Baker R. Encoding of Eye Position in the Goldfish Horizontal Oculomotor Neural Integrator. J Neurophysiol. 2011; 105:896–909. [PubMed: 21160010]

Durstewitz D, Seamans JK, Sejnowski TJ. Neurocomputational models of working memory. Nat Neurosci. 2000; 3(Suppl):1184–1191. [PubMed: 11127836]

Egorov AV, Hamam BN, Fransen E, Hasselmo ME, Alonso AA. Graded persistent activity in entorhinal cortex neurons. Nature. 2002; 420:173–178. [PubMed: 12432392]

Eliasmith, C.; Anderson, CH. Neural Engineering. Cambridge, MA: MIT Press; 2003.

Fall CP, Rinzel J. An intracellular Ca2+ subsystem as a biologically plausible source of intrinsic conditional bistability in a network model of working memory. J Comput Neurosci. 2006; 20:97–107. [PubMed: 16511655]

Gold JL, Shadlen MN. The neural basis of decision making. Annu Rev Neurosci. 2007; 30:535–574. [PubMed: 17600525]

Goldman MS. Memory without Feedback in a Neural Network. Neuron. 2009; 61:621–634. [PubMed: 19249281]

Goldman, MS.; Compte, A.; Wang, XJ. Neural Integrator Models. In: Squire, LR., editor. Encyclopedia of Neuroscience. Oxford: Academic Press; 2009. p. 165-178.

Goldman MS, Golowasch J, Marder E, Abbott LF. Global structure, robustness, and modulation of neuronal models. J Neurosci. 2001; 21:5229–5238. [PubMed: 11438598]

Goldman MS, Levine JH, Major G, Tank DW, Seung HS. Robust persistent neural activity in a model integrator with multiple hysteretic dendrites per neuron. Cereb Cortex. 2003; 13:1185–1195. [PubMed: 14576210]

Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Canada: Springer; 2009.

Jain V, Seung HS, Turaga SC. Machines that learn to segment images: a crucial technology for connectomics. Curr Opin Neurobiol. 2010; 20:653–666. [PubMed: 20801638]

Koulakov AA, Raghavachari S, Kepecs A, Lisman JE. Model for a robust neural integrator. Nat Neurosci. 2002; 5:775–782. [PubMed: 12134153]

Lisman JE, Fellous JM, Wang XJ. A role for NMDA-receptor channels in working memory. Nat Neurosci. 1998; 1:273–275. [PubMed: 10195158]

Machens CK, Brody CD. Design of continuous attractor networks with monotonic tuning using a symmetry principle. Neural Computation. 2008; 20:452–485. [PubMed: 18047414]

Machens CK, Romo R, Brody CD. Flexible control of mutual inhibition: a neural model of two-interval discrimination. Science. 2005; 307:1121–1124. [PubMed: 15718474]

Machens CK, Romo R, Brody CD. Functional, but not anatomical, separation of “what” and “when” in prefrontal cortex. J Neurosci. 2010; 30:350–360. [PubMed: 20053916]

Major G, Polsky A, Denk W, Schiller J, Tank DW. Spatiotemporally graded NMDA spike/plateau potentials in basal dendrites of neocortical pyramidal neurons. J Neurophysiol. 2008; 99:2584–2601. [PubMed: 18337370]

Major G, Tank D. Persistent neural activity: prevalence and mechanisms. Curr Opin Neurobiol. 2004; 14:675–684. [PubMed: 15582368]

Micheva KD, Smith SJ. Array Tomography: A New Tool for Imaging the Molecular Architecture and Ultrastructure of Neural Circuits. Neuron. 2007; 55:25–36. [PubMed: 17610815]

Miri A, Daie K, Arrenberg AB, Baier H, Aksay E, Tank DW. Spatial gradients and multidimensional dynamics in a neural integrator circuit. Nat Neurosci. 2011; 14:1150–1159. [PubMed: 21857656]

Mongillo G, Barak O, Tsodyks M. Synaptic Theory of Working Memory. Science. 2008; 319:1543–1546. [PubMed: 18339943]

Prinz AA. Computational exploration of neuron and neural network models in neurobiology. Methods Mol Biol. 2007; 401:167–179. [PubMed: 18368366]

Prinz AA, Bucher D, Marder E. Similar network activity from disparate circuit parameters. Nat Neurosci. 2004; 7:1345–1352. [PubMed: 15558066]

Renart A, Song P, Wang XJ. Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. Neuron. 2003; 38:473–485. [PubMed: 12741993]

Robinson DA. Integrating with neurons. Annu Rev Neurosci. 1989; 12:33–45. [PubMed: 2648952]

Schoonheim PJ, Arrenberg AB, Del Bene F, Baier H. Optogenetic Localization and Genetic Perturbation of Saccade-Generating Neurons in Zebrafish. J Neurosci. 2010; 30:7111–7120. [PubMed: 20484654]

Seung HS, Lee DD, Reis BY, Tank DW. Stability of the memory of eye position in a recurrent network of conductance-based model neurons. Neuron. 2000; 26:259–271. [PubMed: 10798409]

Shen L. Neural integration by short term potentiation. Biol Cybern. 1989; 61:319–325. [PubMed: 2550085]

Sklavos SG, Moschovakis AK. Neural network simulations of the primate oculomotor system IV. A distributed bilateral stochastic model of the neural integrator of the vertical saccadic system. Biol Cybern. 2002; 86:97–109. [PubMed: 11908843]

Song P, Wang XJ. Angular path integration by moving “hill of activity”: a spiking neuron model without recurrent excitation of the head-direction system. J Neurosci. 2005; 25:1002–1014. [PubMed: 15673682]

Sporns O, Chialvo DR, Kaiser M, Hilgetag CC. Organization, development and function of complex brain networks. Trends Cogn Sci. 2004; 8:418–425. [PubMed: 15350243]

Wang XJ. Synaptic reverberation underlying mnemonic persistent activity. Trends Neurosci. 2001; 24:455–463. [PubMed: 11476885]

Wang Y, Markram H, Goodman PH, Berger TK, Ma J, Goldman-Rakic PS. Heterogeneity in the pyramidal network of the medial prefrontal cortex. Nat Neurosci. 2006; 9:534–542. [PubMed: 16547512]

Wei DS, Mei YA, Bagal A, Kao JP, Thompson SM, Tang CM. Compartmentalized and binary behavior of terminal dendrites in hippocampal pyramidal neurons. Science. 2001; 293:2272–2275. [PubMed: 11567143]

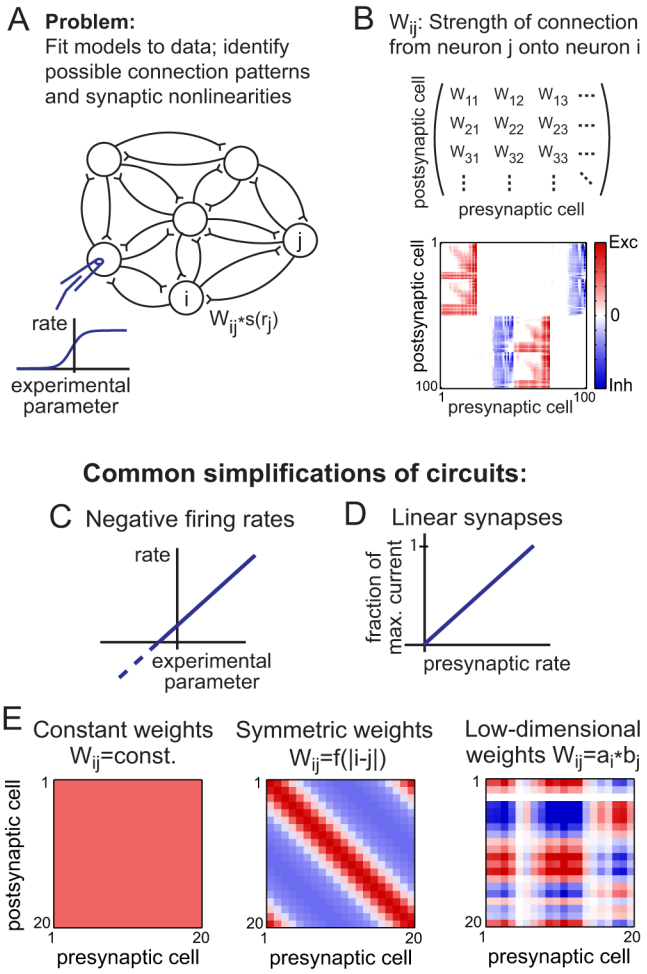

Figure 1. The challenge of understanding memory circuits

(A,B) In recurrent circuits, the activity of a given neuron both influences and depends upon the activity of other neurons in the circuit. Circuit-level interactions between cells can be characterized by two factors, the matrix of connection strengths W_{ij} (panel B) and the transformations in the synaptic pathways from one cell to another defined by the synaptic activation functions $s(r)$. Many possible sets of connection strengths and synaptic activations could support persistent firing. Here we show how combining multiple types of experimental measurements can be used to constrain the set of possible circuit models. **(C–E)** To model recurrent memory circuits, previous studies were forced to make strong simplifications such as allowing negative firing rates (C, dashed line) and assuming linear synaptic activation functions (D) to permit the use of linear systems analysis (e.g. Goldman, 2009; Machens et al., 2010; Robinson, 1989), or assuming simplified connectivities like constant (left), symmetric (middle, see Machens and Brody, 2008; Renart et al., 2003 for review), or low-dimensional (right, see Eliasmith and Anderson, 2003; Seung et al., 2000) weight matrices.

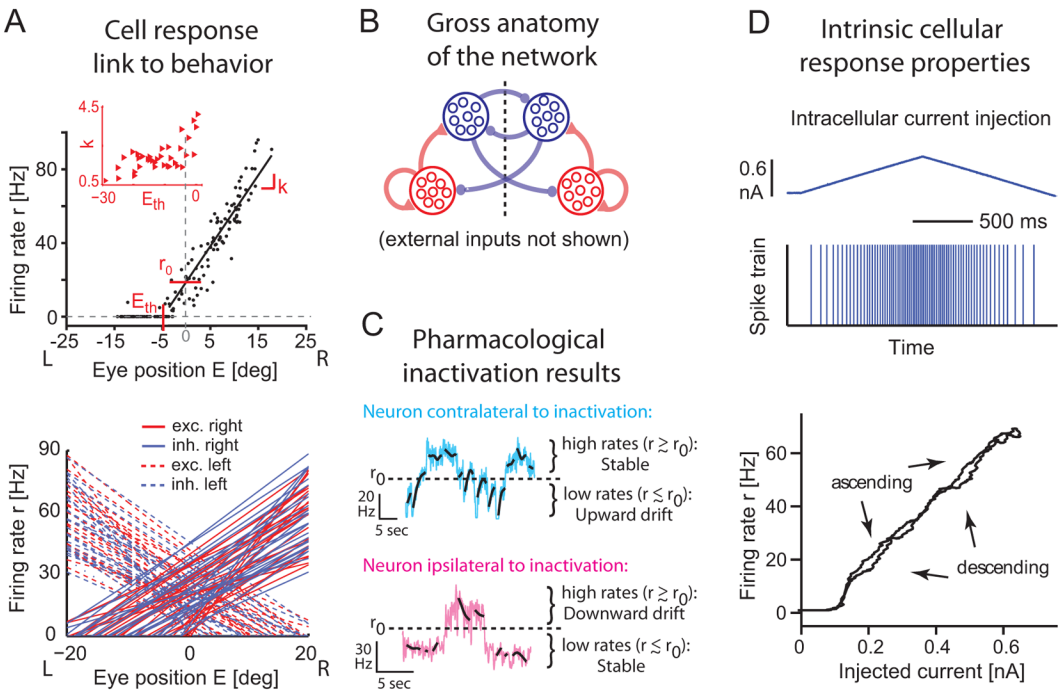

Figure 2. Data used to constrain a model of short-term memory in the oculomotor neural integrator

(A) (Top) Firing rate vs. eye position tuning curve of a right-side integrator neuron. Leftward (L) eye positions are defined as negative and rightward (R) as positive. Data points show time-averaged values for individual fixations. Inset, tuning curve parameters for 37 recorded neurons; to pool all data together, left-side neurons are plotted with $-k$ and $-E_{th}$. For our modeling, the measured tuning curves were supplemented with 63 others resampled from the original 37 (Experimental Procedures) to generate a bilateral population of 100 neurons (bottom). **(B)** Schematic of the recurrent anatomical connectivity suggested by experiments in the goldfish oculomotor integrator. The dashed line indicates the midline. Excitatory neurons (red) project ipsilaterally. Inhibitory neurons (blue) project contralaterally. **(C)** Example firing rate traces illustrating drift in firing rates following complete contralateral (blue) or partial ipsilateral (red) inactivation (adapted from Aksay et al., 2007). In each case, minimal drift occurs over the firing rate range corresponding to when the inactivated population would have been firing at low rates. For population averages, see Figure 5D. **(D)** Response of an integrator neuron to a current ramp injection during a fixation (top: injected current; middle: spike train; bottom: firing rate). Panels A (top), C, and D adapted with permission from Aksay et al. (2000), Aksay et al. (2007), and Aksay et al. (2001), respectively.

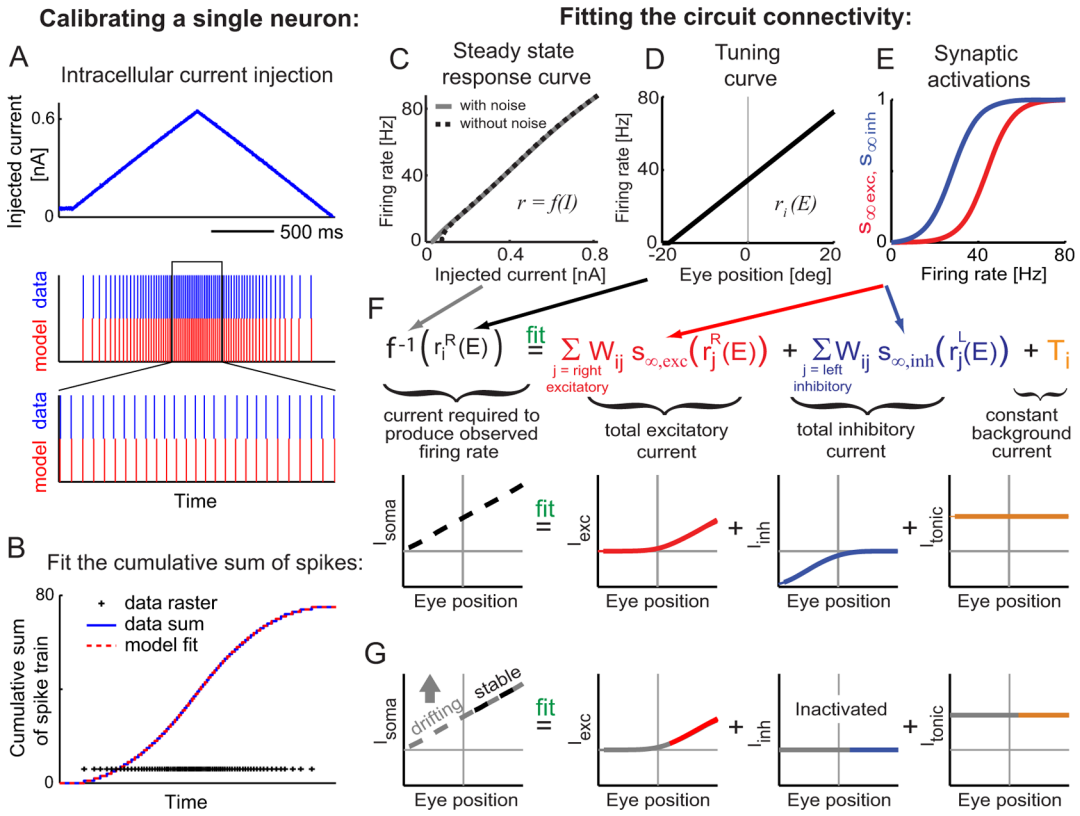

Figure 3. Deriving model parameters directly from experimental data

(A,B) Fit of the single-neuron model. (A) Experimental current injection profile (top), and experimental (blue) and calibrated model (red) neuron spike trains (middle and bottom). (B) Model neuron parameters were determined by fitting the cumulative sum of spikes as a function of time (+'s indicate spike times). **(C–E)** Key components of the circuit model: (C) Steady-state response curve of the calibrated spiking neuron in the presence of noise (without-noise case shown for comparison). (D) Tuning curve of a neuron. (E) Example activation curves for excitatory (red) and inhibitory (blue) synapses. **(F)** From the steady-state response curve (C) and the tuning curve of each neuron (D), we determine the somatic current necessary to maintain the observed persistent firing rate at any eye position (black). This current is provided by a weighted combination of the recurrent excitatory (red) and inhibitory (blue) synaptic activation functions s_{∞}, plus tonic background currents T (orange) (Figure S1). For a given form of the synaptic activations, and considering different (discretized) eye positions as different data points, the weights W_{ij} and background current T_i are determined by a constrained linear regression. **(G)** Because firing rates corresponding to eye positions greater than 5^0 past the primary eye position (black or colored portions of graphs) are maintained following total silencing of inhibitory inputs by unilateral inactivation, we required that these high rates be maintained even when the recurrent inhibitory input is set to zero. See also Figures S1 and S8.

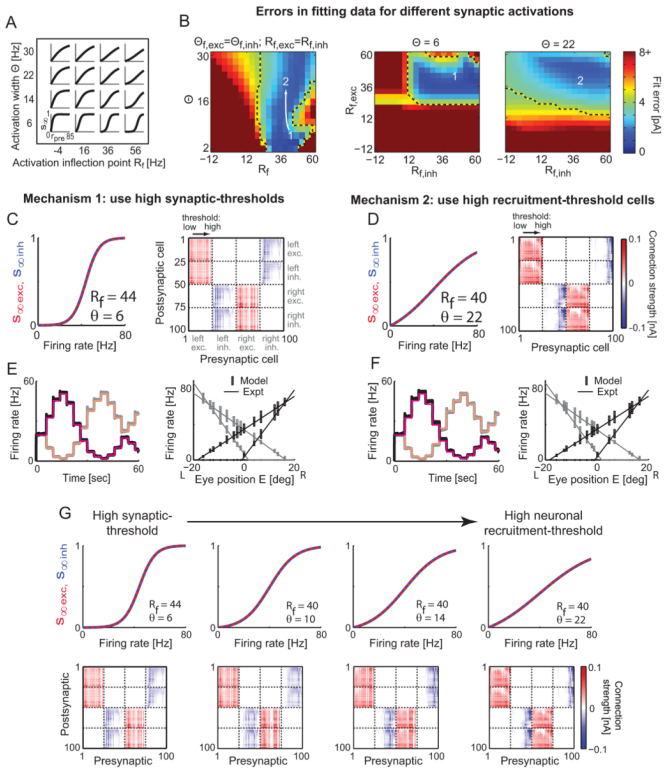

Figure 4. Circuit mechanisms enabling all experiments to be fit: control data

(A) Examples from the parameter grid of tested synaptic activation curves $s_{\infty\infty}(r)$. **(B)** Accuracy of the circuit fitting for different choices of synaptic activation functions $s_{\infty,exc}(r)$ and $s_{\infty,inh}(r)$. Three example planes in the 4-dimensional parameter space ($R_{f,exc}$, $R_{f,inh}$, Θ_{exc}, Θ_{inh}) are shown. (Left) $s_{\infty,exc}(r) = s_{\infty,inh}(r)$ plane; (middle) plane of relatively steep activation curves, $\Theta_{exc} = \Theta_{inh} = 6$ Hz; (right) plane of relatively gentle activation curves, $\Theta_{exc} = \Theta_{inh} = 22$ Hz. Main grids show model fits without the third term of the cost function; black and yellow dashed line indicates location of a 5 pA fit error for fits that included the third term. See Figure S3 for calibration of color scale. **(C,D)** Synaptic activation curves and weight matrices for example circuits using synaptic (C) or neuronal recruitment-threshold (D) mechanisms. Weight matrices W_{ij} are plotted with cells grouped by side and synaptic polarity (dashed lines) and, within each group, ordered by increasing eye-position threshold $-|E_{th}|$. See text for details. **(E,F)** Circuits using each mechanism integrate arbitrary sequences of saccadic inputs (not shown) into stable sequences of fixations (left: mean rate of right-side (black) and left-side (gray) neurons and corresponding rates for a perfect integrator (magenta, orange)). At each fixation, every model neuron's firing rate (boxes: mean±25th percentiles) precisely reproduces its corresponding experimental neuron's tuning curve (solid lines, 4 examples). **(G)** Illustration of an insensitive direction (from points 1 to 2 in panel B) in the parameter space of synaptic activations. See also Figures S2–S5 and S8.

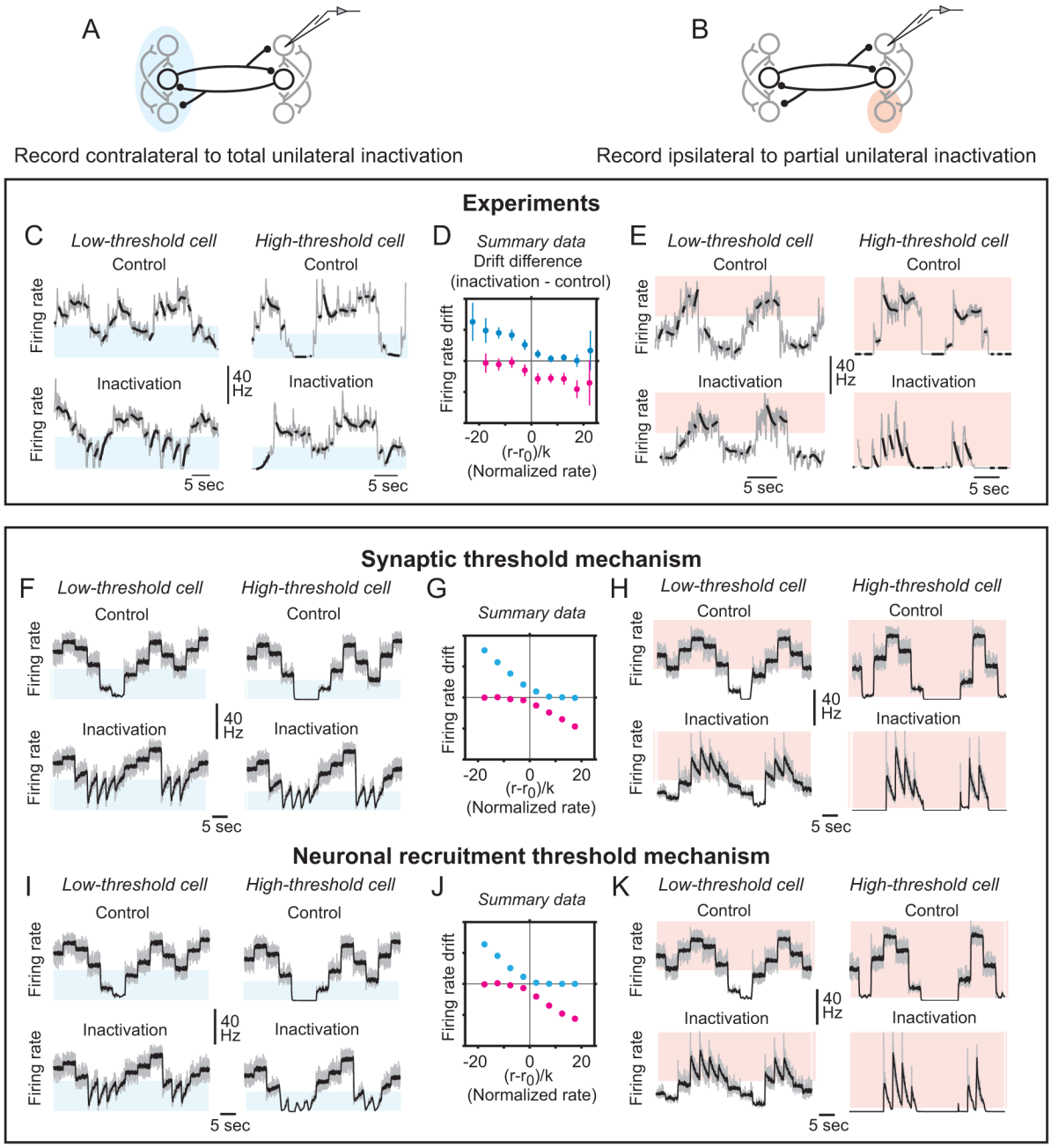

Figure 5. Circuit mechanisms enabling all experiments to be fit: inactivation data

Comparison of the effects of selective inactivations on the goldfish oculomotor integrator in the experiments (C–E) and in model circuits using each mechanism (F–K). **(A,B)** Schematics of contralateral (left) and ipsilateral (right) inactivation experiments. **(C)** Experimental firing rates before (first row) and after (second row) complete contralateral inactivation. Here and below, we show both firing rate (gray) and smoothed firing rate (black), and highlight in blue the regions where the firing rate for the shown neuron is below its primary rate r_0. **(D)** Population average drift rates. Both firing rate and drift were normalized to allow data from different fish to be pooled (Experimental Procedures). Individual points: difference between average drift in control fish and following inactivation, calculated in discretized bins. Vertical lines: 95% confidence intervals. **(E)**

Same layout described for C, but for partial ipsilateral inactivation (and with different recorded neurons). Red regions highlight where firing rate is above the primary rate r_0. **(F–K)** Results of inactivation of the simulated spiking networks of Figures 4C(F–H) and 4D (I–K). Layout is identical to C–E above. Panels C and E adapted with permission from Aksay et al. (2007). See also Figure S2.

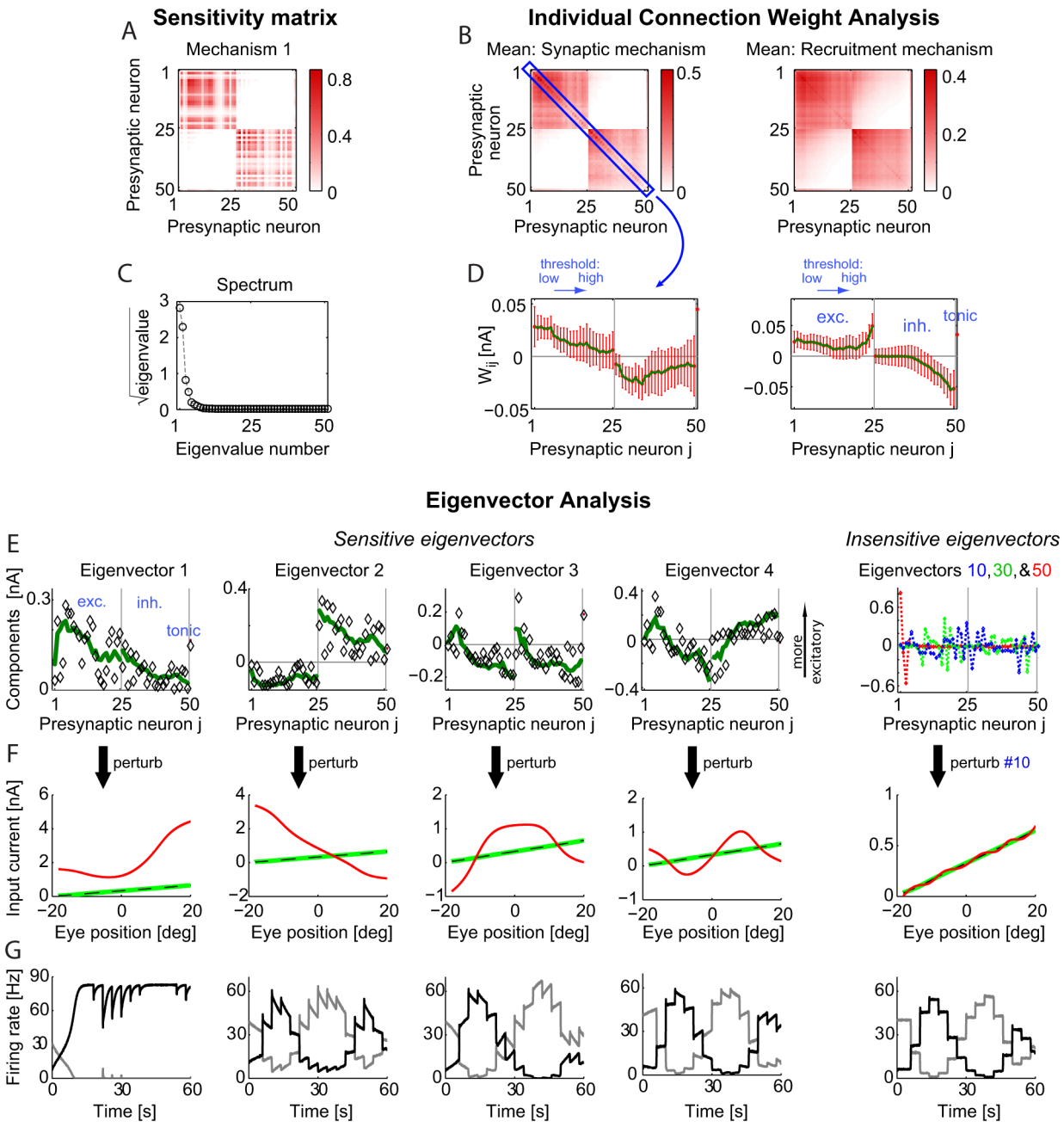

Figure 6.

Sensitivity of fits to different perturbations.

(A) Sensitivity (Hessian) matrix $H_{ij}^{(k)} = \frac{\partial^2 \varepsilon_k}{\partial W_{ki} \partial W_{kj}}$. Example corresponds to a low-threshold neuron from the circuit of Figure 4C. **(B)** Average Hessian matrix, obtained from averaging Hessians of a low-threshold neuron in 100 different circuits generated by random draws of tuning curve parameters from the experimental distribution of Figure 2A (inset). **(C)** Square roots of the eigenvalues of the Hessian matrix of panel A, which give sensitivity to perturbations along the corresponding eigenvector directions. **(D)** Tolerance (red bars) to changes in individual connection strengths away from their mean best-fit values (green). Bar

lengths were found from the diagonal elements of the average Hessian matrix and indicate the weight change required to produce a ±5 pA change in the cost function that lead to noticeably disrupted circuit performance. **(E–G)** Sensitivity of the fits to perturbations of combinations of connection strengths corresponding to the 1^{st} – 4^{th} most sensitive, and a selection of the insensitive, eigenvectors of the Hessian in panel A. **(E)** Eigenvectors of the Hessian. Symbols: eigenvector elements for the Hessian of panel A. Dark green: ensemble average of the eigenvectors over 100 different circuits generated from the experimental tuning curve data distribution. **(F)** Change in the recurrent input fit produced by a 1 nA perturbation along the respective eigenvector. Black: current required for a perfect integrator; green: current in tuned circuit; red: current in perturbed circuit. **(G)** Changes in the spiking network performance produced by a 30 pA perturbation along the respective eigenvector (black, mean right side rate; gray, mean left side rate). See also Figures S6 and S7.

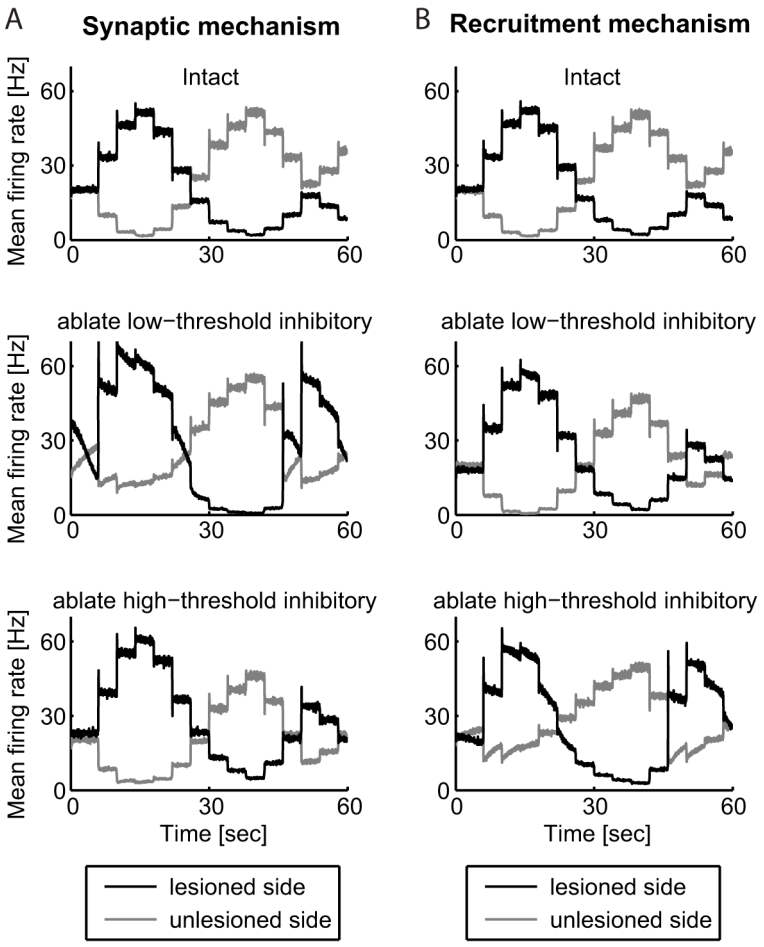

Figure 7. Predictions for selective ablation experiments

Simulation results for the two circuits of Figure 4 before and after a selective ablation of low- or high-threshold inhibitory neurons. **(A)** Circuit with high-threshold synapses.(Top) Mean rates of left-side (gray) and right-side (black) neurons in the intact circuit. (Middle) Ablation of the 12 lowest threshold inhibitory neurons on the right side of this circuit results in a strong drift towards the primary rate r_0. (Bottom) Ablation of the 12 highest-threshold inhibitory neurons on the right side has little effect. **(B)** The opposite behavior is predicted for circuits based upon the high neuronal threshold mechanism. See also Figure S2.

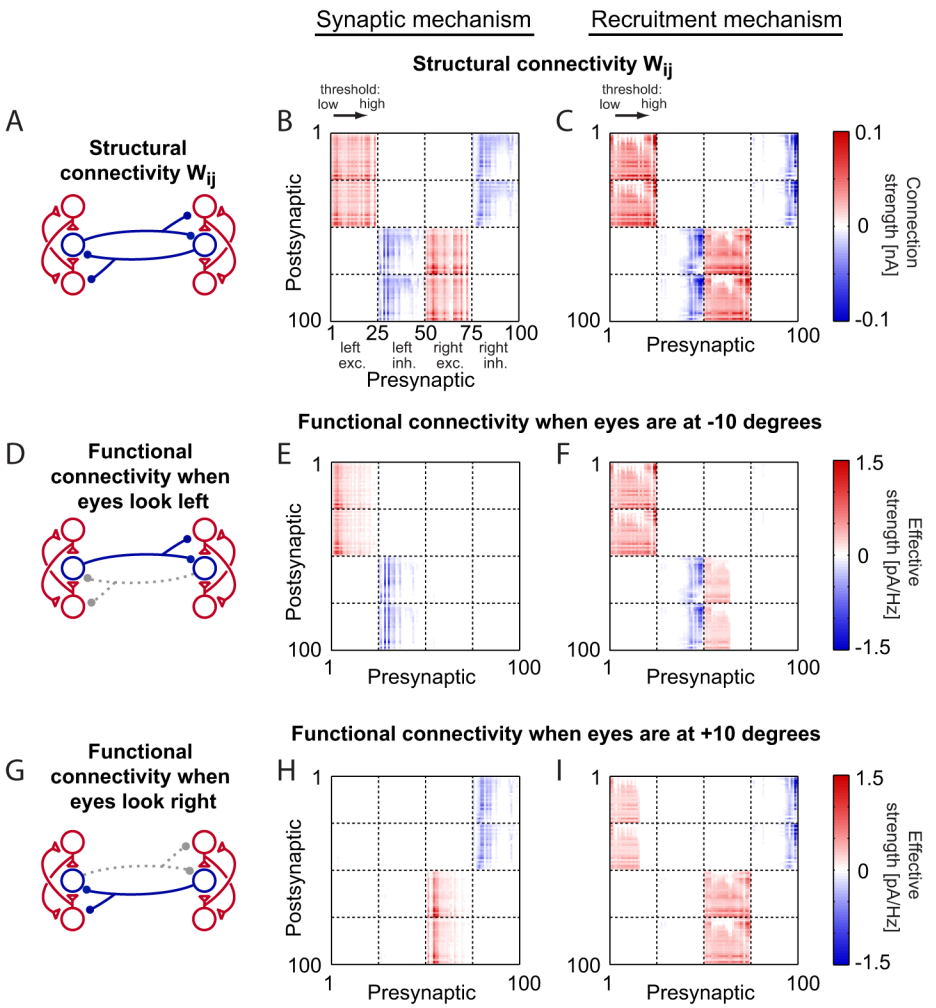

Figure 8. Different structural connectivities hide similarities in functional connectivity

Structural and functional connectivities for the two circuits of Figure 4C,D. **(A)** Schematic of gross anatomy. **(A–C)** Connectivity matrices W_{ij} provide the structural connectivity of the circuits, but do not reflect nonlinearities in intrinsic and synaptic properties. Indexing of neurons by location and eye-position recruitment threshold is as in Figure 4C. **(D–F)** Functional connectivity when the eyes are directed leftward at $E = -0$ deg. Shown are the transmitted currents per spike, $W_{ij}\, s_{\infty,\,j}\,(r_j)/(r_j)$ from neuron j to i, for all i and j. Note the absence of inhibitory functional connectivity from the right side to the left side in both circuit types (E,F, presynaptic neurons 76–100; D, gray dotted lines). **G–I**: Functional connectivity when the eyes are directed rightward, at $E = +10$ deg.