E/I balance enables adaptation to coordinate a neural population for a stable sensory representation

Gabrielle J. Gutierrez[1,2] and Sophie Denève[3]

[1] Department of Applied Mathematics, University of Washington, Seattle, WA

[2] Department of Physiology and Biophysics, University of Washington, Seattle, WA

[3] Group for Neural Theory, Ecole Normale Supérieure, Paris, France

Correspondence should be sent to Gabrielle J. Gutierrez : ellaG9@uw.edu

Abstract

Adaptation is a key component of efficient coding in single neurons. However, it is unclear how a population of adapting neurons manage to accurately and stably encode their inputs. We start with an efficient coding framework and show that realistic spike-frequency adaptation emerges as a mechanism that enables a neural population to solve a global cost-accuracy tradeoff. We learn that adaptation is managed by E/I balanced recurrent connectivity. Such coordinated population adaptation re-distributes activity from highly responsive neurons to less responsive neurons, rather than causing a global response suppression. As a result, the decoded representation remains stable despite changing activities in each neuron. In applying this framework to a model that encodes orientation, we replicate experimental findings, such as apparently-Poisson variability and the tilt illusion. Our results indicate the potential mechanisms behind these statistical and perceptual effects and underscore the diversity of neural adaptation and its role in producing a stable representation of the stimulus. We find that facilitation and dis-inhibition emerge along with the more predictable effects of neural adaptation, such as suppression.

Introduction

The range of firing rates that a sensory neuron can maintain is limited by biophysical constraints and available metabolic resources. Yet, these same neurons have to represent sensory inputs whose strength varies by orders of magnitude. Early work by Barlow and Laughlin (Barlow, 1961; Laughlin, 1981) hypothesized and demonstrated that sensory neurons in early processing stages adapt their response threshold and gain as a function of the range of inputs that they recently received. A particularly striking example of such gain modulation at the single cell level has been shown in the fly H1 neuron (Brenner et al., 2000). Gain adaptation to input has been observed in other early sensory circuits (Solomon and Kohn, 2014; Wark et al., 2007), such as in the retina (Gollisch and Meister, 2010; Kastner and Baccus, 2014), and auditory hair cells (Wen et al., 2009), and in developing cortical neurons which acquire this property during development {Mease:2013hd}.

The work of Laughlin and Barlow was instrumental in uncovering a principle of neural encoding where adapting neural responses maximize information transfer. However, the natural follow-up question concerns the decoding of neural responses after they've been subject to adaptation. Even if such adaptation might be involved in maximizing information transfer in later processing stages (Adibi et al., 2013; Wainwright, 1999), it also results in profound changes of the mapping of neural responses to stimuli in a history-dependent manner. This raises the issue of how such adapting responses are

interpreted and reinterpreted by downstream sensory areas (Series et al., 2009). One possibility, of course, is that downstream areas do not change their decoding strategy, thus introducing systematic biases in perception that persist for as long as the adaptive effects are present. This has been interpreted as the source of perceptual illusions such as the tilt after-effect or the waterfall illusion (Barlow and Hill, 1963; Wainwright, 1999). However, such illusions are classically triggered by long presentations of salient stimuli. Neural adaptation or repetition suppression can affect neural responses even at short time scales or after only one presentation of a stimulus (REFS). If not taken into account by downstream areas, the resulting representational biases could outweigh any advantage in terms of information transfer. For example, it could become impossible to maintain a stable representation of successive sensory patterns in a population of adapting neurons. An example is given in Figure [[1]] where we presented successive stimuli corresponding to a spatial pattern of digital numbers to a population of integrate and fire neurons with receptive fields that tile the space. An optimal linear decoder was trained on a set of stimulus patterns that did not include the test patterns. Without adaptation, the neurons were easily decoded and the test patterns were recovered with the linear decoder. When the neural responses were subject to adaptation, however, the responses became strongly history dependent, and the linear decoder could not decode the test patterns as accurately. Successive presentations of the same digital number result in vastly different representations. This is because the linear decoder does not take past stimuli into account. Over long timescales, it is reasonable to consider a linear decoder that is updated to produce the most accurate stimulus representation possible. However, this requires the decoder to have additional information about the stimulus or information about the changes inherent in the stimulus. In this study, we show that this apparent dilemma is resolved when coding and adaptation are understood at the level of the population, and not at the level of individual cells.

Sensory neurons in cortex are embedded in highly recurrent networks with each cell receiving strong inhibitory currents that co-vary with excitatory currents (Graupner and Reyes, 2013) and are thus E/I balanced. Here, we show that in balanced networks, heterogeneous sensory neurons with activity dependent suppression can be seen as solving a global cost/accuracy tradeoff rather than a local tradeoff at the level of each neuron. In that case, adaptation at the level of individual neurons co-exists with a largely stable representation at the population level. Rather than being globally suppressed by adaptation, E/I balance indirectly ensures that the neural activity is redistributed from highly responsive neurons to less responsive neurons without changing the interpretation of this activity by downstream areas.

Our approach suggests that, given adaptation, neural coding cannot be understood at the level of a single neuron, except in cases where a unique sensory feature is solely encoded by a single neuron (like the H1 neuron). In areas containing large numbers of interconnected neurons with redundant selectivity, many questions about neural coding and adaptation are only meaningful when applied to whole populations. We show that the adapting tuning curves of a single neuron can reflect a collective, flexible solution found by the network in particular contexts.

Results

Derivation of a generically efficient population

We start from first principles to arrive at a network that efficiently encodes a sensory stimulus, s(t). For simplicity, we assume an arbitrary monotonic stimulus space such as luminance or color saturation. The stimulus will be decoded from the firing activity of the network neurons by summing their responses, r(t), with their respective readout weights, w. Thus, a linear decoder is defined as $\hat{s}(t) = \sum_n w_n r_n(t)$. We wish to construct a network that will minimize the difference between s and s-hat so as to produce an accurate representation of the stimulus. Additionally, we wish to impose efficiency in the neural representation. For real neurons, spiking comes with inherent metabolic costs and it is clear that neurons regulate their activity so that their firing rates don't approach infinity given a very large input. To pose this problem more formally, we define an objective function composed of two terms, one representing the precision of the representation, and the other the cost of neural activity (Boerlin et al., 2013):

$$E = (s - \hat{s})^2 + \mu \sum_n \tilde{r}_n^{\,2}$$

The μ parameter weighs the efficiency cost relative to the error. Note that the contributions of the neurons are squared, penalizing both total activity and high individual firing rates, and encouraging the neurons to share the burden of representing the stimulus. The firing rate $\tilde{r}(t)$ corresponds to the output spike trains integrated at an "adaptation" time scale, τ_a. We assume that the adaptation time scale is much longer than the decoder time scale ($\tau_a >> \tau$), meaning that the cost of firing accumulates much more slowly than the time-scale at which the signal is represented in the spikes, or would be extracted by downstream synapses. The neuron takes longer to recover from a spike than the typical time-scale of stimulation, a hall-mark of adaptation. In terms of the objective function, this expresses the fact that

maintaining high firing rates for long periods of time is far more metabolically costly than short bursts of spiking.

Note that the decoder weights will be fixed. Instead of updating the weights to better represent stimuli over several iterations, as is done for the perceptron and convolutional neural networks {REFS}, we derive a prescription for the voltage dynamics so that the network neurons can produce a reconstruction of any stimulus within the stimulus space. The objective function is minimized to obtain the following dynamical equation for the network neurons (see methods for full derivation):

$$\tau \dot{V}_i = -V_i + g_i \; w_i(\dot{s} + \tau s) - \tau g_i \sum_j w_i w_j \, o_j - \tau g_i \mu o_i + g_i \left(\frac{\tau}{\tau_a} - 1\right) \mu \tilde{r_i}$$

Where o_j is the spike train of neuron j, $g_i = 2/(w_j^2 + \mu)$ is the gain of neuron i, and τ is a membrane time constant. Each neuron is effectively representing the population decoding error, $s - \hat{s}$ given that we can express its membrane potential as $V_i = g_i w_i (s - \hat{s}) - g_i \mu \tilde{r_i}$. In other words, the membrane potential is proportional to the coding error, penalized by the past activity through an adaptive current corresponding to the past integrated activity. The neuron fires if and only if it decreases the coding error to a larger extent than it increases the cost, which is equivalent to the membrane potential exceeding the firing threshold (equal to 1). It is followed by a reset of the membrane potential to -1. Together, the network neurons perform a greedy minimization in that they fire as soon as doing so benefits the objective.

Spike-frequency adaptation emerges in the network solution

The form of the voltage equation is amenable to interpreting its terms as currents to the neuron. The $-\sum_j w_i w_j \, o_j$ term indicates that neurons are connected by mutually inhibitory synapses when their decoder weights share the same sign. The reset of the membrane potential after each spike is included as an autapse. Each neuron receives information about the stimulus as weighted feedforward input. This feedforward input includes a differentiated input signal.

The final current term corresponds to an adaptation current that, in addition to the reset current, depresses the voltage as a function of its recent activity [[Fig. 2]] but it does so on a longer time scale than the reset which is immediate. We emphasize that this spike-frequency adaptation current emerged solely as a result of the efficiency imposed in the objective (equation [1]). This indicates that spike-frequency adaptation is not only a solution to the problem of efficiency in a single, isolated neuron as has been shown before, but spike-frequency adaptation is also a solution to the efficiency that is imposed on an

entire population of neurons. We next illustrate how this solution prescribes the intrinsic properties of individual neurons.

Properties of individual neurons

The network solution obtained provides feedforward, recurrent, and autapse weights in terms of the parameters that are present in the starting assumptions. Each of these weights are factored by the intrinsic gain of each neuron, which can be used as a metric of characteristic excitability. In our network paradigm, gain is inversely related to the decoding weight of the neuron. Neurons with the smallest decoding weights have the largest gain, and vice-versa [(Fig. 3)]. Assuming all neurons have the same baseline firing threshold, the feed-forward weights from input i to neuron j are set to $w_{ij}g_i = 2w_{ij}/(\sum_j w_{ij}^2 + \mu)$. Thus, the feedforward weights are inversely related to decoding weights equal to or greater than 1 [[Fig. 3B]].

Each neuron discretizes its response with spikes, therefore the neurons have a precision that is proportional to their decoding weights. The contribution of each spike to the estimate corresponds to the discretization error. In this regime, a neuron's firing rate is effectively inversely proportional to its decoding weight. Neurons with small decoding weights represent smaller changes in signals. As a result, they require more spikes than neurons with larger decoding weights to represent the same stimulus (Fig.[[3A]]). These "High gain" neurons are precise, but due to their high excitability, they are costly. In contrast, neurons with large decoding weights (referred to as "Low gain" neurons) bring less precision to the estimate, but they are more efficient since they can track the stimulus with relatively few spikes. From now on we will refer to neurons by their gains rather than decoding weights to draw a connection to the concept of neural excitability.

Our model is general and can include many forms for the cost. We chose a cost that is the sum of squared filtered spike trains, more generally known as an L2 cost or quadratic penalty. This cost implements an exponential increase in spiking threshold after each spike. This determines, in turn, the response properties of individual neurons when recorded in isolation without any contribution from recurrent connections. Model neurons given a constant stimulus input fire at a rate that decreases exponentially over time [(Fig. 3C)]. The time constant of this adaptation is ultimately determined by the neuron's decoding weight. High gain neurons are more strongly adapted because their excitable responses lead to a rapidly increasing firing threshold after every spike fired in quick succession. Thus, high gain neurons are penalized by their own past activity more than their low gain counterparts.

There are very few other plausible biological mechanisms besides spike-frequency adaptation that could serve to make the network more efficient and it seems natural that the efficiency solution for an individual neuron should generalize to that for a network. But it again raises the question of how adapting neuron responses are decoded unambiguously from a network of such neurons by a static decoder. We illustrate how this conundrum is resolved by our model using a 2-neuron network as an example.

Spike-frequency adaptation is mitigated by mutual dis-inhibition

[[Figure 4]] shows the activity of two neurons that are reciprocally connected as prescribed in the derivation (schematized in Fig. [[4A]]). They receive a constant stimulus. If one looks at the responses of each neuron individually [[Fig. 4B]], one finds that their response levels fluctuate dynamically despite the fact that the signal is constant. How, then, is the network able to maintain an invariant representation of the signal? It is because the network dynamics coordinate the two neurons such that the weighted sum of their responses (the signal estimate) remains accurate. The contribution of each neuron to the estimate is shown in the middle panel of Figure [[4B]]. Initially, neuron 1 is solely responsible for maintaining the network estimate but after a brief period of activity, neuron 2 becomes active and neuron 1's activity is reduced. The accumulating cost [[bottom panel of Fig. 4B]] results in a gradual transfer of activity from the higher gain neuron 1 to the lower gain neuron 2.

Thus, while the stimulus is constant, and while the activity of the two neurons vary, their response remains around a line defined in activity space by $s = w_1r_1 + w_2r_2$. The movement of the activity along the manifold defined by the stimulus reflects a progressive redistribution of activity to satisfy the unfolding cost-accuracy tradeoff, but without affecting the stability of the representation [[Fig. 4C]].

Contrast this to the 2-neuron network that is not recurrently connected. The firing activity doesn't follow the manifold but rather, the neuron activity rates are more correlated. Both neurons are highly active at the onset of the stimulus but their activity decays due to adaptation [[Fig. 4D]]. The network estimate decays with the activity of the two neurons, leading to an unstable representation of a constant stimulus. This is worsened with stronger adaptation (higher μ).

This example illustrates the principle of how a population representation can remain stable, even while single neurons adapt, through an interplay of activity-dependent suppression and lateral inhibition through which neurons compete to represent the stimulus.

Coordinated adaptation of a neural population

Within a network with several neurons [[Fig. 5]], the recurrent connections interact with the intrinsic properties of the neurons. As in the 2-neuron example, the first neurons to be recruited are those with high gains, providing an initially very precise representation of the signal, but a costly one. These neurons inhibit the low gain neurons, preventing them from firing early in the stimulation period. Inhibition from neurons with a higher gain delays the responses of other neurons with lower gains. As time goes on, however, the high gain neurons adapt and their response starts decaying. This is enough to disinhibit neurons with slightly lower gains. These neurons fire in turn, inhibiting the high gain neurons and shortening their transient. For low gain neurons, the delay is longer and highly dependent on stimulus strength. The rise in activity is due to progressive disinhibition, not feedforward excitation. This process continues as more and more low gain neurons are recruited while high gain neurons are increasingly inhibited, until this recruitment ceases for $t \gg \tau_a$. Note that each part of the response is shaped by the activity of other neurons, and thus, these neurons would behave differently for another distribution of input gains.

Because the disinhibition of low gain neurons automatically compensates for the decay in high gain neural responses, the stimulus representation remains stable during the whole period [(Fig. 5, bottom)]. However, its precision degrades as more low gain neurons contribute to the representation. As a result, the variance of the representation increases.

Coordinated adaptation of tuning curves

An argument frequently used to mitigate the influence of neural adaptation on perception is that while this may affect the perceived strength of the stimulus, it does not affect the neural coding of the stimulus identity. For example, visual luminance is normalized in early processing stages and visual features are represented in a largely contrast-invariant way by later processing stages. However, as we can already see in Figure [[1]], adaptation could still have deep effects on the coding of stimulus patterns, if those are presented in close temporal proximity.

To illustrate what coordinated adaptation implies for population coding, we took the example of a population of visual neurons that code for the orientation of a grating. We derived a neural population that receives an oriented stimulus in the form of two input signals $s_1(t) = Ccos(2\theta)$ and $s_2(t) = Csin(2\theta)$ where C is the stimulus strength and θ is the stimulus orientation. The factor of 2 accounts for the circular symmetry of orientation, with 0 = 180 degree. Each neuron has a decoding vector in the 2D signal space, with an angle corresponding to its preferred orientation, and an amplitude

proportional to its gain. In particular, feedforward connections will maximally excite a model neuron at its preferred orientation, and will maximally inhibit it for orthogonal orientations. The cost, as before, is assumed to be proportional to the sum of squared filtered spike trains. Neurons have equally spaced preferred orientations and a partner neuron that shares the same preference. Each pair of neurons with identical preferences has two fixed gains per orientation. For example, there are two neurons that prefer vertically oriented stimuli, one has a decoding weight of 3 and the other a decoding weight of 9. The resulting network has a double-wheeled structure, schematized on Figure [[6A]].

Figure [[6B]] illustrates the spiking response of the dual-gain network to a prolonged oriented stimulus. As seen in the simpler model from Figure [[5]], high gain neurons respond first, then adapt. As the responses of the high gain neurons decay, the low gain neurons are recruited to maintain the representation. This is quantified in the tuning curves in Figure [[6C]]. The late responses of the high gain neurons are suppressed relative to their early responses. Adaptation results in a reduced gain around the adapted orientation for high gain neurons. This is due to the suppression of their transient. Conversely, the responses of the low gain neurons are facilitated in the later part of the presentation and their tuning curves are broadened. Here, the network interactions overrode the intrinsic drive for the low gain neurons to adapt. The disinhibition from the high gain neurons combined with the constant feedforward drive to these neurons results in facilitated activity rather than the suppressed activity one would expect to be caused by adaptation. Note that the tuning curves for the high gain neurons are broader than those for the low gain neurons. This is naturally the case because the high gain neurons are more excitable. They are more likely to fire in response to oriented stimuli that are near their preference than low gain neurons. It should be acknowledged that the facilitation is a parameter-dependent effect, though a complete analysis of those regimes is beyond the scope of this paper.

These features are reminiscent of the response to a one-dimensional stimulus (Fig. [[5]]). However, we also observe non-trivial effects on the neurons' tuning curves. These are illustrated in Figure [[7]]. The network has been adapted to an orientation in the middle of the range for 2 seconds and then presented a test orientation. The tuning curves in the middle of the plot echo those from [[Figure 6]]. All neurons have bell-shaped tuning curves covering a limited range of orientations, with narrower tuning curves for the low gain neurons. These curves show in greater detail why the low gain tuning curves widen. There is a flank effect for both populations of neurons where adaptation at the flank of the control tuning curve leads to a facilitated response to the adapting orientation. For stronger/weaker adaptation/stimulus, we find that the low gain neuron that prefers the adapting orientation has a

suppressed response after adaptation unlike the facilitated effect shown in Figure [[6C]]. Interestingly, however, we observe a flank effect where low gain neurons that have a similar orientation preference are facilitated. This results in a stronger contribution of low gain neurons for test stimuli near the adapted orientation. Similar observations have been made in experiments of mouse visual cortex area V1 (Jeyabalaratnam et al., 2013).

Variability

The same qualitative effects are observed in the network with random preferred orientations and gains (Fig. [[8]]). Tuning curves are suppressed around the preferred orientation, but can be either facilitated or suppressed when the adapted stimulus falls on the flank of the tuning curve, accompanied (or not) by a shift toward the adapted stimulus. However, the effect on single neurons is otherwise very variable. In fact, adaptation in one neuron is impossible to predict quantitatively without observing the rest of the network.

Furthermore, this variability extends in a trial-to-trial manner. The history-dependence of the network ensures that a given stimulus is never represented the same exact way twice by the same neurons, even if the decoded representation is stable from trial-to-trial. This is illustrated in Figure [[9]] showing the response of a network to a given stimulus during three different trials. The network is the same and the stimulus is the same. The only change is the sequence of stimuli that preceded the test stimulus. The sequence of adapting orientations all had the same magnitude and duration but their order was shuffled in each of the three trials. In each trial, the first adapting stimulus activates a set of neurons with matching preferred orientations. Some of those neurons will become fatigued and the following adapting orientation would activate a different set of neurons that may or may not overlap, and so on. After the adapting stimuli, the network will be in a different state than it was originally and in a different state than the network following a differently sequenced set of adapting stimuli, resulting in a different representation of the test stimulus. This kind of trial-to-trial variability is a common occurrence in experimental studies of individual neuron activity.

Perceptual adaptation

We have stressed the accuracy of the stimulus representation in the face of time-varying activity due to adaptation. While this kind of activity could be interpreted as leading to a stable percept in spite of adaptation, we acknowledge that perceptual errors and biases are abundant in the natural world. Our network is capable of emulating these errors and it is able to do so in a manner that is consistent with experimental findings. The network will produce a stable representation as long as there are a sufficient

number of neurons to maintain it. If the adaptation is too strong or the stimulus presentation too long, and there aren't enough neurons to capture that range of stimulus presentation variables, then the network estimate will degrade. This degradation can lead to a bias in the decoder [[Fig 10]]. An oriented, strong, adapting stimulus is presented for 2 seconds followed by a test orientation, as schematized in Figure [[10A]]. An example of the resulting network activity is shown in Figure [[10B]]. Before adaptation takes hold, the adapting stimulus activates the high gain neurons with preferences at and near the stimulus orientation. Because the adapting stimulus is strong, high gain neurons with similar preferences are quickly recruited. As the stimulus persists, the most strongly activated high gain neurons fatigue and the low gain neurons with matching preferences are recruited. Some high gain neurons with opposing preferences are also recruited due to the strong excitatory input coming from the newly activated low gain neurons. After the 2 second presentation of the adapting orientation, a weaker peripherally oriented test stimulus is delivered. The response distribution and dynamics are markedly different. Instead of a widely-tuned response, the weaker stimulus produces a more narrowly distributed response. The decoded orientation is offset from the test stimulus orientation, indicating a bias in the perceived orientation.

A classical study of such perceptual bias is the tilt illusion (Gibson and Radner, 1937). In the tilt illusion, the orientation of a test grating is perceived incorrectly after adaptation to a differently oriented stimulus. Experimental studies report that the perceived orientation is repulsed away from the adapted orientation, the effect being maximal for stimuli tilted around 15-20 degrees. This effect has been confirmed in the visual cortex (Jin et al., 2005). Our findings replicate this effect [[(Fig. 10C,D)]]. The test stimulus is a vertical grating. It is perceived to be repulsed from vertical when the adaptor is approximately 15 degrees from vertical [[Fig. 10C, middle panel]]. However, when the adaptor is quite obliquely oriented, the vertical test grating is perceived to be oriented in a direction that is attracted to the adaptor [[Fig. 10C, right panel]]. Test stimuli within a range of 0-45 degrees difference from the adaptor orientation are repulsed whereas test stimuli with a greater than 45 degree difference from the adaptor orientation are attracted [[Fig. 10D]]. In accordance with experimental findings, the repulsion effect has a greater amplitude than the attraction effect.

Discussion

As individual neurons adapt, their responsiveness varies over time. This poses a potential problem by causing time-varying activity over the course of a constant stimulus presentation because the stimulus

may not be encoded properly over the course of the presentation. A second problem arises in the trial-to-trial variability produced by adaptation that is observed at the single neuron level. The context dependence caused by adaptation begs the question of how a consistent representation can be decoded from a network in which all, or most, neurons are subject to adaptation. Our study shows that the potentially harmful effects of adaptation on the individual neuron's ability to encode a stimulus can be mitigated by a coordinated population response. Other studies that have addressed this issue propose updating the decoder or have considered synaptic plasticity mechanisms (Hosoya et al., 2005). Our study offers an alternative, plausible framework for resolving the cost-accuracy tradeoff on a shorter time scale than the operating time scale for synaptic plasticity.

Our model is developed from an efficient predictive coding framework (Boerlin et al., 2013; Olshausen and Field, 1996; Spratling, 2010) in which we enforce efficiency in the encoder and accuracy in the decoder. The imposed efficiency condition was presented in the form of a metabolic cost term in the objective function. This produced spike-frequency adapting activity in the encoding population. Meanwhile, the accuracy condition ensures that an accurately decoded representation will be produced even as neural responses are subject to adaptation and the decoder remains unchanged. Our normative approach allows for an investigation of the possible neural mechanisms that brain circuits might employ to solve this cost/accuracy tradeoff. We found that E/I balanced recurrent connectivity permits a network to manage the deleterious effects of adapting neural activity. An E/I balanced connectivity scheme maintains network neurons near their firing threshold. Disinhibition from adapting neurons can then quickly recruit neurons with similar preferences so as to maintain a stable network output. This approach can be generalized to many other types of cost, arbitrary weights and number of neurons.

Single neuron coding is dynamic rather than a static property

Our model suggests that diverse adaptation properties within a population can be an asset. A heterogeneous population of neurons is able to better distribute the cost to maximize efficiency in different contexts. Studies in the retina show that retinal ganglion cells with different adaptive properties complement each other such that sensitizing cells can improve the encoding of weak signals when fatiguing cells adapt (Kastner and Baccus, 2011). This arrangement is particularly advantageous for encoding contrast decrements which would be difficult to distinguish from the prior stimulus distribution if only suppressive adaptation prevailed. At the same time, these heterogeneities contribute to complex dynamics in the neural spike trains, obscuring the relationship between neural activity and neural coding

for an observer of single neuron activity. We make the prediction that neurophysiological studies where single neuron activity is recorded may exhibit an experimental bias that results in highly responsive neurons being overrepresented in the sample. However, more recent methods based on recording large, dense population of neurons may not suffer as much from this selection. In those studies, the observed variability of adaptation effects are in line with our predictions {REFS}.

Moreover, our study challenges the notion that tuning is a static characteristic of neurons. Experiments increasingly reveal that neurons change their tuning dynamically with changing stimulus statistics (Hollmann et al., 2015; Hosoya et al., 2005; Nagel and Doupe, 2006; Smirnakis et al., 1997; Solomon and Kohn, 2014; Wark et al., 2007). In the visual cortex, it has been shown that the tilt after effect is not only an effect of response suppression but that it also has the effect of shifting the tuning curves of neurons away from their preferred orientations (Jin et al., 2005). While it may be possible to predict some aspect of the tuning change from measurements of intrinsic neuron properties, our study shows that a great deal of the change may be a network effect rather than an intrinsic neuronal effect. Thus, the extent of adaptation for a single neuron may be difficult to predict without considering the properties of the rest of the network. Such unpredictable adaptation could be a problem for the interpretation by downstream readouts, however, we show that when the network is considered as a whole, the adaptive effects in one neuron can be compensated for by another neuron that reports to the same readout. In other words, the apparently complex adaptation at the single neuron level and the poisson-randomness of spike trains (see (Boerlin et al., 2013)) is not an impediment to the network but rather an indicator of the manner in which the signal is encoded by the network as a whole.

Variable population codes

Our study provides a possible explanation for the trial-to-trial variability seen in recordings of neural populations. It is possible that the population as a whole is producing the most accurate encoding of its inputs that it can manage given a history-dependent adaptation of parts of the population.

Validating the framework experimentally.

How could this framework be tested experimentally, given that single neuron dynamics is impossible to predict in isolation? Our model applies at the level of relatively densely connected (and thus, local) populations. Observing the organized transfer of responses between neurons through adaptation and E/I balance would require one to record a significant proportion of these neurons locally (neurons that are likely to be interconnected directly or through interneurons). Recent experimental techniques

render such recordings possible (REFS), bringing an experimental validation of this framework within grasp. These recording could be compared before and after adaptation, over the duration of prolonged stimuli, or over many repetitions of the same stimulus. What we expect to see is a generalization of the effect illustrated in figure [[4B,C]] to larger neural populations. First of all, there should exist a decoder of neural activity, independent of stimulus history that can detect the stimulus despite large changes in neural activity over time. Second of all, shuffling the neural responses, for example between early and later part of the responses to a prolonged stimulus, should have detrimental effects on such stable decoding. And finally, over the course of adaptation, the activity of the different neurons should not vary independently. For example, if we performed dimensionality reduction (such as principle component analysis) of the neural population activity during a prolonged stimulus presentation, we might be able to observe that neural responses over time (and trials) is constrained on a subspace where the stimulus representation is stable. Another, more direct way of testing our framework would be to activate or inactivate a part of the neural populations. This could be done optogenetically for example (REFS). Such manipulations could change the way neurons (whose activity remained unperturbed) adapt. In particular, weakly responsive (low gain) neurons could start exhibiting some of the properties of high gain neurons, such as stronger, earlier transient response responses and a more pronounced subsequent decay in firing rates.

METHODS

Network model

We provide here a brief description of the network structure and the objective function it minimizes. We consider a spiking neural network composed of *N* neurons that encodes a set of *M* sensory signals, $\mathbf{s} = [s_1 \dots s_M]$. Estimates of these input signals, $\hat{\mathbf{s}} = [\hat{s}_1 \dots \hat{s}_M]$, are decoded by applying a set of decoding weights $\hat{s}_m = \sum_{j=1}^{N} w_{mj}\, r_j$, where r_j is the filtered spike train of neuron *j*. The filtered spike train corresponds to a leaky integration of its spikes, $r_j = \tau o_j * e^{\frac{-t}{\tau}}$, with $o_j = \sum_{t_j^k} \delta(t - t_j^k)$, with t_j^k the spike time of the kth spike in neuron j and τ the time scale of the decoder. As we will see, τ will correspond to the membrane time constant of the model neurons.

The decoding weights W_{ij} are given a priori (they determine the selectivity and gain of the model neurons). We want to construct a neural network that represents the signals most efficiently, given the fixed decoding weights. Efficiency is defined as the minimization of an objective function composed of two terms, one penalizing large coding errors, and the other penalizing high sustained firing rates:

$$E(t) = \left\| s(t) - \widehat{s(t)} \right\|^2 + \mu \sum_n \tilde{r}_n^{\,2}$$

The sustained firing rates are defined as an integration of the spike trains at a slow time scale, with $\tilde{r}_j = \tau_a o_j * e^{\frac{-t}{\tau_a}}$, $\tau_a > \tau$, and μ is a positive constant regulating the cost/accuracy tradeoff.

In order to minimize this objective function, we define a spiking rule that performs a greedy minimization. Thus, neuron j fires as soon as this results in a minimization of the cost, i.e. as soon as $E^{\text{spike in } j}(t) < E^{\text{no spike in } j}(t)$. Solving this equation leads to the following spike rule: neuron j spikes if

$$g_j \left(\sum_i W_{ij}(s_i(t) - \hat{s}_i(t)) - \mu r_j^a(t) \right) > 1 \qquad \text{ZEqn2} \qquad (1)$$

With $g_j = \frac{2}{\sum_i W_{ij}^2 + \mu}$ being the "gain" of neuron j. We interpret the lefthand side of this equation as the membrane potential of neuron j, and the right hand side as its firing threshold. The membrane potential dynamics are obtained by taking the derivative of the voltage expression with respect to time.

$$\tau \dot{V}_j = -V_j + g_j \sum_i W_{ij}(s_i + \tau \dot{s}_j) - \sum_k \Omega_{jk} o_k - g_j \mu r_j^a \qquad \text{ZEqn3} \qquad (2)$$

The lateral connection between neuron k and neuron j is equal to $\Omega_{jk} = g_j \sum_i W_{ik} W_{ij}$. Thus, the lateral connections measure to what extent the feed-forward connections of two neurons are correlated, and remove these correlations to obtain the most efficient code. The firing threshold of all neurons is equal to 1, while the reset is performed by the "self-connection" term $\Omega_{jj} = -2$. Thus, after each spike, the membrane potential is simply reset to -1. Note that g_j is a multiplicative term applied to all the connections (feedforward and lateral) as well as on the spike-based adaptation term. Moreover, the gain is approximately inversely proportional to the norm of the decoding weights. Generally, the feedforward connections of the neuron will scale inversely with the strength of its contribution to the decoded signals.

Digital number encoding network

The network used in Figure [[1]] is a generic recurrent network of 400 neurons with random recurrent and feedforward weights. The feedforward weights are a 7x400 matrix of values drawn from a uniform distribution in the [-1,1] range. The recurrent weights are drawn from a Gaussian distribution

with mean = 0, std = 0.87 (close to 1) and are a 400x400 matrix, however, all neurons had an autapse that was the sum of the negative squares of its feedforward weights. The network was trained on 100 stimulus examples of 300 ms each that were generated randomly from a uniform distribution of input values from [0,4]. An optimal linear decoder was obtained from this training by taking the inverse of the responses and multiplying them by the stimulus training examples: decoder = pinv(r(t))s(t). The trained network was then presented with a sequence of 8 digitized patterns of 200 ms each separated by 100ms of no stimulus input. To demonstrate the effect of adaptation, the trained network was run on the same stimulus sequence and with the same linear decoder but this time the spiking threshold was dynamically regulated by past spiking activity such that $threshold(t) = 1 + \mu\tilde{r}(t)$, where $\dot{\tilde{r}}(t) = -\frac{1}{\tau_a}\tilde{r}(t) + o(t)$.

Orientation model

The network follows the same derivation as outlined for the network model. It has 2 dimensions and 200 neurons. There are two subpopulations of neurons such that 100 neurons are high gain neurons with a feedforward gain of 3 and the remaining 100 neurons are low gain neurons with a feedforward gain of 9. Because the gain inverts the feedforward weights, low gain neurons have a low intrinsic gain and vice versa. Both populations span the unit circle evenly such that one low gain and one high gain neuron share the same preferred orientation. Tuning curves in Figure [[6]] were generated by presenting the network with a full range of stimulus orientations of gain=50. Neuron responses were centered on their preferred orientation and the mean was taken for each subpopulation. Control tuning curves (no adaptation) were normalized to one, tuning curves after adaptation were normalized to control tuning curves. Tuning curves after adaptation were made by lining up neuron responses to an adapting stimulus that corresponded with its preferred orientation. Standard deviations were computed on these centered data.

The random gain network was identical to the above with the exception that the feedforward weight gains were randomly selected from a uniform distribution in the range [3,9].

The tilt illusion curve was generated by presenting the network with an adaptor orientation (duration = 2s) and a subsequent test orientation (250ms). The test orientation remained the same while a series of adaptor orientations were used. The encoded angle was decoded from the network output by taking the arc tangent of the mean output over the 250ms presentation of the test stimulus. The difference between the decoded orientation and the test orientation was plotted. The adaptor had a stimulus gain of 25 while the test had a gain of 5.

References

Adibi, M., McDonald, J.S., Clifford, C.W.G., Arabzadeh, E., 2013. Adaptation Improves Neural Coding Efficiency Despite Increasing Correlations in Variability. Journal of Neuroscience 33, 2108–2120. doi:10.1523/JNEUROSCI.3449-12.2013

Barlow, H.B., 1961. Possible Principles Underlying the Transformations of Sensory Messages, in: Sensory Communication. Sensory Communication, pp. 216–234. doi:10.7551/mitpress/9780262518420.003.0013

Barlow, H.B., Hill, R.M., 1963. Evidence for a Physiological Explanation of the Waterfall Phenomenon and Figural After-Effects. Nature 200, 1345–1347.

Boerlin, M., Machens, C.K., Denève, S., 2013. Predictive Coding of Dynamical Variables in Balanced Spiking Networks. PLoS Comput Biol 9, e1003258–16. doi:10.1371/journal.pcbi.1003258

Brenner, N., Bialek, W., van Steveninck, R.D., 2000. Adaptive rescaling maximizes information transmission. Neuron 26, 695–702. doi:10.1016/S0896-6273(00)81205-2

Gibson, J.J., Radner, M., 1937. Adaptation, after-effect and contrast in the perception of tilted lines. I. Quantitative studies. Journal of Experimental Psychology 20, 453–467. doi:10.1037/h0059826

Gollisch, T., Meister, M., 2010. Eye Smarter than Scientists Believed: Neural Computations in Circuits of the Retina. Neuron 65, 150–164. doi:10.1016/j.neuron.2009.12.009

Graupner, M., Reyes, A.D., 2013. Synaptic input correlations leading to membrane potential decorrelation of spontaneous activity in cortex. J. Neurosci. 33, 15075–15085. doi:10.1523/JNEUROSCI.0347-13.2013

Hollmann, V., VH, Lucks, V., VL, Kurtz, R., Engelmann, J., JE, 2015. Adaptation-induced modification of motion selectivity tuning in visual tectal neurons of adult zebrafish. Journal of Neurophysiology 114, jn.00568.2015–2902. doi:10.1152/jn.00568.2015

Hosoya, T., Baccus, S.A., Meister, M., 2005. Dynamic predictive coding by the retina. Nature 436, 71–77. doi:10.1038/nature03689

Jeyabalaratnam, J., Bharmauria, V., Bachatene, L., Cattan, S., Angers, A., Molotchnikoff, S., 2013. Adaptation Shifts Preferred Orientation of Tuning Curve in the Mouse Visual Cortex. PLoS ONE 8, e64294–8. doi:10.1371/journal.pone.0064294

Jin, D.Z., Dragoi, V., Sur, M., Seung, H.S., 2005. Tilt aftereffect and adaptation-induced changes in orientation tuning in visual cortex. Journal of Neurophysiology 94, 4038–4050. doi:10.1152/jn.00571.2004

Kastner, D.B., Baccus, S.A., 2014. Insights from the retina into the diverse and general

computations of adaptation, detection, and prediction. Current Opinion in Neurobiology 25, 63–69. doi:10.1016/j.conb.2013.11.012

Kastner, D.B., Baccus, S.A., 2011. Coordinated dynamic encoding in the retina using opposing forms of plasticity. Nat Neurosci 14, 1317–1322. doi:10.1038/nn.2906

Laughlin, S., 1981. A Simple Coding Procedure Enhances a Neurons Information Capacity. Z. Naturforsch., C, Biosci. 36, 910–912. doi:10.1515/znc-1981-9-1040

Nagel, K.I., Doupe, A.J., 2006. Temporal processing and adaptation in the songbird auditory forebrain. Neuron 51, 845–859. doi:10.1016/j.neuron.2006.08.030

Olshausen, B.A., Field, D.J., 1996. Natural image statistics and efficient coding. Network 7, 333–339. doi:10.1088/0954-898X/7/2/014

Series, P., Stocker, A.A., Simoncelli, E.P., 2009. Is the Homunculus “Aware” of Sensory Adaptation? Neural Comput 21, 3271–3304. doi:10.1162/neco.2009.09-08-869

Smirnakis, S.M., Berry, M.J., Warland, D.K., Bialek, W., 1997. Adaptation of retinal processing to image contrast and spatial scale. Nature 386, 69–73. doi:10.1038/386069a0

Solomon, S.G., Kohn, A., 2014. Moving Sensory Adaptation beyond Suppressive Effects in Single Neurons. Current Biology 24, R1012–R1022. doi:10.1016/j.cub.2014.09.001

Spratling, M.W., 2010. Predictive Coding as a Model of Response Properties in Cortical Area V1. Journal of Neuroscience 30, 3531–3543. doi:10.1523/JNEUROSCI.4911-09.2010

Wainwright, M.J., 1999. Visual adaptation as optimal information transmission. Vision Res. 39, 3960–3974. doi:10.1016/S0042-6989(99)00101-7

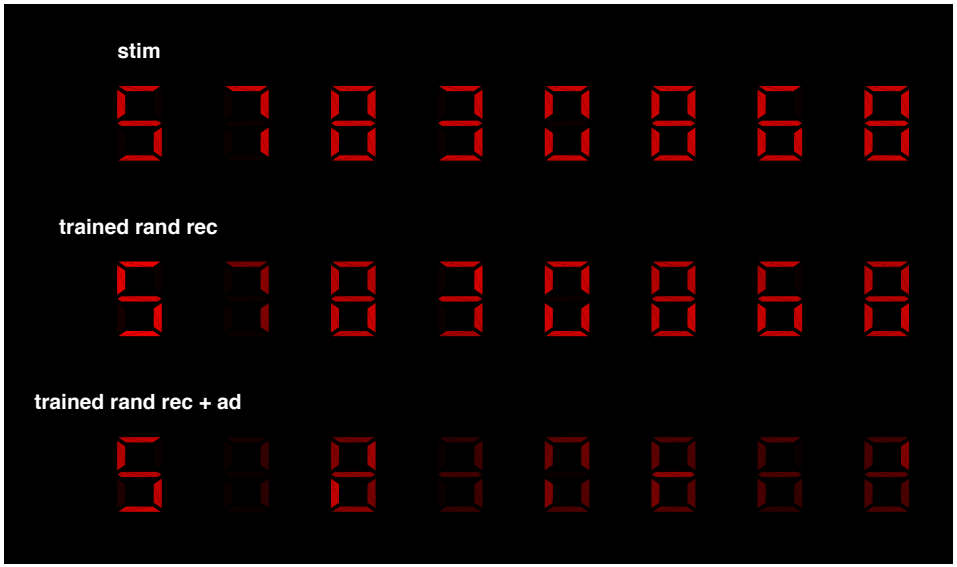
Wark, B., Lundstrom, B.N., Fairhall, A., 2007. Sensory adaptation. Current Opinion in Neurobiology 17, 423–429. doi:10.1016/j.conb.2007.07.001

Wen, B., Wang, G.I., Dean, I., Delgutte, B., 2009. Dynamic range adaptation to sound level statistics in the auditory nerve. J. Neurosci. 29, 13797–13808. doi:10.1523/JNEUROSCI.5610-08.2009

Figure 1

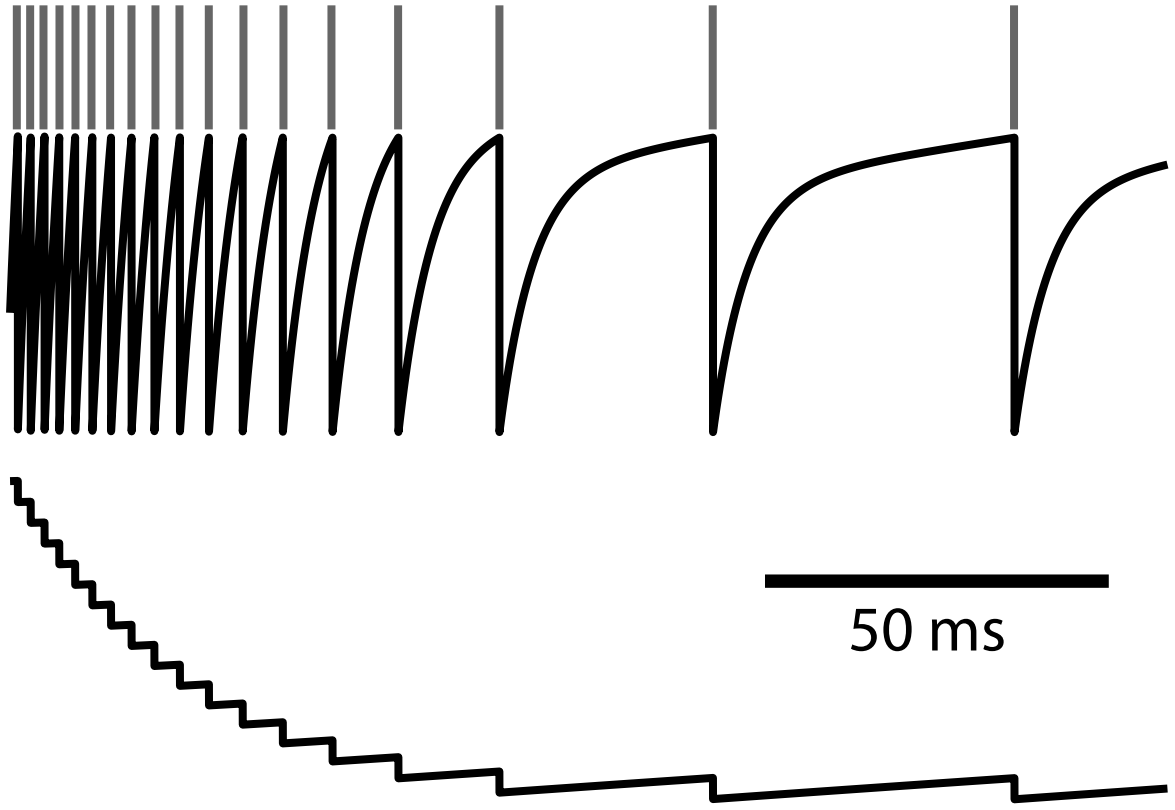
A.

B.

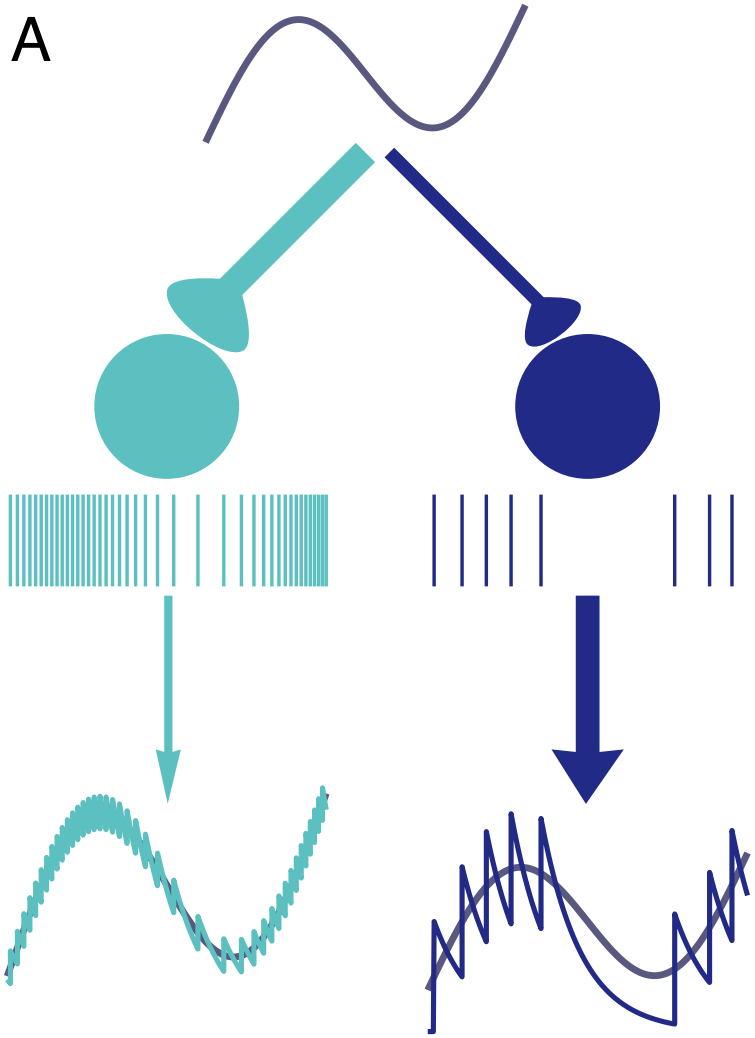


(A) schematic. (B) top, sequence of patterned stimuli, middle, trained random recurrent network output, bottom, output of same trained random recurrent network but with adapting neuron responses.

Figure 2



SFA fig. Top, voltage trace of isolated neuron with adaptation. Bottom, adaptation current of same neuron.



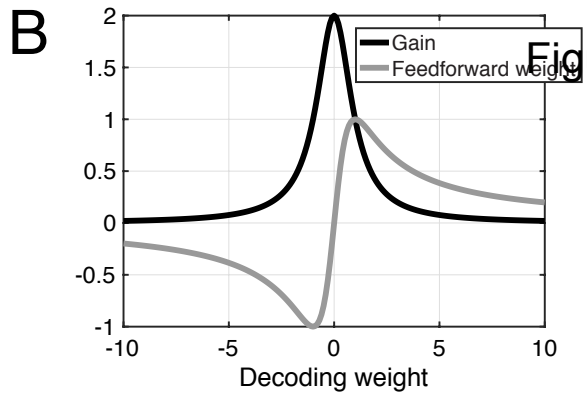
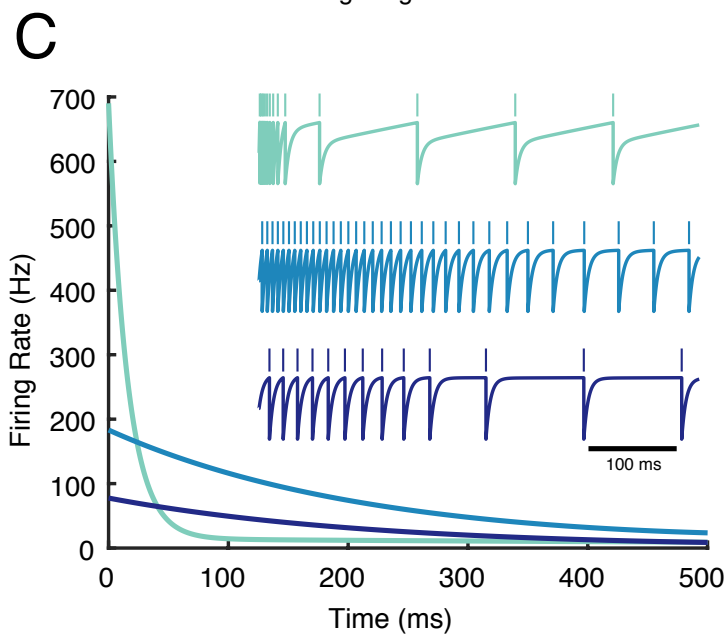


Figure 3



Intrinsic neuron properties. (A)High gain neurons (light blue) are precise while low gain neurons (dark blue) have less precision. (B) Relationship between gain and feedforward weight and decoding weight (mu=1). (C) High gain neurons have the steepest adaptation whereas low gain neurons do not adapt as rapidly given the same input.

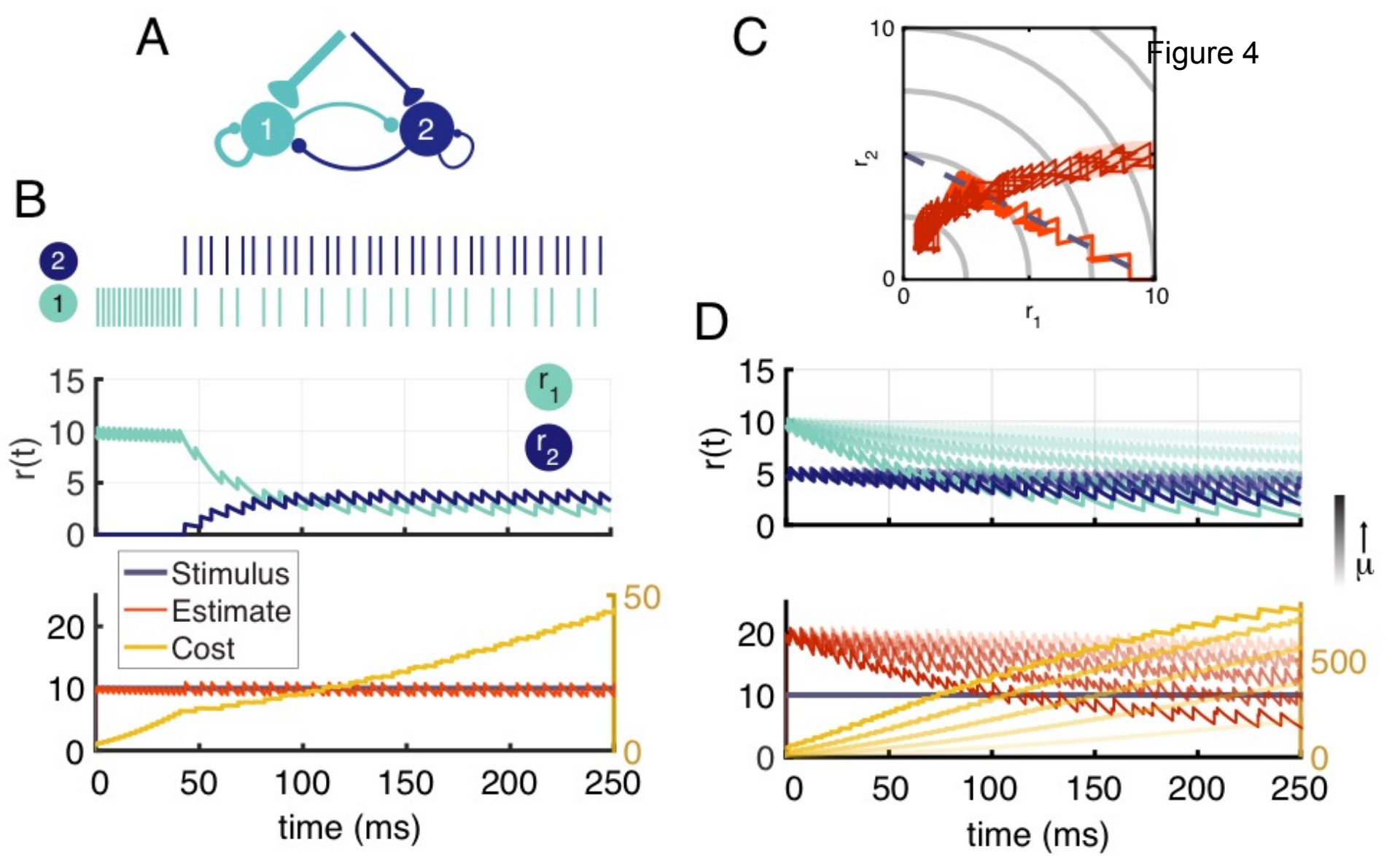
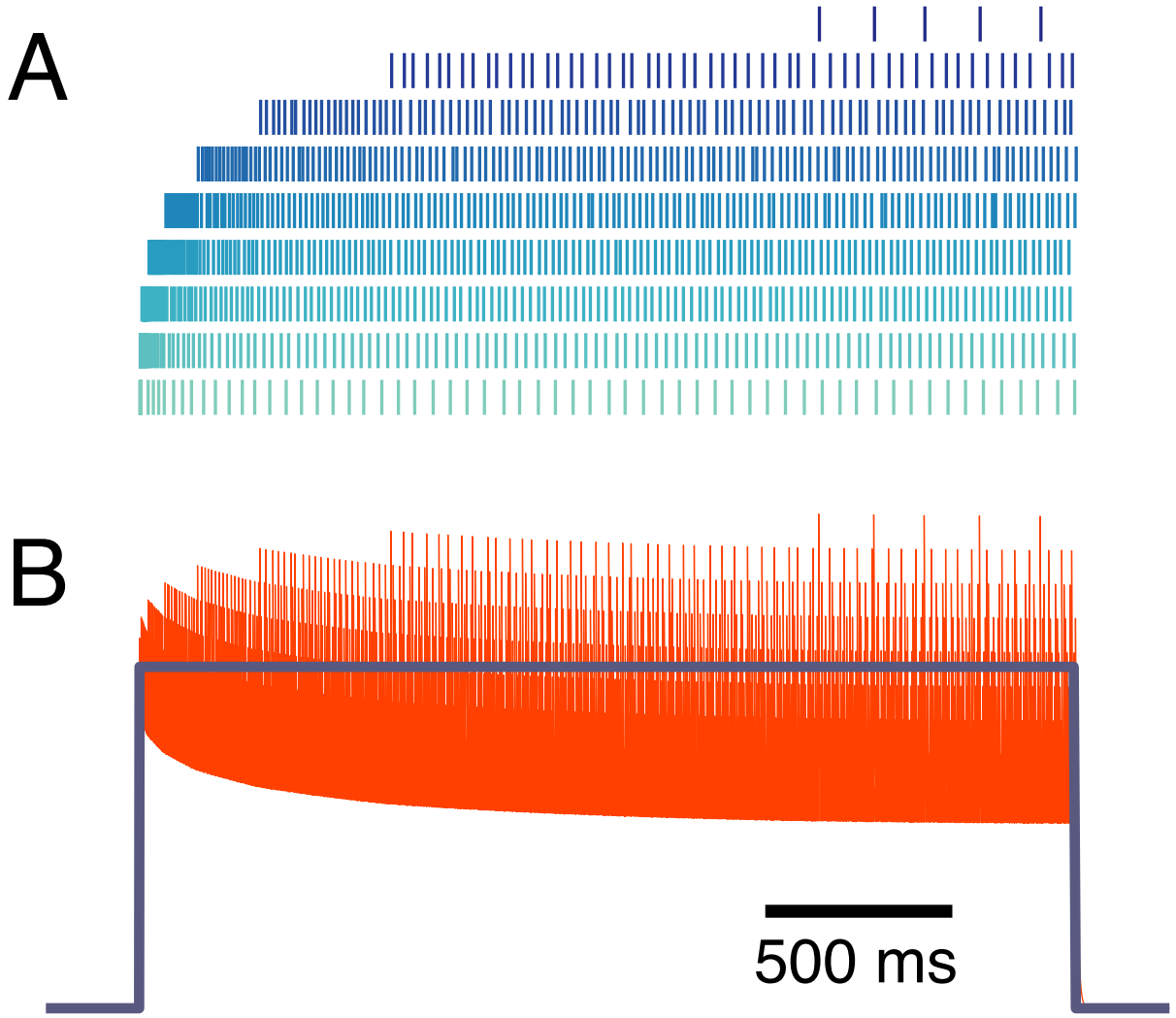


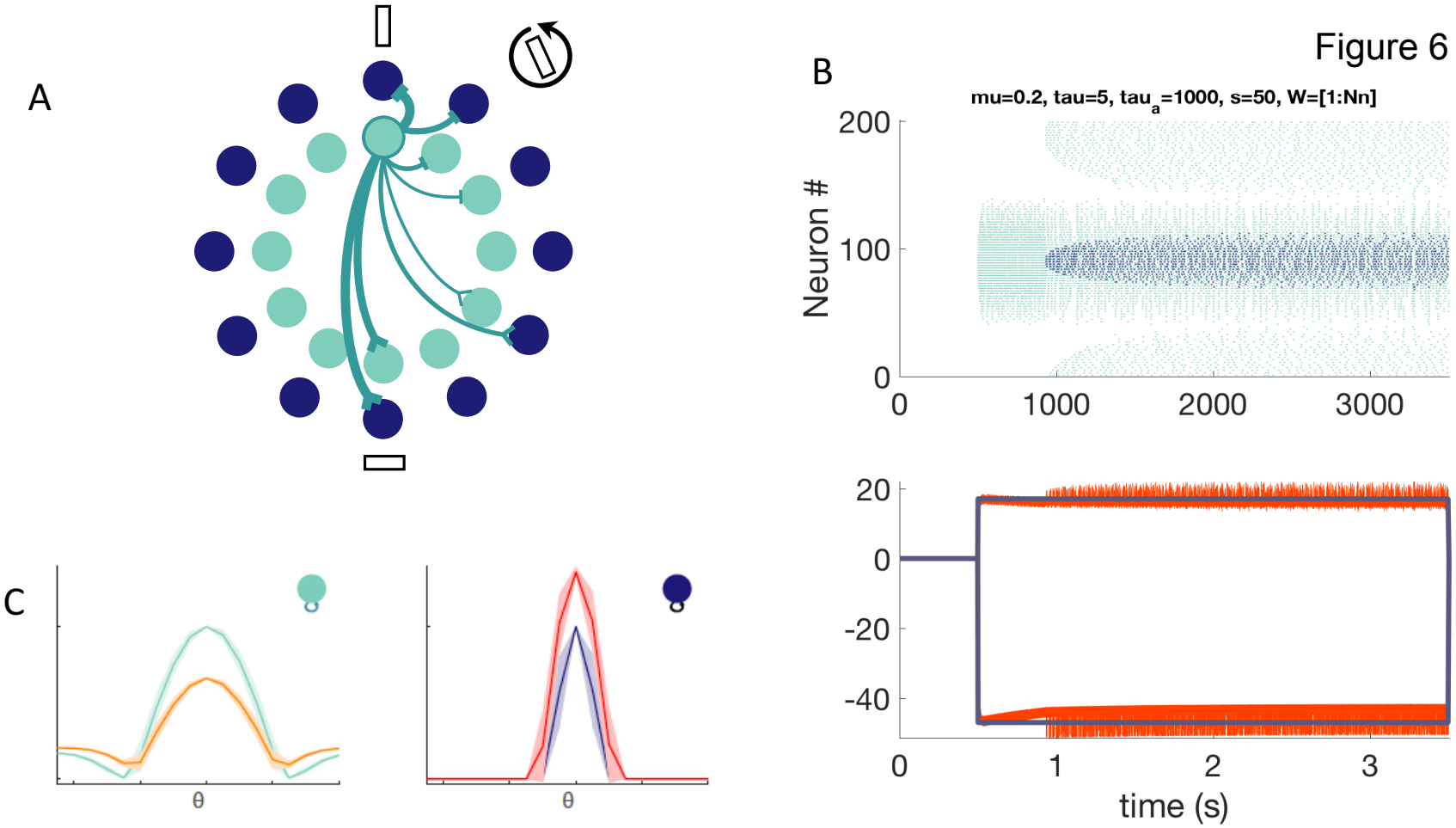
Figure 4

2-neuron fig. (A) Schematic. (B) Top, spikes from neuron 1 (light blue) and neuron 2 (dark blue), middle, respective postsynaptic variables r(t), bottom, stim (grey), estimate (orange), cost (yellow). (C) balanced network follows a manifold. (translucent orange, no recurr, mu = 0.02; dark orange, no recurr, mu=0.4).(D) No recurrent connections. Top, r(t) for successively greater mu with darker shades (mu=0.05,0.1,0.2,0.4). Bottom, estimates and cost with greater mu as color deepens.

Figure 5

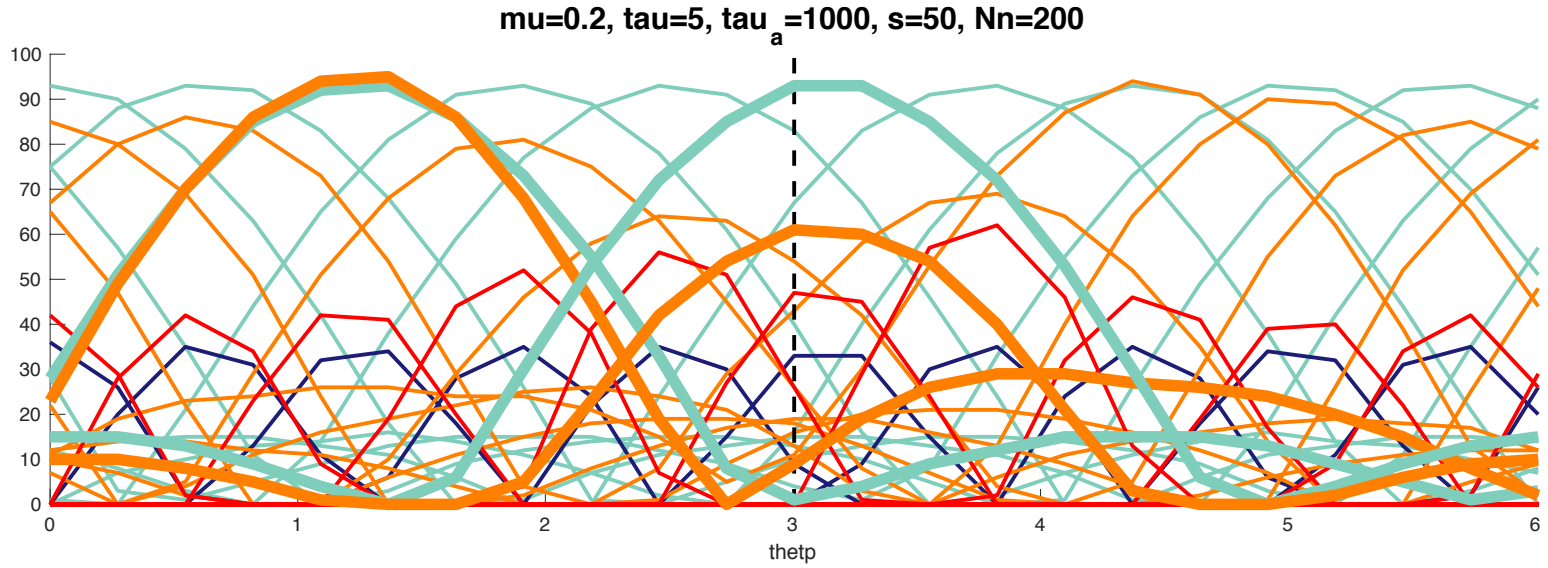


Adapting population of heterogeneous neurons. A, raster of spikes from 10 neuron balanced network in response to a pulse stimulus. B, stimulus (gray) and network estimate (orange).



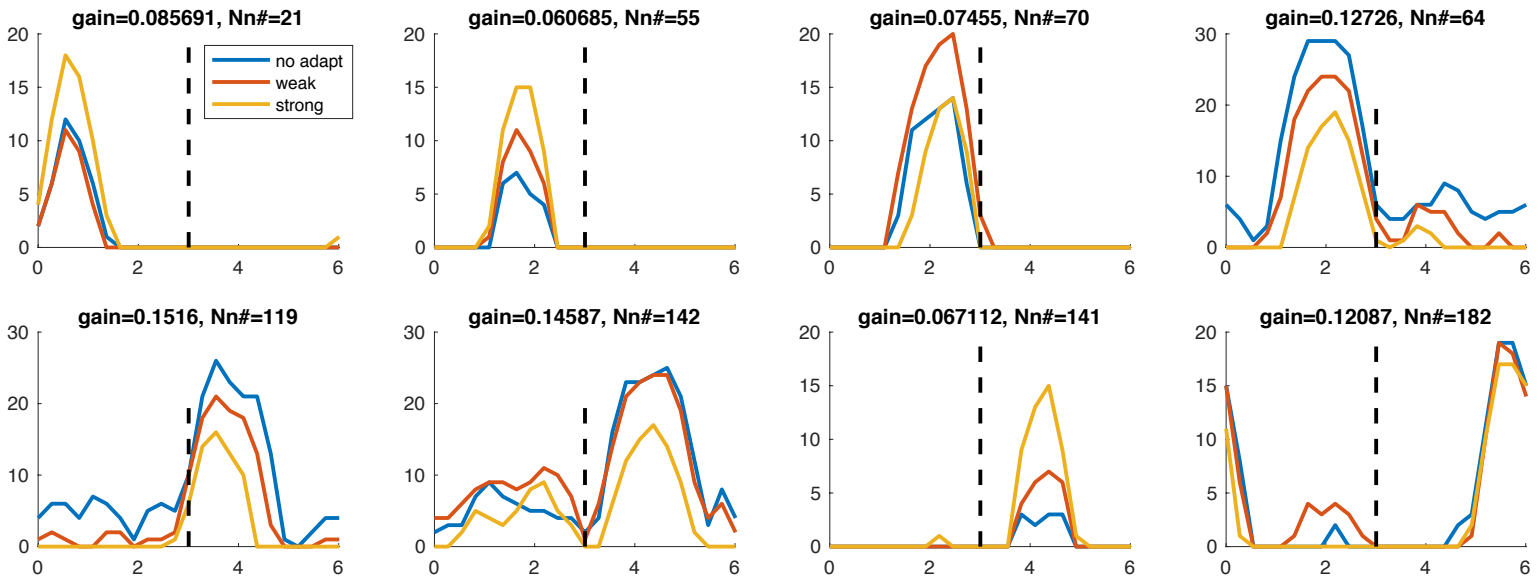
Orientation fig. (a) schematic. (b) network activity showing early and late response to a prolonged stim. (c) tuning curves of high gain and low gain neurons during early response and late part of response to orientation.

Figure 7



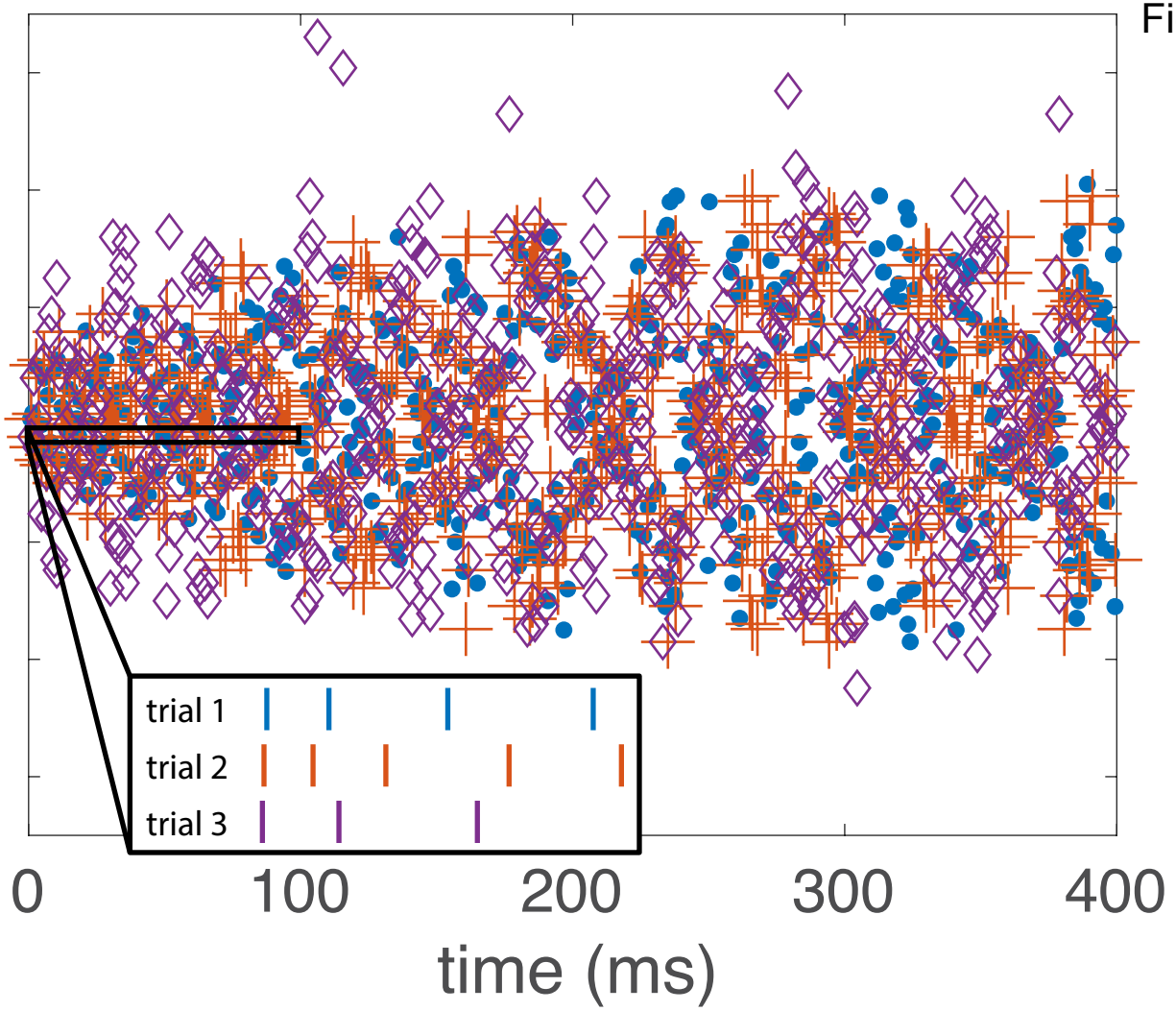
Tuning curves showing neuron responses to a full range of test orientations (x-axis) after adaptation to a single orientation indicated by black dashed line. Light and dark blue curves are those for the high gain and low gain neurons, respectively, in control (i.e. before adaptation). Orange and red curves are high and low gain neurons tuning curves after adaptation to orientation indicated by black line.

Figure 8

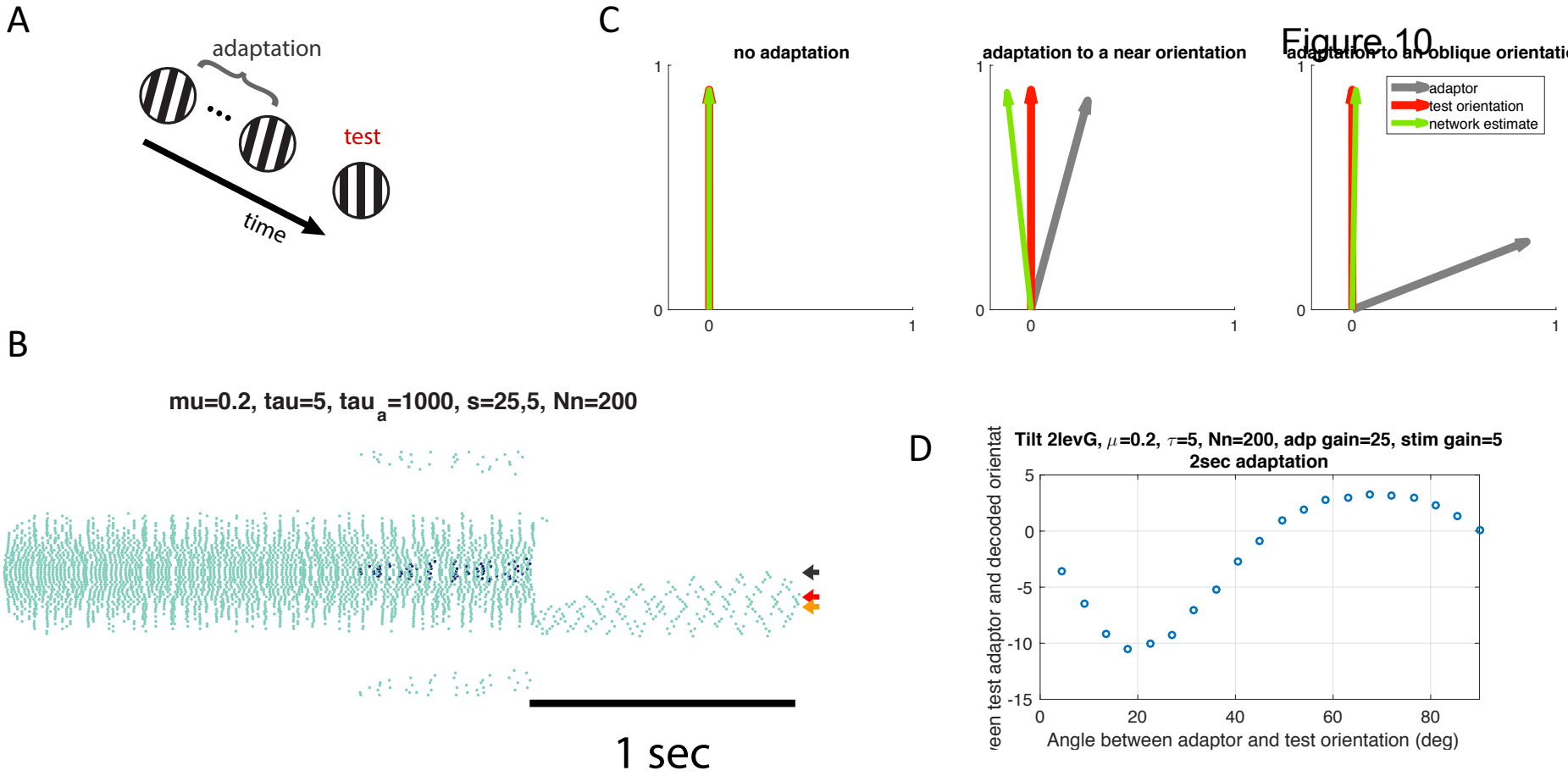


Selected tuning curves from orientation network with random neuron gains. Blue curves, before adaptation; red curves, after weak adaptation; orange curves, after strong adaptation. Some neuron responses are suppressed after adaptation while others are facilitated, and some tuning curves shift laterally after adaptation.

Figure 9



Variable population activity in response to the same test stimulus over 3 trials. Each trial differs in the sequence of stimuli presented before the test stimulus. Popout shows detail over a 100ms window at the onset of the test stimulus for the 3 trials.



(A) Schematic of tilt adaptation protocol. (B) Network activity in response to an adapting stimulus followed by a test stim. Black arrows, neurons that prefer adapting orientation; red arrow, test orientation; orange arrow, decoded orientation. (C) Examples of tilt bias: (left) no bias before adaptation, (middle) network estimate is biased away from test stimulus and adaptor when adaptor is near test orientation, (right) estimate is biased towards adaptor when adaptor is at large angle to test stimulus. (D) estimate bias is repulsive for near adaptation and attractive for oblique adaptation.