Population adaptation in efficient balanced networks.

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

Funding Sources:

Abstract

Adaptation is a key component of efficient coding in sensory neurons. However, it remains unclear how neurons can provide a stable representation of external stimuli given their history-dependent responses. Here we show that a stable representation is maintained if efficiency is optimized by a population of neurons rather than by neurons individually. We show that spike-frequency adaptation and E/I balanced recurrent connectivity emerge as solutions to a global cost-accuracy tradeoff. The network will re-distribute sensory responses from highly excitable neurons to less excitable neurons as the cost of neural activity increases. This does not change the representation at the population level, despite causing dynamic changes in individual neurons. By applying this framework to an orientation coding network, we reconcile neural and behavioral findings. Our approach underscores the common mechanisms behind the diversity of neural adaptation and its role in producing a reliable representation of the stimulus while minimizing metabolic cost.

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 represent sensory inputs whose strength varies by orders of magnitude. Seminal work by Barlow and Laughlin (Barlow, 1961; Laughlin, 1981) demonstrated that sensory neurons in early processing stages adapt their response threshold and gain to 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 has been observed in other early sensory circuits (Blakemore and Campbell, 1969; Fairhall et al., 2001; Solomon and Kohn, 2014; Wark et al., 2007), such as in the retina (Kastner and Baccus, 2014), auditory hair cells (Nagel and Doupe, 2006; Wen et al., 2009) and is also present in later sensory stages (Adibi et al., 2013; Wainwright, 1999). Moreover, Cortical neurons acquire this property during development (Mease et al., 2013).

The work of Laughlin and Barlow was instrumental in uncovering a principle of neural adaption as maximizing information transfer. However, the natural follow-up question concerns the decoding of such neural responses after they have been subject to adaptation. Indeed, such changes in neural gains may result 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 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. 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). Such illusions are classically triggered by long presentations of particularly strong or repetitive stimuli (Maffei et al., 1973). However, adaptation deeply affects neural responses even at short time scales or after only one repetition of the same stimulus (Patterson et al., 2013).

Adaptation could make it impossible to recognize a visual object independently of the stimuli presented previously. An example is given in Figure 1 where we present successive visual patterns to a population of randomly connected leaky integrate-and-fire (LIF) neurons. For simplicity and for the sake of illustration, the network takes a vector of 7 time-varying inputs interpreted as temporal sequences of digital numbers (Fig. 1B, top row). An optimal linear decoder was trained to reconstruct the patterns from the spike counts during the presentation of the patterns. Not surprisingly, the decoder could reconstruct the patterns accurately, regardless of their place in the sequence (Fig. 1B, 2nd row). We then tested the network in the presence of spike-based adaptation in the LIF neurons. We induced spike-based adaptation by temporarily hyperpolarizing the neuron after each spike. The time scale of this adaptation was chosen to be long enough to cover several visual patterns. When subjected to this spike-time dependent adaptation, the responses became strongly history dependent, resulting in a highly inaccurate decoding (Fig. 1B, 3rd row). This would suggest that activity in downstream areas and perceptual interpretations should be based not only on the current sensory responses, but also on the recent history of neural activity. In this study, we show that this is not necessarily the case. Recurrent connections can be tuned such that spike-dependent adaptation will not impair the stability of the representation (Fig. 1B, bottom row).

Results

Neural network solving a global cost-accuracy tradeoff

We will start from an objective function quantifying the efficiency of a population of spiking neurons in representing a time varying sensory stimulus, s(t). We will then show that appropriate recurrent connections between the neurons, namely connections that maintain a tight balance between the excitation and inhibition received by each neuron, will minimize this objective function and thus, maximize the efficiency of the neural code. For the sake of illustration, we hereby assume that the stimulus is unidimensional and positive, as for luminance or color saturation (see methods for

multidimensional stimuli). The stimulus will be decoded from the firing activity of the neurons by summing their responses with their respective readout weights, i.e. $\hat{s}(t) = \sum_i w_i r_i\,(t)$. The neural response r_i is defined as the spike train integrated at a short time scale, $\dot{r}_i = -\frac{1}{\tau} r_i + o_i$, where o_i corresponds to the spike train of neuron i. The readout weight of neuron i is denoted as w_i and it is a fixed parameter. One may choose to include a wide range of readout weights in the network. The output estimate, $\hat{s}(t)$, can be interpreted as a postsynaptic integration of the output spike trains of the population, weighted by synaptic weights w_i .

We wish to construct a network that will minimize the difference between *s(t)* and $\hat{s}(t)$, ensuring an accurate representation of the stimulus. Additionally, we wish to impose, not only accuracy, but also cost efficiency in the neural representation. For biological neurons, spiking comes with inherent metabolic costs. For example, resources are depleted after each spike and neurons or neural populations may need some time to recover from a period of strong activity. Albeit many different types of cost can be incorporated into our approach, here we summarize these constraints as a cost term representing the sum of all squared firing rates. Thus, 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):

$$E = (s - \hat{s})^2 + \mu \sum_i {f_i}^2$$

Where f_i is the firing history of neuron n and the parameter μ weights the relative contribution of error and metabolic costs. The firing history f_i is defined as the spike train integrated with a time constant, τ_a, such that $\dot{f}_i = -\frac{1}{\tau_a} f_i + o_i$. Typically, the adaptation time scale is assumed to be significantly longer than the decoder time scale ($\tau_a >> \tau$). A short τ (e.g. of the order of 10ms) ensures that fast changes in the stimulus can be represented accurately. However, the metabolic cost of spiking accumulates and recovers at slower time scales (e.g. τ_a corresponds to hundreds of ms). The underlying assumption is that the dynamics allowing metabolic resources to be replenished is slower than the time scale at which neural populations transmit information. The sum of squared firing history will encourage, not only small activity at the level of the population, but also small activity in single neurons. As a result, neurons will share the burden of the representation.

From these assumptions, we derive a prescription for the voltage dynamics of LIF neurons performing a greedy minimization of the objective function E (see methods for full derivation):

$$\tau \dot{V}_i = -V_i + g_i w_i (\tau \dot{s} + s) - \tau g_i \sum_j \Omega_{ij} o_j - \kappa_i f_i$$

Where $g_i = 1/(w_i^2 + \mu)$ is the gain of neuron i, the lateral connections are given by $\Omega_{ij} = w_i w_j + \mu\delta_j$, where δ_j is the delta function (equals one only if j=i, zero otherwise) and $\kappa_i = \mu g_i(1 - \frac{\tau}{\tau_a})$. Membrane potentials are normalized such that each LIF neuron has a threshold equal to 1/2 and reset potential equal to -1/2.

The form of the voltage equation is amenable to being interpreted as currents to a neuron embedded in a recurrent network. Neurons are connected by mutually inhibitory synapses determined by their decoding weights. The final term corresponds to an adaptation current that depresses the voltage as a function of its recent activity (see Fig. 2C). This indicates that spike-frequency adaptation in single neurons is part of the solution to the cost-accuracy tradeoff. However, we will show that it cannot work alone; it needs to be associated with appropriately tuned recurrent connections.

It is easier to interpret the network function if we consider that the membrane potentials are effectively proportional to the global coding error penalized by the past activity of the neuron, i.e. $V_i = g_i w_i(s - \hat{s}) - \mu g_i f_i$ (see methods). A neuron that reaches the firing threshold is guaranteed to contribute a decrease of the error term in the objective function (eqn. 1). As a whole, the population performs a greedy minimization of the objective function, or, in other terms, a greedy maximization of the coding efficiency. Finally, we note that since the integrated excitatory input, $g_i w_i s$, is cancelled as precisely as possible (except for the cost penalty) by the recurrent inhibition, $-g_i w_i \hat{s}$, the network can be considered as balancing feedforward excitation and recurrent inhibition. The second ingredient for population efficiency (in addition to spike-based adaptation) is thus to maintain a tight E/I balance in the network. In other words, we show that a memoryless decoder will be able to reconstruct the stimulus from the output spike trains of a population of adapting neurons. This is shown in the bottom row of Figure 1B. Before we investigate the network dynamics and performance, we first describe the properties of single neurons and the relationship between their gain and their coding precision.

Single neuron properties

Let us first consider the case without quadratic cost (i.e. $\mu = 0$). In that case, each neuron effectively has identical voltage and spiking dynamics. Neurons are differentiated only by their gain, g_i, and their decoding weight, w_i (Fig. 2). The strength of the feedforward gain is inversely related to the strength of the output weight for each neuron. As a result, neurons with the smallest decoding weights (and thus, the highest precision in representing the input) tend to respond most strongly to the stimulus (Fig. 2A). We will refer to these costly but reliable neurons as “strongly excitable”. In contrast, neurons with large decoding weights and small input weights (thus “Low gain” neurons) bring

less precision to the estimate but are metabolically cheap. We will refer to these neurons as "weakly excitable".

Note that if μ=0, the cost is not taken into account by the network. Thus, it will always favor precision over cost. In that case, only the most excitable neuron (with the smallest decoding weight) will respond to the stimulus while completely inhibiting the other neurons. However, with the addition of a cost (μ>0), adaptive currents contribute to the voltage dynamics, penalizing neurons with large firing rates. Moreover, the ratio of excitability, $g_i\, w_i = \frac{w_i}{|w_i|^2 + \mu}$, is not necessarily decreasing monotonically with the decoding weights (Fig. 2B). For very small decoding weights, $|w_i|^2 \ll \mu$, thus neurons with decoding weights smaller than $\sqrt{\mu}$ are penalized. These neurons would simply be too costly to participate meaningfully in the cost/accuracy tradeoff solved by the population.

Model neurons in isolation (without contribution from recurrent connections) would respond to a step-like input with a rate that decreases exponentially with time before reaching a plateau (Fig. 2C), a classic signature of activity-dependent suppression. The time constant of this adaptation is determined by τ_a, while strength of this adaptation increases with the gain. Highly excitable neurons adapt strongly, while less excitable neurons adapt weakly.

However, these intrinsic properties of single neurons will be deeply affected by the dynamic introduced by recurrent connections. To gain a better understanding of population adaptation, we will investigate how inhibitory connections orchestrate the relative contribution of different neurons over the duration of a long stimulus.

Network activity is distributed on a manifold in neural activity space

We first illustrate the effect of recurrent connections with an example network composed of only two neurons (Figure 3). The two neurons are reciprocally connected with an inhibitory connection, as prescribed in the derivation (schematized in Fig. 3A). They receive a constant stimulus, but have different input weights.

During sustained stimulation, the response of each neuron fluctuates dynamically despite the fact that the stimulus is constant (Fig. 3B). This would be expected given their spike-time dependent adaptation. However, if one removes the recurrent connections and plots the response of one neuron as a function of the other (Fig. 3C, right), we discover that the population response wanders from the iso-coding line $s = w_1r_1 + w_2r_2$ (i.e. the manifold in activity space were the stimulus would be decoded properly). In contrast, the intact network with its recurrent connections coordinates the two neurons

such that the weighted sum of their responses remains accurate. The movement of the activity along the manifold defined by the constant stimulus and the decoding weights reflects a progressive redistribution of activity to satisfy the unfolding cost-accuracy tradeoff, as the cost slowly accumulates (Fig. 3D).

While to a naïve observer, the first neuron may appear to adapt while the second neuron has a sustained response and a longer delay, in fact both contribute to population adaptation, i.e. to limit the metabolic cost of the representation while maintaining its accuracy. Recurrent connections deeply affect the dynamics of each neuron. For example, the inhibition from the more excitable neuron is responsible for the response delay of the other neuron.

Coordinated adaptation in a neural population

Within a network with many neurons (Fig. 4), recurrent connections interact with the intrinsic properties of the neurons in a similar manner as in the previous example. The first neurons to be recruited are strongly excitable and provide an initially very precise representation of the signal. These neurons inhibit the less excitable neurons, preventing them from firing early in this stimulation period. As the cost accumulates, however, the response of the high gain neurons decays due to spike-frequency adaptation. This is compensated by less excitable neurons who become disinhibited, fire, and then adapt in their turn. The less excitable a neuron is, the later it will be recruited, resulting in strong response delays. The dynamic response properties of individual neurons are thus dominated by network interactions and are markedly different from their intrinsic adaptive properties (Fig. 2).

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. 4D). However, note that its precision degrades as more low gain neurons contribute to the representation. As a result, the bias and standard deviation of the representation increases as imposed by the global cost/accuracy tradeoff.

Coordinated adaptation of tuning curves

To illustrate what coordinated adaptation would predict for tuning curves measured experimentally, we constructed a population of neurons that code for visual orientation, i.e. V1 simple cells. The input to the network takes the form of a two-dimensional signal with a cosine and a sine of the presented orientation (see methods). Each neuron has a preferred orientation that is given by the

combination of input weight strengths in the two input dimensions. In turn, the network orientation estimate can be decoded with a population vector (see methods).

In contrast to the previous model, the lateral connections derived from the model maximally inhibit neurons with similar preferred orientation and excite neurons with orthogonal orientations (see schematic in Fig. 5A) due to the choice of decoding weights which can be positive and negative. To observe the effects of adaptation on a diverse population, we constructed our network so that neurons have equally spaced preferred orientations and a partner neuron that shares the same preference but has a different gain. There is a high gain and a low gain neuron among each pair of neurons that code for the same orientation.

Figure 5B illustrates the spiking response of the network to a prolonged oriented stimulus. As seen in the simpler model from Figure 4, high gain neurons respond first, then adapt. As the responses of highly excitable neurons decay, less excitable neurons are recruited, maintaining the representation. This results in systematic changes in the tuning curves from early in the response to later in the response Fig. 6. Highly excitable neurons are suppressed relative to their early responses (Fig. 6, top). In contrast, weakly excitable neurons see their tuning curves widen when the adapting stimulus is similar to their preferred orientation (Fig. 6, bottom). At the flank of the adapting orientation, low gain neurons see an increase in their responsiveness. Here, the network interactions override the intrinsic adapting currents in the weakly excitable neural population. In other words, the disinhibition from excitable neurons combined with the constant feedforward drive to these low gain neurons results in facilitated activity rather than the suppressed activity one would expect to be caused by adaptation. Finally the tuning curves for the most excitable neurons are broader than those for weakly excitable neurons. These neurons are more likely to fire first in response to oriented stimuli that are near their preferred orientation and prevent the low gain neurons from doing the same.

Heterogeneity in a more plausible model

The same qualitative effects are observed in a more realistic network where the preferred orientation of different neurons and their decoding weights were taken from a random distribution (Fig. 7), rather than regularly spaced with two levels of excitability. The tuning curves are more heterogeneous not because of noise but because of the randomness of the decoding weights. Tuning curves can be either facilitated or suppressed by adaptation. When the adapted stimulus falls on the flank of the tuning curve, it can be accompanied by a shift toward or away from the adapting stimulus. The effect of adaptation on single neurons is variable not because of noise (we did not introduce any) but because of local heterogeneity in the competition they receive from other neurons, itself due to

the random choices of weights and gains. In fact, adaptation in one neuron would be impossible to predict quantitatively without observing the rest of the network.

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 is designed to negotiate the tradeoff between accuracy and efficiency and it will prioritize the production of a stable representation if μ is small. If μ is large, the network will favor cost over accuracy. Thus, strong adaptation and a prolonged stimulus presentation can produce a representation that degrades over time. This degradation can lead to a bias in the decoder. In Figure 8, an oriented, strong, adapting stimulus is presented for 2 seconds followed by a test orientation (this is schematized in Figure 8A). An example of the resulting network activity is shown in Figure 8B. Before adaptation takes hold, the adapting stimulus activates the high gain neurons that have preferences at or 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. 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 test 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 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 often repulsed away from the adapted orientation, the effect being maximal for adapting stimuli tilted around 15-20 degrees from the test stimulus. If the adapting stimulus is oriented around 60 degrees from the adaptor, a repulsive effect is observed instead. This effect has been confirmed in the visual cortex (Jin et al., 2005). Our model replicates this effect (Fig. 8C,D). The test stimulus is decoded at an orientation that is repulsed from its actual orientation away from the adaptor when the adaptor is approximately 15 degrees from vertical (Fig. 8D, middle). However, when the adaptor is obliquely oriented from the test orientation, the test stimulus is perceived to be oriented in a direction that is

attracted to the adaptor (Fig. 8D, right). 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. 8D). In accordance with experimental findings, the repulsion effect has a greater amplitude than the attraction effect.

Discussion

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 solve a global cost/accuracy tradeoff rather than a local tradeoff at the level of each neuron. In our 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 (such as 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.

Another problem arises in the trial-to-trial variability produced by adaptation that is observed at the single neuron level. As mentioned in the introduction, the history-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 (Benucci et al., 2009) 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. Instead of updating the weights to better represent stimuli over several iterations, as is done for the perceptron and convolutional neural networks (Fukushima, 1980; LeCun et al., 1999; Olshausen and

Field, 1996; Rosenblatt, 1958), we derive a prescription for the voltage dynamics so that the network neurons can produce a reconstruction of any stimulus with fixed decoder and recurrent weights.

Our model is developed from a normative encoding framework (Boerlin et al., 2013; Druckmann and Chklovskii, 2012; Olshausen and Field, 1996; Spratling, 2010) in which we enforce efficiency in the encoder and accuracy in the decoder. The new contribution compared to (Boerlin et al., 2013) is to allow more flexibility in the form that the metabolic cost can take. In particular, the time scales of the cost and of the representation are disassociated which leads to distinct dynamics for the cost and for the neural firing rate. 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. The variability of adaptation effects has been observed in V1 neurons (Jeyabalaratnam et al., 2013). 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.

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; Hong et al., 2008; Hosoya et al., 2005; Nagel and Doupe, 2006; Smirnakis et al., 1997; Solomon and Kohn, 2014; Wark et al., 2007; 2009). 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 (Fairhall, 2014). 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 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.

Validating the framework experimentally.

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 (Buzsáki, 2004), bringing an experimental validation of this framework within grasp. These recordings 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 the 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 are 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 population. This could be done optogenetically, for example (Okun et al., 2015).

METHODS

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 the threshold was $1 + \mu f_i(t)$, where $\dot{f}_i(t) = -\frac{1}{\tau_a} f_i(t) + o_i(t)$. For the example of the balanced network with adaptation, the network was derived using the framework described below [$\tau_a = 2000ms, \mu = 0.02, \tau = 5ms$].

Network model

We provide here a brief description of the network structure and the objective function it minimizes. A detailed version of this derivation is found in (Boerlin et al., 2013). In contrast to the study, our derivation utilizes distinct variables for the neural response and the spike history. 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 w_i, where r_i is the filtered spike train of neuron *i*. The filtered spike train corresponds to a leaky integration of its spikes, , with $o_i(t) = \sum_k \delta(t - t_i^k)$, with t_i^k the spike time of the kth spike in neuron i 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 chosen 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 coding errors, and the other penalizing firing rates:

$$E(t) = ||s(t) - \hat{s}(t)||^2 + \mu \sum_i r_i^2 \qquad \text{eqn. 1}$$

The firing rates, f, are defined as an integration of the spike trains at a slow time scale, with $\tau_a > \tau$, 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 i 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 i spikes if

$$g_i(\sum_j w_{ij}(s_j(t) - \hat{s}_j(t)) - \mu f_i) > \frac{1}{2} \qquad \text{eqn. 2}$$

With g_i being the "gain" of neuron i. We interpret the left-hand side of this equation as the membrane potential of neuron i, 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.

$$\tau \dot{V}_i = -V_i + g_i w_i(\tau \dot{s} + s) - \kappa f_i - \tau g_i \sum_j \sum_k w_{ij} w_{jk} o_k - \mu \tau g_i o_i \qquad \text{eqn. 3}$$

The lateral connection between neuron i and neuron k is equal to $\tau g_i \sum_j w_{ij} w_{jk}$. 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.

Orientation model

The orientation-coding network follows the same derivation as outlined for the generic network model. It has two input dimensions and 200 neurons. There are two subpopulations of neurons such that 100 neurons are high gain neurons with a feedforward weight of 3 and the remaining 100 neurons are low gain neurons with a feedforward weight of 9. Both populations span the unit circle evenly such that one low gain and one high gain neuron share the same preferred orientation.

More precisely, we endowed each neuron with a decoding vector $[w_i^1, w_i^2] = g_i[cos(\theta_i), sin(\theta_i)]$. Feedforward inputs corresponded to two time-varying inputs $s_1 = Csin(\theta), s_2 = Ccos(\theta)$, where C is the stimulus magnitude and theta is the stimulus orientation. The orientation estimate is decoded from the population as $\hat{\theta} = arctan(1/2 * \hat{s}_1/\hat{s}_2)$. The factor of 1/2 in front of theta represents the fact that LGN inputs are not directionally tuned (i.e. inputs for orientation 0 degree = 180 degree). The spiking threshold included an additional term, ηw_i^2, that ensured that neurons of opposing preferences would not be activated to spike so easily by the excitation from opposing neurons.

Tuning curves in Figure 6 were generated by presenting the network with a full range of stimulus orientations of (stimulus magnitude = 50). Neuron responses were centered on their preferred orientation and the mean was taken for each subpopulation. 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 was generated by presenting the network with an adaptor orientation (duration = 2s) and a subsequent test orientation (250ms). The perceived angle was decoded from the mean network output over the 250ms presentation of the test stimulus. 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.

Benucci, A., Ringach, D.L., Carandini, M., 2009. Coding of stimulus sequences by population responses in visual cortex. Nat Neurosci 12, 1317–1324. doi:10.1038/nn.2398

Blakemore, C., Campbell, F.W., 1969. On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. J. Physiol. (Lond.) 203, 237–260. doi:10.1113/jphysiol.1969.sp008862

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

Buzsáki, G., 2004. Large-scale recording of neuronal ensembles. Nat Neurosci 7, 446–451. doi:10.1038/nn1233

Druckmann, S., Chklovskii, D.B., 2012. Neuronal circuits underlying persistent representations despite time varying activity. Curr. Biol. 22, 2095–2103. doi:10.1016/j.cub.2012.08.058

Fairhall, A., 2014. The receptive field is dead. Long live the receptive field? Current Opinion in Neurobiology 25, ix–xii. doi:10.1016/j.conb.2014.02.001

Fairhall, A.L., Lewen, G.D., Bialek, W., van Steveninck, R., 2001. Efficiency and ambiguity in an adaptive neural code. Nature 412, 787–792. doi:10.1038/35090500

Fukushima, K., 1980. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern 36, 193–202.

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

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

Hong, S., Lundstrom, B.N., Fairhall, A.L., 2008. Intrinsic Gain Modulation and Adaptive Neural Coding. PLoS Comput Biol 4. doi:10.1371/journal.pcbi.1000119

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
LeCun, Y., Haffner, P., Bottou, L., Bengio, Y., 1999. Object Recognition with Gradient-Based Learning, in: Shape, Contour and Grouping in Computer Vision, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, Berlin, Heidelberg, pp. 319–345. doi:10.1007/3-540-46805-6_19
Maffei, L., Fiorentini, A., Bisti, S., 1973. Neural correlate of perceptual adaptation to gratings. Science 182, 1036–1038.
Mease, R.A., Famulare, M., Gjorgjieva, J., Moody, W.J., Fairhall, A.L., 2013. Emergence of adaptive computation by single neurons in the developing cortex. J. Neurosci. 33, 12154–12170. doi:10.1523/JNEUROSCI.3263-12.2013
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
Okun, M., Steinmetz, N.A., Cossell, L., Iacaruso, M.F., Ko, H., Bartho, P., Moore, T., Hofer, S.B., Mrsic-Flogel, T.D., Carandini, M., Harris, K.D., 2015. Diverse coupling of neurons to populations in sensory cortex. Nature 521, 511–515. doi:10.1038/nature14273
Olshausen, B.A., Field, D.J., 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381, 607–609. doi:10.1038/381607a0
Patterson, C.A., Wissig, S.C., Kohn, A., 2013. Distinct effects of brief and prolonged adaptation on orientation tuning in primary visual cortex. J. Neurosci. 33, 532–543. doi:10.1523/JNEUROSCI.3345-12.2013
Rosenblatt, F., 1958. The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review 65, 386–408. doi:10.1037/h0042519
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., Fairhall, A., Rieke, F., 2009. Timescales of Inference in Visual Adaptation. Neuron 61, 750–761. doi:10.1016/j.neuron.2009.01.019

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

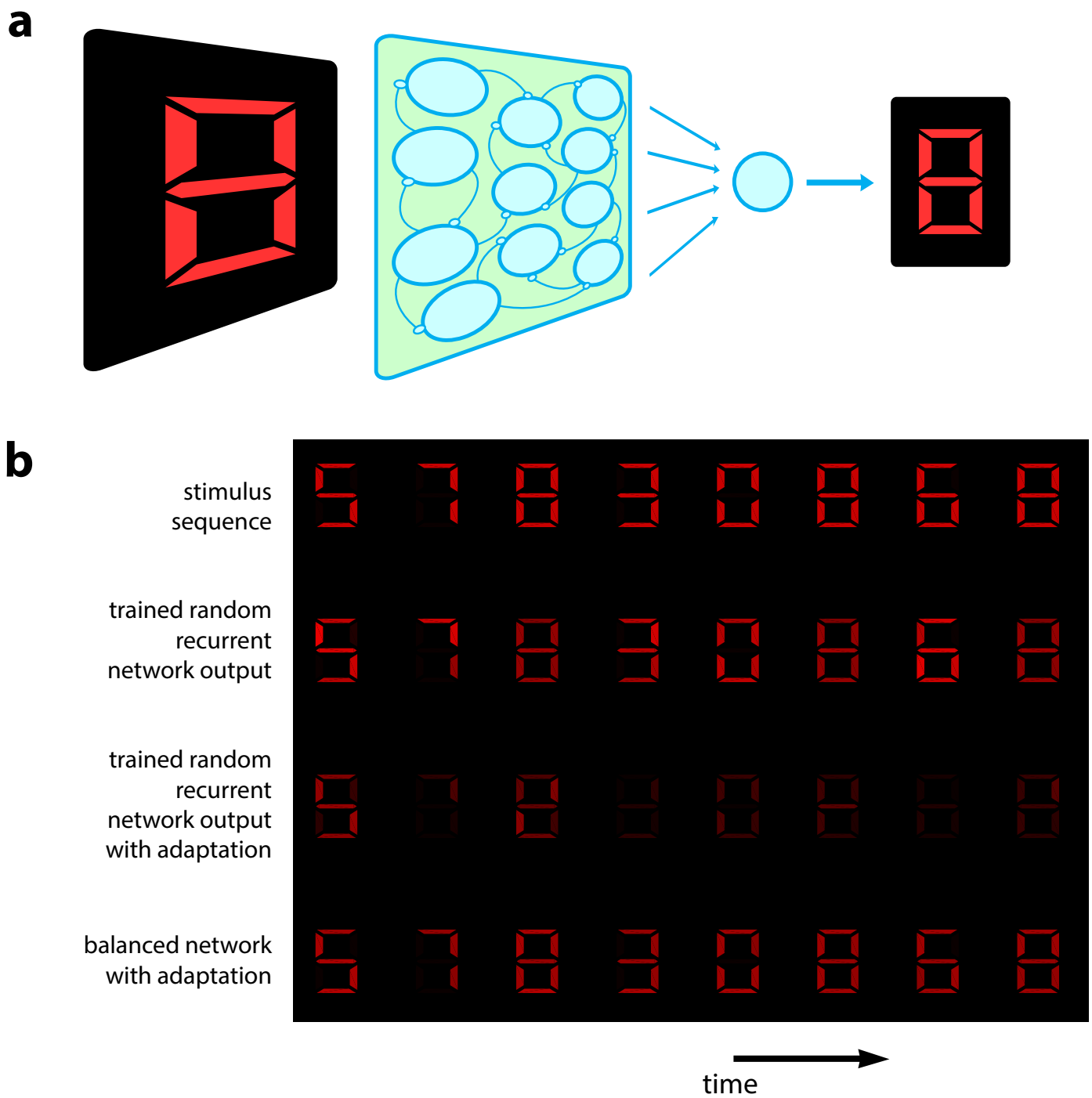

Figure 1:
(a) Schematic of a 7-dimensional input (one dimension for each bar position of a digital interface) being presented to a random recurrent network that sends input to a readout layer (here represented by a single neuron).
(b) Top, a sequence of digits that serve as stimuli (presented for 200ms each, spaced by 100ms between digits). Second row, decoded output of random recurrent network with optimal decoder (trained on 100 samples of completely random patterns). Third row, decoded output of same random recurrent network as above but with adapting neuron responses (mu = 0.02; tau_a = 2000ms for the adaptive firing rates). Bottom row, balanced network with adaptation derived from efficient coding framework. [400 neurons, tau=5ms for neural responses integrated by decoder].

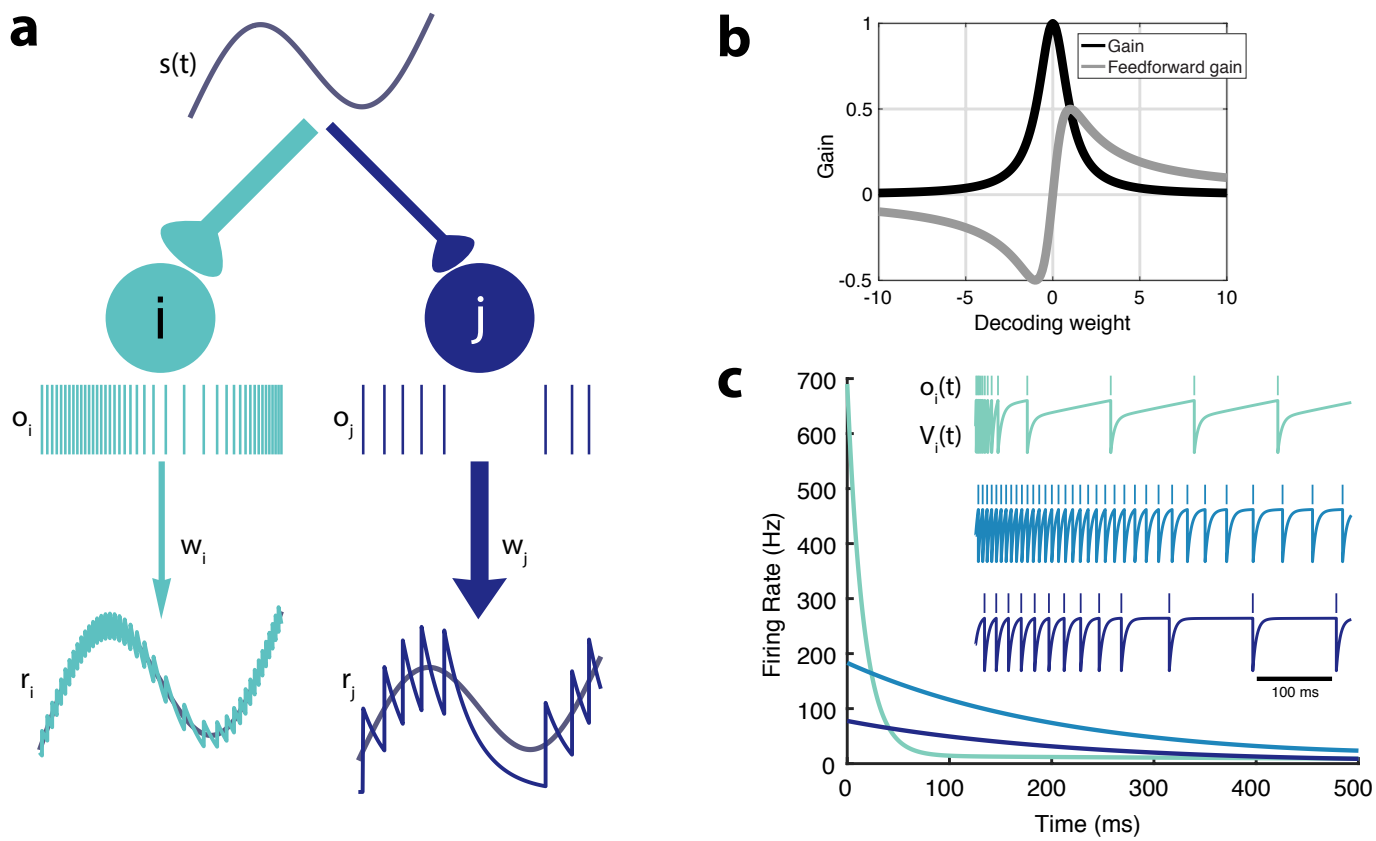

Figure 2:
Intrinsic model neuron properties.
(a) High gain neurons (light blue) are intrinsically excitable and due to their small decoding weight they are precise while low gain neurons (dark blue) are less excitable but also less precise. An arbitrary input, s(t), elicits distinct responses from the two neurons (spikes train o_i and o_j, respectively). Neurons send a filtered response, r_i , r_j, to the decoder weighted by w_i and w_j, respectively.
(b) Relationship between gain, $g_i = 1/(w_i^2 + \mu)$, feedforward gain, $g_i = w_i/(w_i^2 + \mu)$, and decoding weight, w_i (μ=1).
(c) Different gains give neurons distinct adaptation dynamics. Instantaneous spiking rates, o(t), are plotted over time for three model neurons with different decoding weights (light blue, w=1; medium blue, w=5; dark blue, w=9). High gain neurons have the steepest adaptation (light blue) whereas low gain neurons (dark blue) do not adapt as rapidly given the same input. Inset shows the voltage trace, V(t), and spike train, o(t), for each example neuron.

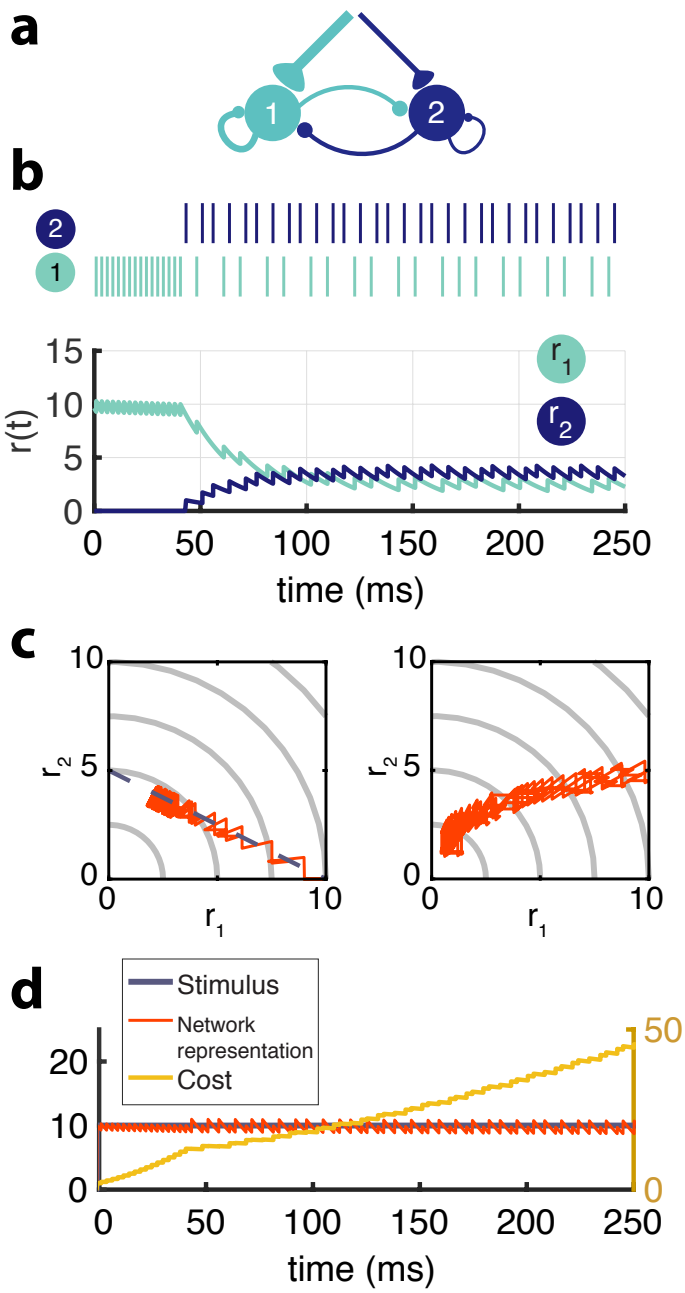

Figure 3:
Two-neuron network.
(a) Schematic of recurrently connected two-neuron network derived from efficient coding framework. Neuron 1 is strongly excitable ($w_i = 1$), while neuron 2 is less excitable ($w_i = 2$).
(b) Spikes from neuron 1 (light blue) and neuron 2 (dark blue) show the transient response of the strongly excitable neuron and the delayed, but sustained response of the weakly excitable neuron (top). Postsynaptic activity, r(t) (bottom) (μ=0.02, τ = 25ms, τ_a = 1000ms, stim = 10).
(c) The balanced network with adaptation follows a linear manifold (left), whereas the network without recurrent connections but with adaptation cannot be linearly decoded (right).
(d) The cost ($\mu\Sigma_n f_n^2$, yellow) accumulates steeply until neuron 1 adapts and neuron 2 is recruited and the cost increases at a slower rate. The network representation (orange) is maintained despite the redistribution of activity among the neurons.

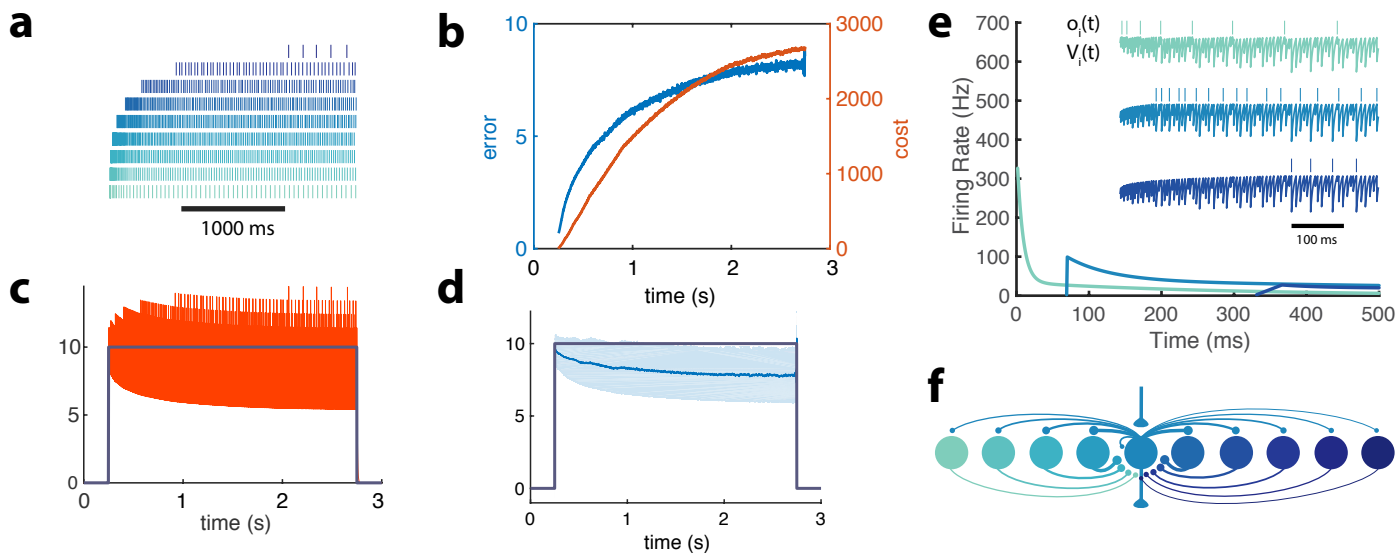

Figure 4:
Adapting population of heterogeneous neurons.
(a) Spike raster of all 10 neurons in a balanced network with adaptation in response to a pulse stimulus (μ=0.2, τ=5, τ_a=1000, s(t)=10, W=[1:10]). Neurons are ordered from weakly excitable (top, dark blue) to highly excitable (bottom, light blue).
(b) Both the error ($||s(t)-\hat{s}(t)||^2$, blue) and cost ($\mu\sum_n f_n^2$, orange) accumulate over time.
(c) The network estimate (s (t), orange) tracks the stimulus (s(t), gray) with increasing variance.
(d) The smoothed network estimate (blue line) shows a biased estimate with increasing variance (blue shade, standard deviation).
(e) Instantaneous spiking rates of 3 example neurons in the network (same neurons as in Fig. 2). Inset shows the voltage trace, V(t), and spike train, o(t), for each example neuron.
(f) Schematic of 10-neuron balanced network showing only connections to and from the middle neuron. Excitatory connections are shown as triangles and in this particular network are only found in the feedforward input and the output. Inhibitory connections are shown with small circles and make up only the recurrent connections.

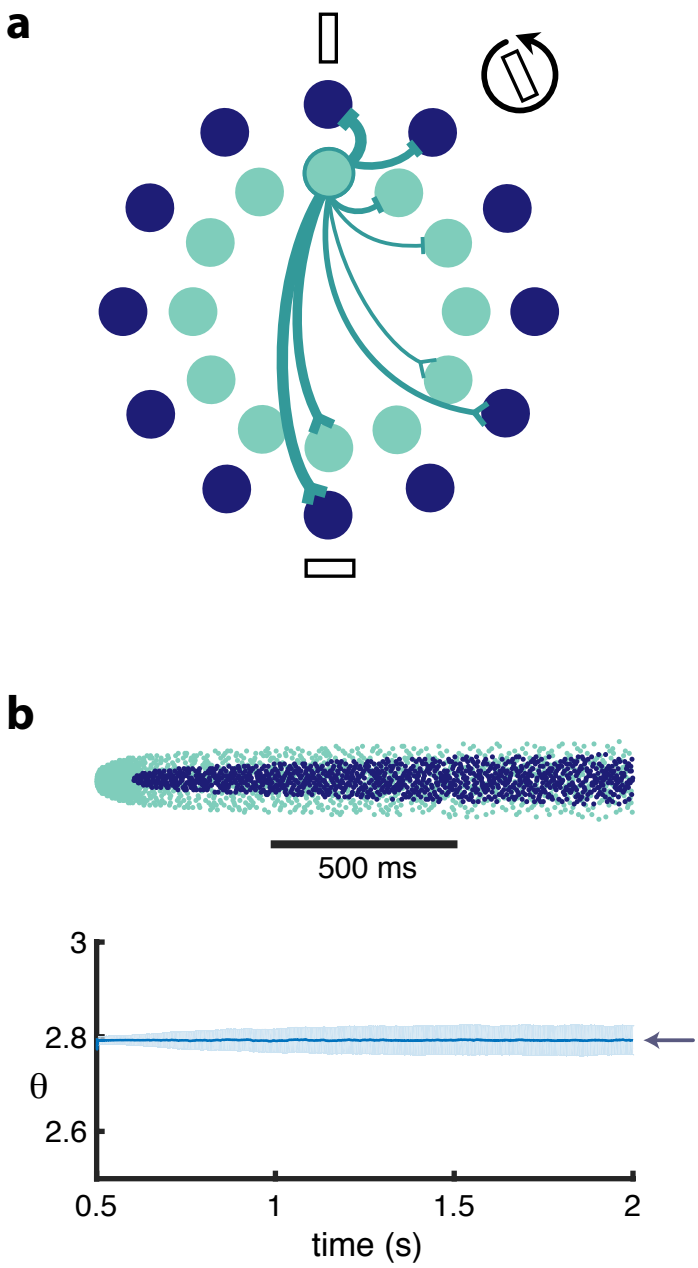

Figure 5:
Orientation coding-network.
(a) Schematic showing the dual-ring structure of the network of high gain (light blue) and low gain (dark blue) neurons. Some of the recurrent connections from the highlighted light blue neuron are illustrated to show that a neuron inhibits its neighbors most strongly and excites neurons with opposing preferences (inhibitory connections are shown as slashes, excitatory connections are shown as chevrons).
(b) Spike raster (top) of population activity showing the evolution of the population response during a prolonged stimulus presentation of a constant orientation. The decoded orientation is steady while the variance increases over time (bottom). Arrow indicates the stimulus orientation.

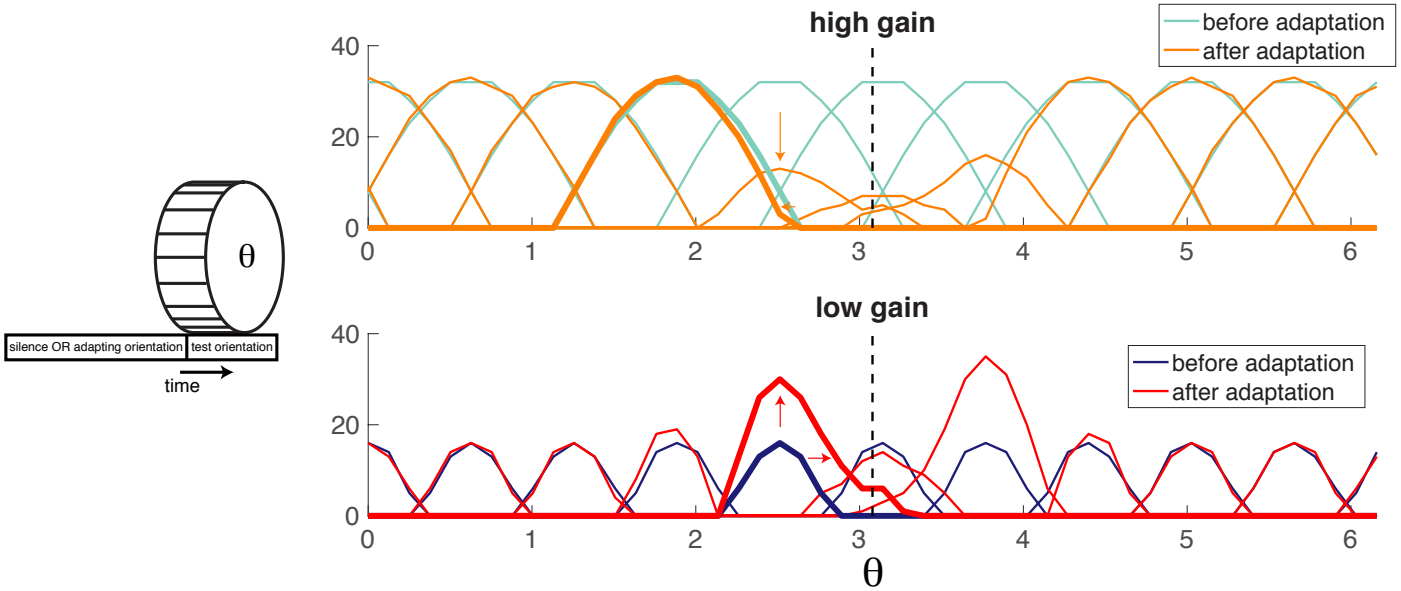

Figure 6:
Population adaptation tuning curves show neuron responses to a full range of test orientations (x-axis) after adaptation to a single orientation (black dashed line). Top, tuning curves for strongly excitable neurons before adaptation (light blue) are broad. After adaptation (orange), tuning curves near the adaptor are suppressed. Bottom, tuning curves for weakly excitable neurons before adaptation (dark blue) show less activation than for strongly excitable neurons and more specific tuning. After adaptation (red), flanking curves are facilitated and shifted toward adaptor. Schematic (left) illustrates how tuning curves were calculated. (μ =0.1, τ=5ms, τ_a=2000ms, s(t)=50, 200 neurons).

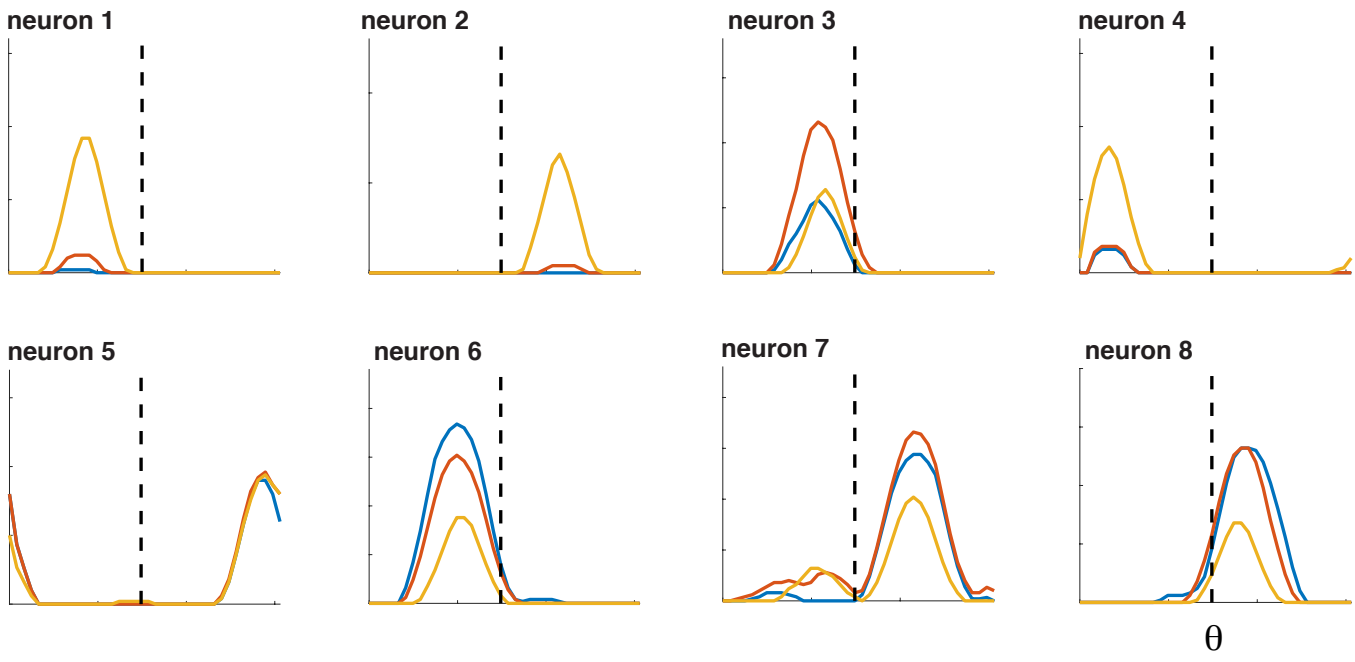

Figure 7:
Selected tuning curves from orientation network with random neuron gains shown top left to bottom right in order of increasing gain (from g=0.067 to 0.127). Blue curves, before adaptation; red curves, after weak adaptation; orange curves, after strong adaptation. Some neuron responses are suppressed after adaptation while others are facilitated, and some tuning curves shift laterally after adaptation.

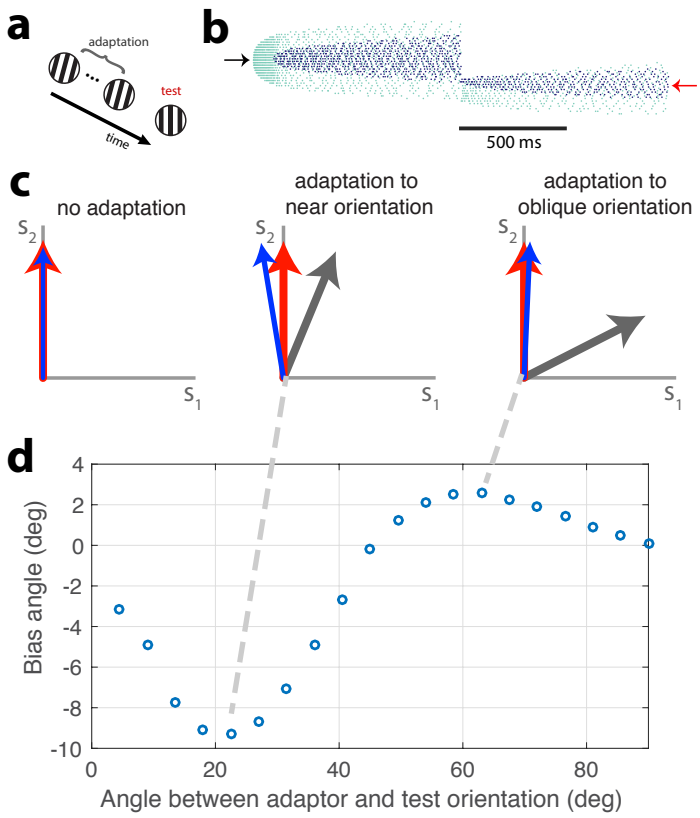

Figure 8:
Tilt illusion.
(a) Schematic of tilt adaptation protocol.
(b) Network activity in response to an adapting stimulus followed by a test stim. Rasters are ordered by neurons' orientation preference. Black arrow, neurons that prefer adapting orientation; red arrow, neurons that prefer test orientation ($\mu = 0.1$, $\tau =$ 5ms, $\tau_a = 2000$ms, 200 neurons, adaptor stim = 50, test stim = 25).
(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. Adaptor is presented for 2 seconds and test orientation is presented for 250ms.