UNIVERSITY OF CALIFORNIA, SAN DIEGO

Information Maximization in Early Sensory Systems

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Physics

by

Yilun Zhang

Committee in charge:

Professor Tatyana Sharpee, Chair
Professor Massimo Vergassola, Co-Chair
Professor Julius Kuti
Professor Terrence Sejnowski
Professor Charles Stevens

2018

The dissertation of Yilun Zhang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____
Co-Chair

_____
Chair

University of California, San Diego

2018

DEDICATION

To life, the universe and everything.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

Thanks to my advisor Tatyana Sharpee, who has been mentoring and supporting me, for serious or fun projects, in-office or remote works, academic or non-academic experience.

Thanks to my family, especially my parents, who have always been loving and supporting me.

Thanks to my future family, who has been loving and supporting me, and gave me chances to come up with ideas on the planes.

Thanks to my collaborators David Kastner and Stephen Baccus, for insightful suggestions and precious data.

Thanks to my amazing labmates and colleagues, for the countless discussions that inspired me.

Thanks to my committee members, for the discussions, suggestions and mentorship that showed me the directions.

Thanks to my friends, for making San Diego a wonderful place to study and live.

Thanks to UCSD and Salk, for providing professional studying and working environment.

Thanks to all my funding supports.

My PhD life wouldn't be better.

Chapter 2, in part, is a preprint of the material as it will appear in 52th Annual Conference on Information Sciences and Systems (CISS). Zhang, Yilun; Kastner, David; Baccus, Stephen; Sharpee, Tatyana, 2018. The dissertation author was the primary investigator and author of this material.

Chapter 3, in full, is a preprint of the material as it will appear in 52th Annual Conference on Information Sciences and Systems (CISS). Zhang, Yilun; Kastner, David; Baccus, Stephen; Sharpee, Tatyana, 2018. The dissertation author was the primary investigator and author of this material.

Chapter 4, in part, is currently being prepared for submission for publication of the

material. Zhang, Yilun; Sharpee, Tatyana. The dissertation author was the primary investigator and author of this material.

Chapter 5, in full, is a reprint of the material as it appears in PLoS computational biology. Zhang, Yilun; Sharpee, Tatyana, 2016. The dissertation author was the primary investigator and author of this paper.

VITA

2013	B. S. in Physics, Peking University, Peking, China
2013-2014	Teaching Assistant, University of California, San Diego
2014	M. S. in Physics, University of California, San Diego
2014-2018	Research Assistant, University of California, San Diego
2018	Ph. D. in Physics, University of California, San Diego

PUBLICATIONS

Yilun Zhang, Tatyana Sharpee, "Optimal Nonlinearity for Neurons Receiving Temporal Correlated Signal", 2018, in preparation.

Yilun Zhang, David Kastner, Stephen Baccus, Tatyana Sharpee, "Optimal Information Transmission by Multiple Mosaics of Retinal Cells", *Conference on Information Sciences and Systems*, 2018

Yilun Zhang, Tatyana Sharpee, "A Robust Feedforward Model of the Olfactory System", *PLoS computational biology*, 2016.

FIELDS OF STUDY

Major Field: Physics

Studies in Computational Neuroscience
Professor Tatyana Sharpee

ABSTRACT OF THE DISSERTATION

Information Maximization in Early Sensory Systems

by

Yilun Zhang

Doctor of Philosophy in Physics

University of California, San Diego, 2018

Professor Tatyana Sharpee, Chair
Professor Massimo Vergassola, Co-Chair

Information maximization is a strong candidate for the design principles of early sensory systems. Yet, previous applications of information maximization are mostly restricted to linear or small neural systems due to difficulty in computing mutual information. To solve this problem, we developed a method that could efficiently compute mutual information provided about high dimensional inputs by responses of a large neural population.

Using our method, we first quantify information transmission by multiple overlapping retinal ganglion cell mosaics. The results reveal a transition where one high-density mosaic becomes less informative than two or more overlapping lower-density mosaics. The results

explain differences in the fractions of multiple cell types and predict the existence of new retinal ganglion cell subtypes.

We then apply our method to neurons receiving time-varying stimuli and producing spike trains. Surprisingly, we found that the optimal nonlinearity for neurons receiving temporal correlated signal has finite slope, quantitatively explaining the ubiquitous sigmoid shape nonlinearity observed in neurons. The optimal nonlinearities we predicted agree well with experimental data without any parameters in our model.

We further investigate the optimal network connectivity for information transmission. Using olfactory system as a model, we analytically compute the optimal connectivity rate that maximize information transmission. The optimal connectivity rate has suprisingly simple expression and is inverse proportional to the input pattern sparsity. Our model also provides a feedforward solution to reconstruct odor signal. Our architecture is shown to be efficient, robust, and account for a number of experimental observations.

Chapter 1

Introduction

Information maximization (sometimes also known as efficient coding) is a strong candidate for the design principles of early sensory systems such as retina [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. Formulated by the language of information theory [11], the principle states that neural systems should maximize mutual information between its input stimulus and neural response, subject to certain biological constraints such as metabolic cost. The intuition behind the principle is straightforward: without distinguishing useful and unuseful information at sensing stage, the best that early sensory system can do is to capture as much information as possible, and transit to downstreaming areas using spike trains.

Information maximization has strong predictive power in sensory systems. For example, the contrast-response curve of fly large monopolar cell agrees with the prediction of information maximization principle [4]. Optimal receptive fields in an array of linear neurons predicted by information maximization have center surround structure [5, 6], consistent with receptive fields of retinal ganglion cells. Receptive fields predicted by information maximization based independent component analysis on natural images are Gabor-like edge filters [9], similar to the ones observed in visual cortex [12]. Information maximization predicts a phase transition from redundant coding to differential coding in a pair of binary neurons receiving the same input, explaining the existence

of adapting and sensitizing cell types [10].

However, previous applications of information maximization principle are mostly restricted to linear systems with Gaussian noise [6], systems using local Gaussian approximation [13], or small neural systems consisting of only few neurons [4, 10, 14, 15] for the following two reasons. First, analytical expression, or even closed-form approximation, of mutual information is hard to obtain with only few exceptions, such as linear channels with Gaussian input and noise. Second, time complexity of computing mutual information numerically is generally exponential in the number of dimensions (neurons), which makes it computationally intractable for (even slightly) larger system size. Biological neural systems, however, have huge dimensions and strong nonlinearity. Many of them cannot be reduced to or approximated by small scale systems or linear networks, preventing us from using information theoretical framework. For instance, there are multiple types of retina ganglion cell (RGC), different cell types have distinct spatial receptive field size, forming overlaping neuronal arrays [16, 17]. Such system cannot be reduced to neuron pairs due to different yet overlapping receptive fields, neither can it be well approximated by linear networks because of the highly nonlinear firing patterns. As a result, despite previous results on multiple RGC types using independent channels [18] and neuron pair [10], we still lack basic understanding of whether such cell type split can be explained and predicted by information maximization principle.

One approach to overcome the computational difficulty in calculating high dimensional mutual information is to use Fisher information as a link [19, 20]. However, this approach is only shown to work for one dimensional stimulus [21], which does not apply to many neural systems such as retina. Moreover, the bias of such approximation can be arbitrarily large if the effective noise in the neural representation is very nongaussian [20]. Another type of approach is based on variational method [22, 23, 24], which provides a lower bound on mutual information and use it as an approximation. The bias, however, depends on the choice of decoder family, and will not vanish in general. Thus, a major open problem remains for how to computed mutual

information for large neural populations receiving high dimensional input signal taking into account the nonlinear aspects of neuronal responses.

We solve the problem by developing a method that could efficiently compute mutual information provided about high dimensional inputs by responses of a large neural population (Chapter Compute Mutual Information for Large Neuronal Arrays). The method works for a correlated Gaussian distribution or any other stimulus distribution that can be factorized, and for neurons with sigmoid nonlinearity. Applying this approach to multiple overlapping retinal ganglion cell mosaics we identify a transition that determines when it is optimal for a system to shift from one high-density mosaic to two or more overlapping lower-density mosaics (Chapter Optimal Information Transmission by Multiple Mosaics of Retinal Cells). We further predict how the relative number of low and high threshold neurons should differ depending on the noise level, which in turn depends on a cell's preferred temporal frequency. These predictions explain experimental measurements showing that salamander ganglion cells with different temporal filters have different fractions of low threshold (sensitizing) and high threshold (adapting) cells.

Our method can be applied to neurons receiving time-varying stimuli and producing spike trains. Using the method, we found that, suprisingly, the optimal nonlinearity for neurons receiving temporal correlated signal has finite slope which is inverse proportional to input standard deviation (Chapter Optimal Nonlinearity for Neurons Receiving Temporal Correlated Signal). This explains the ubiquitous sigmoid shape nonlinearity observed in neural systems. The optimal nonlinearities we predicted agree well with experimental data without any parameters in our model. On the other hand, our results indicate that stochastic resonance in neural system [25, 26, 27, 28] is much more common than previously thought, where stochastic resonance is only restricted to limited cases such as neurons receive subthreshold signal [29, 30], multiple neurons have the same nonlinearity [31] or slope [32], only temporal independent stimuli are considered [10].

In Chapter Optimal Connectivity Rate for Information Transmission in Olfactory System, we extend our search for neural system optimal for information transmission to optimal network

connectivity. Using olfactory system as a model, we analytically compute the mutual information between two layers of randomly connected binary neurons and compute the optimal connectivity rate that maximize information transmission. The optimal connectivity rate has suprisingly simple expression and is inverse proportional to the input pattern sparsity. Our model provides a biological plausible and computational efficient feedforward solution to reconstruct signal in compressive sensing framework. Our architecture is shown to be robust to noise and account for a number of experimental observations.

Chapter 2

Compute Mutual Information for Large Neuronal Arrays

For a large group of linear-nonlinear (LN) neurons encoding a high dimensional stimulus $\vec{s}$ using population neural response $\vec{r}$, the mutual information between stimulus and neural response $I(\vec{s},\vec{r})$ is computationally intractable due to its high dimensionality. We show that the problem can be converted to a sum of one-dimensional problems provided one can identify a basis where the input distribution factorizes. Importantly, the neural responses depend simultaneously on multiple stimulus components, so this problem is more complex than that of independent response channels encoding independent inputs.

To model neural responses, we assume that their responses can be approximately described by a saturating logistic nonlinearity as a function of the stimulus component along the neuron's receptive field (RF). Specifically, the probability to observe a spike from the ith neuron (in a small time window) is given by:

$$p(r_i = 1|\vec{s}) = \frac{1}{1+e^{-\frac{u_i-\mu_i}{v_i^{\text{eff}}}}}, \tag{2.1}$$

where $u_i = \vec{f}^i \cdot \vec{s}$ represents the component of the stimulus $\vec{s}$ along the ith neuron's RF $\vec{f}^i$. The

parameter μ_i represents the mid-point (threshold) of the spiking nonlinearity where the spiking probability is 0.5. The parameter ν_i^{eff} is the inverse slope of the nonlinearity which represents the cumulative effect of multiple noise sources in the input neural circuitry [10, 33, 34, 35]. The response across the population is described by a response vector $\vec{r}$ whose components r_i take values 1 or 0 corresponding to the presence or absence of a spike from the ith neuron. The probability to observe the response vector $\vec{r}$ across the population is then given by:

$$p(\vec{r}|\vec{s}) = \frac{1}{Z(\vec{s})} \prod_i \exp\left(r_i \frac{u_i - \mu_i}{\nu_i^{\text{eff}}}\right), \tag{2.2}$$

with the normalization factor $Z(\vec{s}) = \prod_i \left[1 + \exp\left(\frac{u_i - \mu_i}{\nu_i^{\text{eff}}}\right)\right]$. We note that although this expression is obtained assuming the neural responses are conditionally independent given the stimulus, the final result will also hold in the presence of stimulus-independent noise correlations as long as the population response function still belongs to an exponential family.

Now if there is a basis within which the input distribution factorizes, $P(\vec{s}) = \prod_k P(s_k)$, then we can use this basis to represent both the stimuli and neural RFs. For concreteness, we use the Fourier components, because this is the basis within which translation invariant signals factorize, and this basis would be the most appropriate for analyzing the retinal arrays. Both stimuli and neural RFs can be described using their Fourier components $S_{\vec{k}}$ and $F_{\vec{k}}^j$ as:

$$s(\vec{x}) = \sum_{\vec{k}} S_{\vec{k}} e^{i\vec{k}\cdot\vec{x}}, \;\; f^j(\vec{x}) = \sum_{\vec{k}} F_{\vec{k}}^j e^{i\vec{k}\cdot(\vec{x}-\vec{x}_j)}, \tag{2.3}$$

where $\vec{x}_j$ marks the coordinates of the peak of jth neuron RF and F_k^j is a Fourier component of the jth neuron RF computed relative to $\vec{x}_j$. In these expressions, the Fourier components should satisfy the following constraints $S_{\vec{k}} = S^*_{-\vec{k}}$ and $F_{\vec{k}}^i = F^{i*}_{-\vec{k}}$ that derive from the fact that stimuli and RFs have to be real-valued functions. The stimulus component along jth neuron RF is given by $u_j = \sum_{\vec{k}} S_{\vec{k}} F_{\vec{k}}^{j*} e^{i\vec{k}\cdot\vec{x}_j}$.

Now, the full response probability from Eq. (2.2) can be re-written as

$$p(\vec{r}|\vec{s}) = \frac{1}{Z(\vec{s})} \exp\left[\sum_{\vec{k}} S_{\vec{k}} T_{\vec{k}}(\vec{r})\right] h(\vec{r}), \tag{2.4}$$

where the quantities $T_{\vec{k}}$ are obtained as weighted averages of neural responses as $T_{\vec{k}} = \sum_j r_j \frac{F_{\vec{k}}^{j*}}{\nu_j} e^{i\vec{k}\cdot\vec{x}_j}$, and $h(\vec{r}) = \exp(-\sum_i r_i \frac{\mu_i}{\nu_i})$. Because of the exponential structure of the full response probability, these quantities comprise the sufficient statistics [36] of the neural responses [37]. In other words, the mutual information between stimuli and the set of neural responses $\vec{r}$ is equal to the mutual information between the stimuli and the set of $T_{\vec{k}}$ values: $I(\{S_{\vec{k}}\}, \vec{r}) = (\{S_{\vec{k}}\}, \{T_{\vec{k}}\})$.

Using the chain rule for mutual information we have

$$\begin{aligned} I(\{S_{\vec{k}}\}, \{T_{\vec{k}}\}) &= \sum_i I(S_i, \{T_{\vec{k}}\} | S_{i-1}, S_{i-2}, ..., S_1) \\ &\geq \sum_i I(S_i, T_i | S_{i-1}, S_{i-2}, ..., S_1) \\ &\geq \sum_{\vec{k}} I(S_{\vec{k}}, T_{\vec{k}}), \end{aligned} \tag{2.5}$$

where we assume a large but finite number of neurons, thus T is discrete. The first inequality becomes equal when $T_{\vec{k}}$ is the sufficient statistics of $S_{\vec{k}}$. Further, the second inequality becomes equal when $S_{\vec{k}}$ is the sufficient statistics of $T_{\vec{k}}$.

One can show that when $\sigma_i \gg \sigma_{i+1}, \forall i$, where σ_i^2 is the variance of S_i, $T_{\vec{k}}$ is sufficient for $S_{\vec{k}}$. This is approximately true for natural stimuli where $\sigma_k^2 \propto k^{-2}$. Thus for natural stimuli we have the approximation

$$I(\vec{s}, \vec{r}) \approx \sum_i I(S_i, T_i | S_{i-1}, S_{i-2}, ..., S_1). \tag{2.6}$$

Note that although the conditionally mutual information above contains high dimensional integral over $S_1, S_2, ..., S_{i-1}$, it is effectively an one dimensional integral over the effective threshold, making it computational efficient.

When the signal correlation length and the system size is much larger than the receptive field size, both $S_{\vec{k}}$ and $T_{\vec{k}}$ becomes Gaussian [38], and both inequality in Eq. 2.5 become equal such that

$$I(\vec{s},\vec{r}) \approx \sum_{\vec{k}} I(S_{\vec{k}}, T_{\vec{k}}), \tag{2.7}$$

regardless of the neuron model. This is more computational efficient as one only needs to compute a series of one dimensional mutual information.

This approach makes it possible to convert a multi-dimensional information calculation into a sum of one dimensional calculation for the mutual information between independent stimulus components $S_{\vec{k}}$ and the corresponding sufficient statistics variables $T_{\vec{k}}(\vec{r})$. The variables $T_{\vec{k}}(\vec{r})$ are computed as linear combinations of the neural responses across the population. Thus the approach does take different patterns of population responses into account. It also indicates which patterns of neural responses carry information about which stimulus components.

Chapter 2, in part, is a preprint of the material as it will appear in 52th Annual Conference on Information Sciences and Systems (CISS). Zhang, Yilun; Kastner, David; Baccus, Stephen; Sharpee, Tatyana, 2018. The dissertation author was the primary investigator and author of this material.

Chapter 3

Optimal Information Transmission by Multiple Mosaics of Retinal Cells

3.1 Efficient Coding in Retina

The efficient coding hypothesis plays an important role in understanding the design principles of neural systems, especially sensory systems [1]. However, in practice the application of efficient coding hypothesis is largely restricted to linear systems with Gaussian noise [6], systems using local Gaussian approximation [13], or small neural systems consisting of only few neurons [4, 10, 14, 15]. The reasons are that, on the one hand, mutual information can be computed analytically for large systems only with strong simplifying assumptions, such as assuming linear input-output functions for individual neurons. On the other hand, numeric computation of mutual information is challenging for high dimensional variables, since the computational complexity is exponential in the number of dimensions. One approach to overcome this problem is to use Fisher information as a bound [19] or approximation [20] to mutual information, but this approach does not work with high dimensional stimuli. Another approach is to use variational approximation with certain decoder [39], but such methods only provide a

bound on mutual information, and the result depends on the choice of a specific decoder. Thus, a major open problem remains for how to characterize information transmitted by large neuronal populations taking into account both the nonlinear aspects of neuronal responses and the fact that these neurons might be sensitive to different components of multi-dimensional input signals.

To test the efficient coding hypothesis in large neural populations, we use the method we developed that could efficiently compute the mutual information provided about the high dimensional inputs by the responses of a large neuronal array Compute Mutual Information for Large Neuronal Arrays. Applying this approach to multiple overlapping retinal ganglion cell mosaics, consisting of ~ 1000 neurons, we identify a transition that determines when it is optimal for a system to shift from one high-density mosaic to two or more overlapping lower-density mosaics. We further show how the relative densities in the two mosaics depend on different types of input noise and their spatiotemporal filtering properties. The results account for the observed properties of retinal ganglion cell types in the salamander cells and predict new cell types in the primate and guinea pig retinas.

3.2 Transitions in the Coordinated Encoding by Retinal Ganglion Cell Mosaics

We now use this approach to compute and analyze information transmission in large arrays of retinal ganglion cells composed of multiple sub-types. The first question that we would like to address is whether it is better to encode the same type of the visual feature with higher spatial resolution or to encode different types of visual features each with lower spatial resolution. "Better" in this context is quantified as conveying greater amounts of mutual information about the full spatial stimulus pattern. We model each retinal array using a hexagonal grid (Fig. 3.1, insets). The spatial profile of each neuron is modeled as a difference of Gaussian to capture the

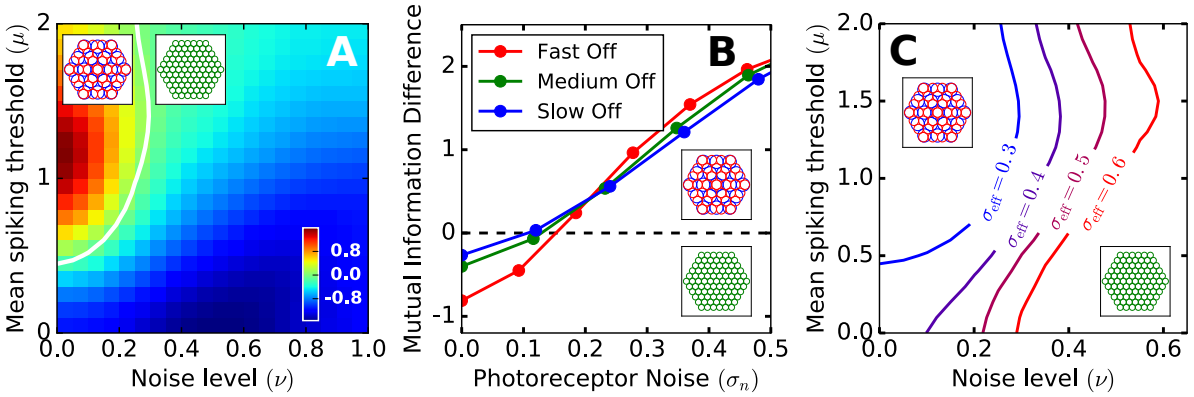

Figure 3.1: (**A**) Mutual information difference between optimal two-array encoding and one-array encoding as a function of neuron intrinsic noise level and mean spiking probability. We map mean spiking probability to mean spiking threshold to better compare with experimental data. White line show zero mutual information difference marking the transition of optimal encoding strategy from one array to two arrays. We use photoreceptor noise $\sigma_{\text{eff}} = 0.3$. (**B**) Mutual information difference between optimal two-array encoding and one-array encoding as a function of photoreceptor noise level. Black dashed line shows the transition from one array to two arrays. The intrinsic noise level and mean spiking threshold for each line are chosen to match the data from salamander [17]. Experimentally measured temporal kernels of each cell type (as shown in Fig. 3.3A) are used to calculate the effective photoreceptor noise. (**C**) Transition boundary for different levels of photoreceptor noise.

center-surround structure [40]

$$f(\vec{x}) = C(\sigma, w, \gamma)\left(\frac{1}{2\pi\sigma^2}e^{-\frac{(\vec{x}-\vec{x}_c)^2}{2\sigma^2}} - \frac{w}{2\pi\gamma^2\sigma^2}e^{-\frac{(\vec{x}-\vec{x}_c)^2}{2\gamma^2\sigma^2}}\right), \tag{3.1}$$

where $C(\sigma, w, \gamma)$ is chosen so that $||f||_2 \equiv \int f(\vec{x})^2 d\vec{x} = 1$, and $\vec{x}_c$ is the receptive field center.

The lattice spacing a, defined as the difference between nearest neighbor RF peaks, relates to the RF center size σ as $a = 2\sigma$, in agreement with experimental observations [41, 42, 40, 43, 44]. This relationship also constrains the number of neurons N in a given lattice as $Na^2 = \text{const}$. Both configurations – either a single array or two overlapping arrays – are constrained to have the same total number of neurons and to produce the same number of spiking responses, on average across the stimulus ensemble.

We find that a transition occurs when noise level v_{eff} decreases. For a given average spiking rate across the population, when noise level is high, a single higher density array conveys more information. When noise level decreases below a certain value, two overlapping mosaics convey more information than one higher-density mosaic. In Fig. 3.1A we plot results as a function of noise component ν associated with spike generation. Here one observes a transition boundary (white line) that separates a range of values for the noise level and neural thresholds where two overlapping mosaics convey more information from those values for which a single higher density mosaic convey more information.

The transition can also be induced by noise in the afferent circuitry. We model this type of noise by a Gaussian white noise added to stimulus, so that the input signal u according to which spikes are produced is described by:

$$u = \vec{f} \cdot (\vec{s} + \vec{n}), \tag{3.2}$$

where $\langle n(\vec{x})n(\vec{x'})\rangle = \delta(\vec{x} - \vec{x'})\sigma_n^2$. We will refer to this noise as photoreceptor noise because studies find it to be the dominant component prior to ganglion cell nonlinearity [33, 34, 35]. The

presence of an independent additive noise does not change the form of sufficient statistics, so our previous information calculation still holds. Yet, because the photoreceptor noise is added before ganglion cell integrate (spatial-temporal) inputs, the magnitude of this noise can be reduced by integration summarized by the spatio-temporal filters. In particular, when neurons form multiple mosaics, they have broader spatial filters compared to the case where there is only one type of neurons. Thus, forming multiple mosaics provides a powerful way to reduce input noise. Correspondingly, increasing input noise can by itself force the transition from encoding using one array to encoding using multiple arrays (Fig. 3.1B).

3.3 Interaction Between Multiple Noise Sources Impacts Optimal Coordination Between Retinal Cell Types

In presence of photoreceptor noise defined by Eq. 3.2, the spiking probability given filtered stimulus $p(r|u)$ is a sigmoid function $p(r=1|u)=\frac{1}{1+e^{-\frac{u-\mu}{\nu'}}}$. The spiking probability given stimulus is a convolution of sigmoid and gaussian $p(r|s)=\int p(r|u)p(u|s)du$, which is approximately another sigmoid function with different slope $p(r=1|s)=\frac{1}{1+e^{-\frac{u-\mu}{\nu^{\text{eff}}}}}$, where the new slope satisfies $(\nu^{\text{eff}})^2=\nu'^2+\frac{\pi}{8}\sigma_n^2$. Thus, given sigmoid spiking probability $p(r=1|u)$ and gaussian noise, $p(r=1|s)$ is a sigmoid function with higher effective ν.

To find out the exact expression for ν^{eff}, we note that ν' scales with the standard deviation of u such that $\nu'=\sigma_u\nu=\nu\sigma_{\vec{f}\cdot\vec{s}}\sqrt{1+\sigma_{\text{eff}}^2}$, where ν is the intrinsic inverse slope in unit of σ_u as we reported in our figures, σ_{eff} is defined in Eq. 3.4, and $\sigma_{\vec{f}\cdot\vec{s}}$ is the standard deviation of $\vec{f}\cdot\vec{s}$. Thus we have $(\nu^{\text{eff}})^2=(1+\sigma_{\text{eff}}^2)\nu^2\sigma_{\vec{f}\cdot\vec{s}}^2+\frac{\pi}{8}\sigma_n^2$, and the effective noise in Eq. 2.1 becomes:

$$\nu^{\text{eff}}=\nu\sigma_{\vec{f}\cdot\vec{s}}\sqrt{1+\sigma_{\text{eff}}^2+\frac{\pi}{8}\frac{\sigma_{\text{eff}}^2}{\nu^2}} \tag{3.3}$$

σ_{eff}^2 is the effective photoreceptor noise that depends on both the power spectrum of stimuli and

neural filters:

$$\sigma^2_{\text{eff}} = \frac{\sigma^2_n}{\sigma^2_{\vec{f}\cdot\vec{s}}} = \frac{\sigma^2_n}{\int \hat{f}(\vec{k})^2 k^{-2} d\vec{k} \int \hat{g}(\omega)^2 \omega^{-2} d\omega}. \tag{3.4}$$

Here f and g are the spatial and temporal Fourier components of the filter describing neuron's spatiotemporal RF. Eq. 3.4 takes into account statistics of natural visual scenes showing $1/f^2$ dependence on both spatial and temporal frequencies [45, 46].

Using Eq. 3.3 and 3.4 we can now analyze how the interaction between these noise sources impacts properties of the optimal mosaics. There are two main effects. The first effect is mainly quantitative, showing that with increasing photoreceptor noise, the transition boundary between one-array and two-array encoding shifts towards larger ν (Fig. 3.1C).

The second effect induced by photoreceptor noise can lead to large changes in the predicted fraction of different neuronal types. For example, for low values of the photoreceptor noise, near the transition from one-array encoding to two-array encoding, when two overlapping mosaics first start to provide more information than a single dense mosaics, the number of neurons is very unevenly distributed between the arrays. Most of the neurons form one (denser) array and only a small fraction of neurons form the other array (Fig. 3.2A, D). Neurons in the lower density array have higher thresholds and spike less often. In contrast, for high photoreceptor noise, neurons in the lower density array have lower thresholds and spike more often (Fig. 3.2C, F). These theoretical observations can help explain the observed differences in the number of adapting and sensitizing neurons for fast Off, medium Off, and slow Off ganglion cells in salamander [17]. Because fast Off neurons have faster temporal filters, according to Eq. 3.4, the effective photoreceptor noise for these neurons will be higher than that for the medium/slow Off neurons which have comparable temporal kernels. As a result, the expected ratio between adapting and sensitizing cells for these three classes will be different, with largest differences between fast and medium/slow subtypes (Fig. 3.3).

The theoretical analyses also predict the existence of new cell types. In Fig. 3.4A we

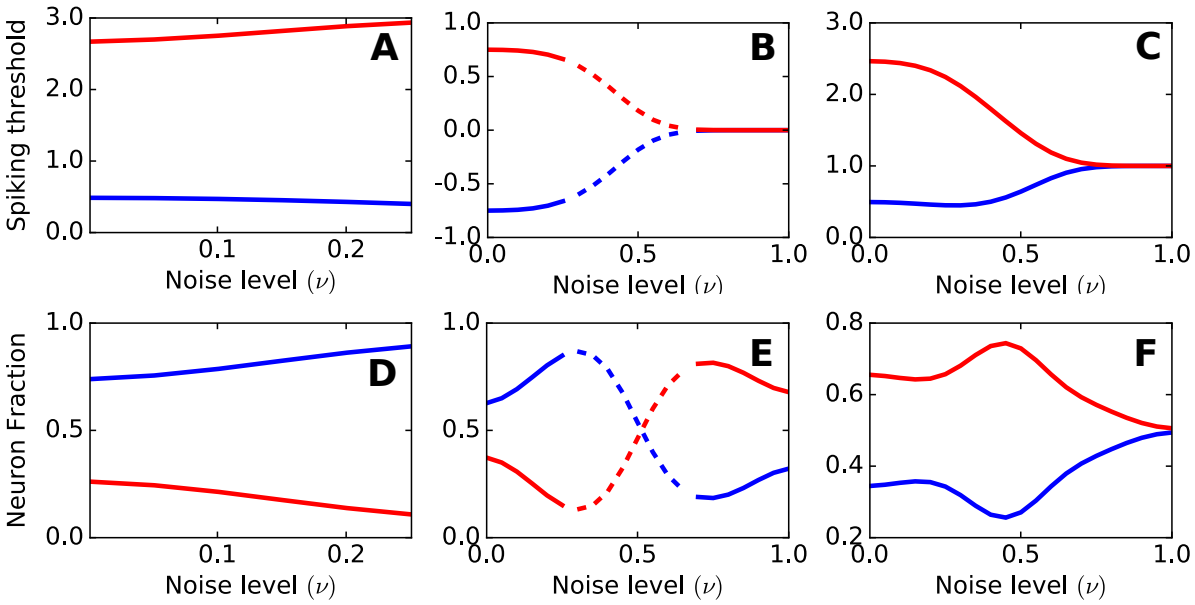

Figure 3.2: Optimal threshold and neuron fraction of high threshold cell (red) and low threshold cell (blue) under low (A, $\sigma_{\text{eff}} = 0.3$), intermediate (B, $\sigma_{\text{eff}} = 0.6$) and high (C, $\sigma_{\text{eff}} = 1.0$) photoreceptor noise. Solid/dashed line shows the regime where two/one array is optimal. For low photoreceptor noise, only the regime where two arrays are optimal is shown, since only the transition from two arrays to one array take places, and redundant coding (two neuron types have the same threshold and neuron fraction) is not optimal. For high photoreceptor noise, two arrays are always preferable than one array, and a transition to redundant coding occurs. For intermediate photoreceptor noise, both transitions could coexist, with the transition from two arrays to one array occurs at low intrinsic noise level, while the transition to redundant coding occurs at high intrinsic noise level. Another noticeable difference between different photoreceptor noise is the ratio of cells of high and low spiking thresholds. High threshold cells have few neurons than low threshold cells when photoreceptor noise is low, while the opposite is true when photoreceptor noise is high. For intermediate photoreceptor noise, both ratio could be optimal, with a potential transition take places as intrinsic noise changes.

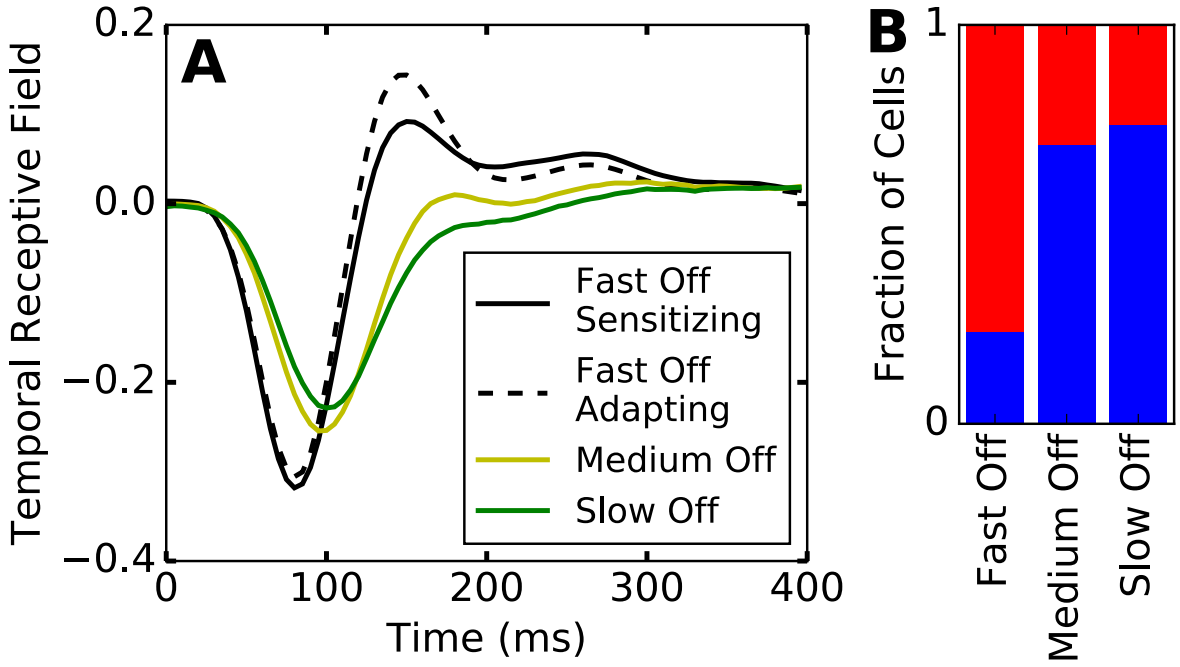

Figure 3.3: (**A**) Average temporal receptive fields measured in experiments. Faster temporal kernel results in higher effective photoreceptor noise, thus have more adapting cells. Similarly, slower temporal kernel results in more sensitizing cells. (**B**) (Adapted from [17]) Fractions of adapting (red) and sensitizing (blue) cells of different cell types measured in Salamander.

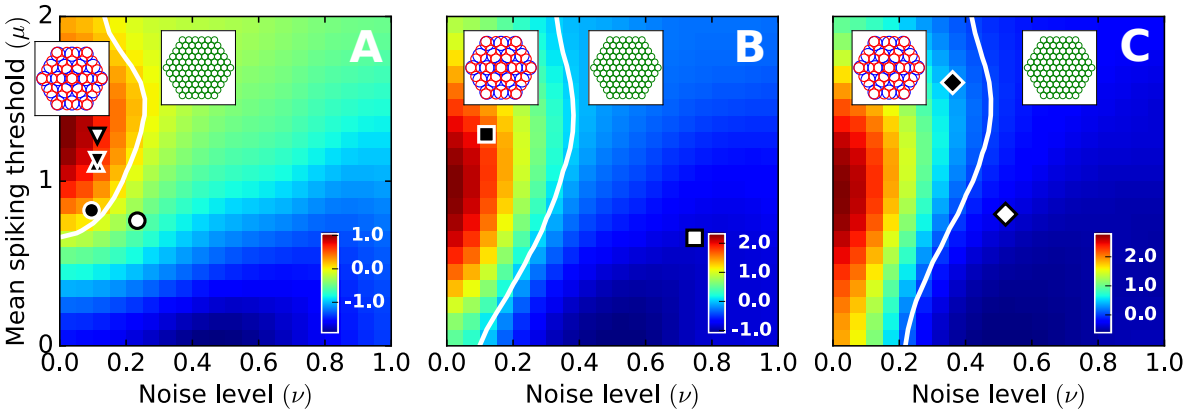

Figure 3.4: Mutual information difference and predicted transition boundary for salamander, primate, and guinea pig, compared with experimental data. Markers with white/black face color indicate the data of ON/OFF cells, respectively. (**A**) Same as Fig. 3.1A but for photoreceptor noise $\sigma_{\rm eff} = 0.25$, and data points from salamander [17]. Circles represent fast cell, up triangle represents medium cell, down triangles represent slow cell. (**B**) Same as (**A**), but for photoreceptor noise $\sigma_{\rm eff} = 0.4$ and data from primate parvo cells [47]. (**C**) Same as (**A**), but for photoreceptor noise $\sigma_{\rm eff} = 0.5$ and data from guinea pig [48]. To compare our predictions with experiments, we transform the measured values to the intrinsic ν and μ that are independent from photoreceptor noise or neuron's filters. The measured inverse slope $\nu_{\rm exp}$ is simply $\nu^{\rm eff}$ in unit of the standard deviation of $\vec{f}\cdot\vec{s}$ such that $\nu_{\rm exp} = \nu^{\rm eff}/\sigma_{\vec{f}\cdot\vec{s}}$, so we have $\nu^2_{\rm exp} = (1+\sigma^2_{\rm eff})\nu^2 + \frac{\pi}{8}\sigma^2_{\rm eff}$. Similarly we have $\mu^2_{\rm exp} = (1+\sigma^2_{\rm eff})\mu^2$. Using these relations, we can transform measured values $\nu_{\rm exp}$ and $\mu_{\rm exp}$ (averaged across all neuron data of the same type) to ν and μ and compare with our calculation.

show data from salamander cell types [17] relative to the transition boundary of the region in the parameter space where one array encoding provides more information than the two array encoding. We find that Off neurons are located on one side of the boundary where two arrays provide more information whereas most On neurons are located on the other side of the boundary. In the case of the salamander neurons with fast time courses these predictions agree with experimental measurements because it is known that the fast-Off neurons split into subtypes whereas fast-On neurons do not. At the same time, theory predicts that slow On neurons should split into subtypes. This is a new prediction to be tested experimentally.

In Fig. 3.4B and C we compare theoretical predictions for the number of cell types with vertebrate data from primate [47] and guinea pig [48], respectively. Here the predictions are that Off neurons from the guinea pig as well as primate parvo cells should split into sub-types whereas the corresponding On neurons in those species should not. The theoretical predictions are made for larger photoreceptor noise than in the salamander because photoreceptor noise increases with temperature. We estimate the photoreceptor noise for guinea pig to be approximately 2 times the value of salamander using linear interpolation [49]. Meanwhile, primate is shown to have smaller noise in single photon response than guinea pig [50], thus we use a smaller photoreceptor noise for primate than for guinea pig.

3.4 Optimal Coordinated Encoding Explains Transient Adaptation

Our results predict that, for larger effective photoreceptor noise where the fraction of adapting cell is larger (Fig. 3.2C), the optimal threshold difference between the adapting and sensitizing cells increase when the intrinsic noise level ν decreases. Such prediction explains the observation that the threshold difference between adaptive and sensitizing cells increases transiently after signal variance suddenly drops [17]. Theoretical results for lower effective

photoreceptor noise, such as for medium and slow Off neurons that have slower temporal filters, indicate the opposite behavior. For example, in Fig. 3.2A the difference between thresholds of the two subtypes would decrease following a decrease in intrinsic noise level ν.

3.5 Optimal Redundant Coding

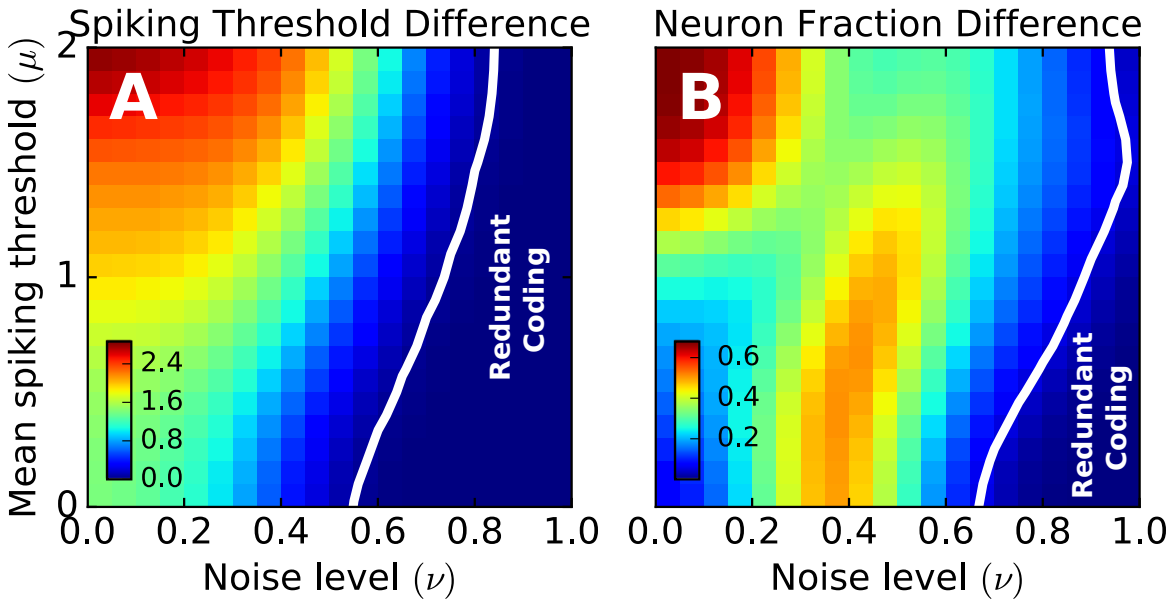

Figure 3.5: Optimal spiking threshold difference (**A**) and neuron fraction difference (**B**) of two neuron sub-types in two arrays. White lines shows the transition boundary from different threshold (**A**) and receptive field size (**B**) to the same threshold/RF size. Redundant coding (two cell types in two arrays have the same spiking threshold/receptive field size) is optimal on the right side of the boundary. Photoreceptor noise is $\sigma_{\text{eff}} = 1.0$.

Finally, it is worth commenting on the possibility that two overlapping arrays can be formed by neurons with identical response characteristics. We find such possibility, but only for relative large noise levels associated with both spiking and in photoreceptors (Fig. 3.5). So far, none of the experimental data we have fell into this region. The observation of optimal redundant coding creates the prediction that the usual association between mosaics and cell types in the retina can be violated in the presence of high noise.

Chapter 3, in full, is a preprint of the material as it will appear in 52th Annual Conference on Information Sciences and Systems (CISS). Zhang, Yilun; Kastner, David; Baccus, Stephen; Sharpee, Tatyana, 2018. The dissertation author was the primary investigator and author of this material.

Chapter 4

Optimal Nonlinearity for Neurons Receiving Temporal Correlated Signal

4.1 Introduction

Experimentally measured neuron nonlinearity (also known as tuning curve or activation function) can usually be well described by sigmoid function. It is commonly believed that the finite slope of sigmoid nonlinearity is due to noise or biological constraint. Indeed, a sigmoid nonlinearity can be well approximated by an effective Gaussian input noise followed by a threshold nonlinearity (see Input Noise and Effective Nonlinearity), and it has been shown that information transmission rate (mutual information between stimulus and neural response) increases monotonically as the slope of the sigmoid nonlinearity increases (see Step Function Nonlinearity Maximizes Mutual Information Without Temporal Correlation). The optimal nonlinearity would simply be a threshold function, far from the measured sigmoid shape.

Meanwhile, it is known that when multiple binary neurons receive the same scalar input, the optimal sigmoid nonlinearity has finite slope [31]. Yet, such model cannot explain the ubiquitous sigmoid nonlinearity since we recently found that when one or multiple neuronal arrays

are encoding high dimensional inputs (such as retina), the optimal nonlinearity again degenerates to a threshold function (see Step Function Nonlinearity Maximizes Mutual Information Without Temporal Correlation), qualitatively different from experimental observation.

Suprisingly, using the method we developed (see Compute Mutual Information for Large Neuronal Arrays), we found that threshold nonlinearity is optimal only when neuron receive white noise as input. In real biological system where neuron receive temporal correlated signal, the optimal sigmoid nonlinearity that transmit most information has finite slope which is inverse proportional to input standard deviation. More importantly, the optimal nonlinearities we predicted agree well with experimental data without any parameters in our model.

4.2 Step Function Nonlinearity Maximizes Mutual Information Without Temporal Correlation

Consider a binary neuron with sigmoid nonlinearity centered at zero $p(r|s) = \frac{\exp(rs/\nu)}{1+\exp(s/\nu)}$, and receiving a symmetric stimulus s without temporal correlation. The mutual information between stimulus s and neural response r is $I(r,s) = H(r) - H(r|s)$. The derivative of $I(r,s)$ with respect to ν is

$$\begin{aligned}
\frac{\partial I(r,s)}{\partial \nu} &= -\frac{\partial H(r|s)}{\partial \nu} \\
&= \int \frac{\partial}{\partial \nu} \sum_{r=0,1} p(r|s) \log p(r|s) ds \\
&= \int \sum_{r=0,1} [1 + \log p(r|s)] \frac{\partial p(r|s)}{\partial \nu} ds \\
&= -\int \frac{s^2}{\nu^3} \frac{e^{\frac{s}{\nu}}}{(1+e^{\frac{s}{\nu}})^2} ds \\
&< 0,
\end{aligned} \tag{4.1}$$

where we use the fact that $H(r)$ is constant given the symmetry $p(s) = p(-s)$. Therefore, mutual information is maximized when $\nu = 0$, corresponding to step function nonlinearity.

This conclusion does not change for one or multiple neuronal array receiving high dimensional signal (like retinal ganglion cell array does) as well as nonzero firing threshold, as shown in Fig. 4.1.

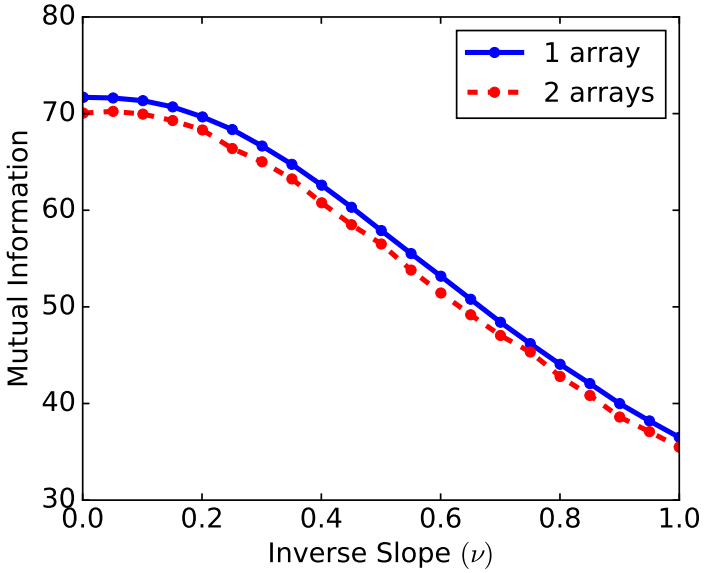

Figure 4.1: Mutual information of neuronal array(s) decreases as ν increases. Results from Optimal Information Transmission by Multiple Mosaics of Retinal Cells. Gaussian stimuli with spatial correlation (but without temporal correlation) are used. Neurons in the same array have the same but shifted receptive field. All neurons are binary neurons with sigmoid nonlinearity.

4.3 Optimal Nonlinearity for Temporal Correlated Stimuli

4.3.1 Compute mutual information for neurons receiving temporal correlated stimuli

Consider the model where a neuron encodes a time-varying stimulus $s(t)$ using its spike train $r(t)$. A general assumption in information theoretical framework is that neuron is optimized for capturing information about stimulus $s(t)$ using its spike train $r(t)$. A quantitative measure of

the information captured is the mutual information between stimulus and spike train

$$I(s,r) = H(s) + H(r) - H(s,r), \tag{4.2}$$

where $H(s)$ and $H(r)$ are the Shannon entropy of stimulus and spike train, and $H(s,r)$ is the joint entropy of stimulus and spike train [51].

We consider a neuron model with finite temporal resolution Δt. We assume Δt is the inverse of the maximal firing rate R

$$\Delta t = \frac{1}{R}. \tag{4.3}$$

Given such temporal resolution, spike train $r(t)$ is represented by binary time series, where $r(t_i)$ is 1 or 0 depending on whether neuron fires in time bin i. Stimulus $s(t)$ is also binned into $s(t_i)$ with same temporal resolution. The probability that neuron fires at t_i is determined by the sigmoid nonlinearity

$$p(r(t_i) = 1 | s(t_i)) = \frac{1}{1 + \exp(-\frac{s(t_i) - \mu}{\nu})}, \tag{4.4}$$

where ν determines the shape of the sigmoid nonlinearity, μ is the firing threshold. Here the temporal filter of neuron doesn't appear because $s(t)$ is the filtered signal. Filtering will not affect mutual information as long as the filter is reversible.

The optimal nonlinearity in information theoretical framework is the one that maximizes the mutual information between stimulus and spike train. For simplicity, we model stimulus $s(t)$ as temporal correlated Gaussian signal with zero mean, unit variance and random Fourier phase such that the stimulus statistics is fully characterized by its power spectrum $C(\omega)$. Thus, the optimal nonlinearity is fully determined by stimulus power spectrum $C(\omega)$ and maximal firing rate R.

Compute mutual information between time-varying stimulus s and spike train r is challanging due to the high dimensionality (number of neurons times number of time bins). To

overcome this problem, we use our recently developed method (see Compute Mutual Information for Large Neuronal Arrays). We first decompose stimulus into Fourier components

$$s(t) = \sum_{\omega} (A_\omega \cos 2\pi\omega t + B_\omega \sin 2\pi\omega t). \tag{4.5}$$

The spiking probability for neuron i at time t_j is determined by sigmoid nonlinearity

$$p(r_{ij}) = \frac{e^{r_{ij}\frac{s_j-\mu_i}{\nu_i}}}{1+e^{\frac{s_j-\mu_i}{\nu_i}}}, \tag{4.6}$$

where r_{ij} is either 1 or 0 representing neuron spiking or not. The probability to generate full spike train r is

$$\begin{aligned} p(r) &= \prod_{i,j} p(r_{ij}) \\ &= e^{\sum_{i,j}\frac{r_{ij}s(t_j)}{\nu_i}} h_1(s)h_2(r) \\ &= e^{\sum_{i,j}\frac{r_{ij}}{\nu_i}\sum_\omega (A_\omega \cos 2\pi\omega t_j + B_\omega \sin 2\pi\omega t_j)} h_1(s)h_2(r) \\ &= e^{\sum_\omega \left(A_\omega \sum_{i,j}\frac{r_{ij}}{\nu_i}\cos 2\pi\omega t_j + B_\omega \sum_{i,j}\frac{r_{ij}}{\nu_i}\sin 2\pi\omega t_j\right)} h_1(s)h_2(r), \end{aligned} \tag{4.7}$$

where $h_1 s$ is some function that does not depend on spike train r, $h_2(r)$ is some function that does not depend on stimulus s. Thus, the sufficient statistics for stimulus Fourier components A_ω, B_ω are

$$\begin{aligned} T(A_\omega) &= \sum_{i,j}\frac{r_{ij}}{\nu_i}\cos 2\pi\omega t_j, \\ T(B_\omega) &= \sum_{i,j}\frac{r_{ij}}{\nu_i}\sin 2\pi\omega t_j. \end{aligned} \tag{4.8}$$

The mutual information between high dimensional stimulus and spike train can be computed using a series of mutual information between two scalars

$$I(s,r) = \sum_\omega I(A_\omega, T(A_\omega)) + I(B_\omega, T(B_\omega)). \tag{4.9}$$

We compute $I(A_\omega, T(A_\omega))$ and $I(B_\omega, T(B_\omega))$ numerically using Non-Parametric Entropy Estimation Toolbox (NPEET) [52].

4.3.2 Input Noise and Effective Nonlinearity

We consider the input noise added to the stimuli $\hat{s} = s + n$, where n is a Gaussian white noise with zero mean and variance σ_n^2. The firing probability given a true signal can be written as a convolution of the sigmoid nonlinearity and a Gaussian

$$\begin{aligned} p(r=1|s) &= \int p(r=1|\hat{s}) p(\hat{s}|s) d\hat{s} \\ &= \int \frac{1}{1+e^{-(\hat{s}-\mu)/\nu}} \frac{1}{\sqrt{2\pi}\sigma_n} e^{-(\hat{s}-s)^2/2\sigma_n^2} du \\ &\approx \frac{1}{1+e^{-(s-\mu)/\nu'}}, \quad \nu' = \sqrt{\nu^2 + \frac{\pi\sigma_n^2}{8}}. \end{aligned} \tag{4.10}$$

In the last line above, we used the relation that the convolution of a sigmoid function with slope ν and a Gaussian with standard deviation σ is approximately another sigmoid function with slope $\nu' = \sqrt{\nu^2 + \frac{\pi\sigma^2}{8}}$ [53]. Thus, adding Gaussian input noise with variance σ_n^2 is approximately equivalent to increasing ν to $\sqrt{\nu^2 + \frac{\pi\sigma_n^2}{8}}$.

4.3.3 Optimal nonlinearity of single neuron

To study how the optimal nonlinearity depends on stimulus power spectrum $C(\omega)$, we compute the mutual information between stimulus and spike train $I(s,r)$ for three different stimuli: white noise corresponding to previous study, white noise filtered by temporal filter to model neuron receiving white noise as input, stimulus with $1/\omega^2$ power spectrum filtered by temporal filter to model neuron receiving natural stimulus as input. We use temporal filter and maximal firing rate R measured in Salamander retinal ganglion cell (fast off adapting cell type) [17]. Natural stimulus is modeled as signal with $1/\omega^2$ power spectrum as observed in experiments [46].

To find the optimal sigmoid nonlinearity, we plot the mutual information $I(s,r)$ as a function of the sigmoid inverse slope ν (Eq. 4.4), as shown in Fig. 4.2. We set $\mu = 0$ since it is optimal in the case of single neuron without any firing rate constraints.

When $s(t)$ is white noise, it is sufficient to study single spike information because nearby stimuli in time are independent, so are nearby spikes. In this case, the mutual information is simply proportional to the mutual information of single spike

$$I(s,r) = RTI(s(t_1), r(t_1)), \tag{4.11}$$

where t_1 is the first time bin, T is the total measurement time, and R is the maximal firing rate. As shown in Fig. 4.2, the mutual information decreases monotonically as ν increases from 0, in agreement with our analytical calculation (see Step Function Nonlinearity Maximizes Mutual Information Without Temporal Correlation). The optimal nonlinearity in this case is a step function, corresponding to $\nu = 0$.

However, we would like to point out that $s(t)$ should almost never be white noise in biological system because of the following reasons: first, stimulus in biological system, either from environments or from other neurons, should have some (finite) temporal correlation (for example, stimuli are often continuous in time); second, stimulus will be filtered over time before passing through nonlinearity to generate spike train, so even white noise input becomes temporal correlated after filtering. In the case of temporal correlated stimulus $s(t)$, the optimal nonlinearity is usually not step function, as shown below.

To show the effect of temporal filtering, we use temporal filtered white noise to model the scenario where a neuron receives white noise as input. We use temporal filter and maximal firing rate R measured in Salamander (fast off adapting) retinal ganglion cell [17]. As shown in Fig. 4.2, the information transmission rate no longer decreases as ν for $\nu < 0.15$ (in unit of stimulus standard deviation). Compared to unfiltered white noise which is only affected by single spike

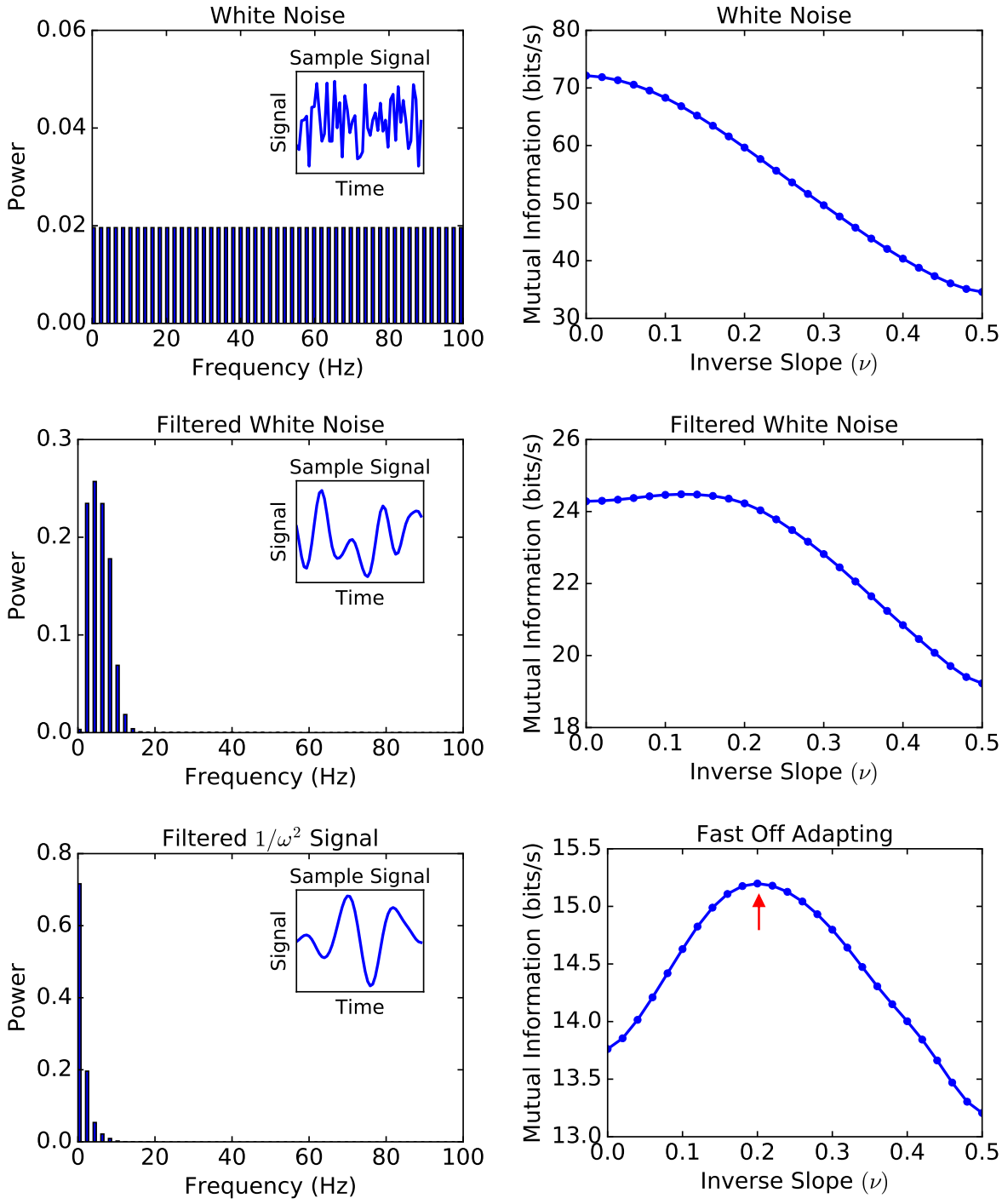

Figure 4.2: Optimal nonlinearity depends on signal temporal power spectrum. Left column shows signal power spectrum and sample signal (inset), right column plots mutual information between signal and neuron response as a function of nonlinearity inverse slope ν. Top, middle, bottom row show white noise, filtered white noise, filtered $1/\omega^2$ signal, respectively. Measured temporal filter and firing rate of fast off adapting ganglion cell in Salamander retina is used. Red arrow indicates experimentally measured value.

information, introducing temporal correlation to stimulus shifts optimal nonlinearity towards higher ν. This effect is more prominent when neuron receives simulated temporal correlated signal as shown below.

To model more realistic case where neuron receives natural stimulus rather than white noise, we simulate stimulus in natural scene using correlated Gaussian signal with unit variance and $1/\omega^2$ power spectrum [46]. As above, we use the same temporal filter and maximal firing rate measured in Salamander retina. Suprisingly, the mutual information between stimulus and spike train $I(s,r)$ is maximized when ν is around 0.2, far from zero which was previously thought to be optimal. More importantly, the predicted optimal value agrees well with experimental data (red arrow) without any parameters in the model, indicating that nonzero ν is optimal to capture information, rather than simply induced by some undesired noise.

Compare white noise input with natural stimulus in Fig. 4.2, we can see a trade-off between single spike information, which decreases as ν increases, and temporal correlation induced information, which increases as ν. When stimulus is continuous in time, spikes in nearby time bins are more likely to be the same when ν is small, since neuron's dynamic range where stimulus change can be effectively detected is proportional to ν. Thus, although single spike transmit more information when ν is small, spike train are more redundant and transmit less information. We can see the trade-off from the definition of mutual information

$$I(s,r) = H(r) - H(r|s), \tag{4.12}$$

where $H(r)$ is the Shannon entropy of spike train, and $H(r|s)$ is the conditional entropy. When we consider single spike or s is white noise, $H(r)$ does not depend on ν. The optimal nonlinearity should minimize $H(r|s)$ which increases monotonically with ν. So $\nu = 0$ is optimal. However, when we consider a spike train with temporal correlated stimulus, $H(r)$ also increases with ν as discussed above. Thus there is a trade-off between maximize $H(r)$ and minimize $H(r|s)$. As a

result, the optimal ν depends on stimulus correlation as well as neuron's maximal firing rate, as we will discuss later.

In addition to data from fast OFF adapting cells in Salamander retina, we also measured temporal filter and maximal firing rate of fast OFF sensitizing cells, medium OFF cells, slow OFF cells and ON cells in Salamander retina (ganglion cells) [17]. We predict the optimal ν of each cell type assuming stimulus has $1/\omega^2$ power spectrum before filtering, as shown in Fig. 4.3. The predictions agree well with measured value for different types of OFF cells.

4.3.4 Optimal nonlinearity of ON/OFF neuron pairs

In the previous section, we only consider the model where a stimulus is received by one neuron. In biological systems, a stimulus is sometimes received by multiple types of neuron to encode different features. For example, in visual system, ON and OFF retinal ganglion cells both receive the same visual stimulus, but respond to different contrast; in auditory system, same sound signal is received by different cochlear hair cells, each encoding different frequency information. The optimal nonlinearity of each neuron type could be different depending on what features neuron ensemble is encoding. Using ON/OFF neurons as example, we investigate how the optimal nonlinearity depends on neuron type as well as energy constraint.

Same as above, we simulate natural stimulus using correlated Gaussian signal with zero mean, unit variance and $1/\omega^2$ power spectrum. Stimulus is received by an ON/OFF neuron pair, which we assume have the same temporal filter and maximal firing rate measured by experiments. Due to symmetry, the optimal ν for ON and OFF neuron should be the same, while the optimal μ should sum to zero. Therefore we only consider the symmetric configuration where $\nu_{\mathrm{ON}} = \nu_{\mathrm{OFF}} = \nu$ and $\mu_{\mathrm{ON}} = -\mu_{\mathrm{OFF}} = \mu$. We compute the mutual information between stimulus and spike train for different ν and μ using the measured temporal filter and maximal firing rate of fast OFF adapting cells, fast OFF sensitizing cells, medium OFF cells, slow OFF cells and ON cells in Salamander retina (ganglion cells) [17], as shown in Fig. 4.4. The measured ν and

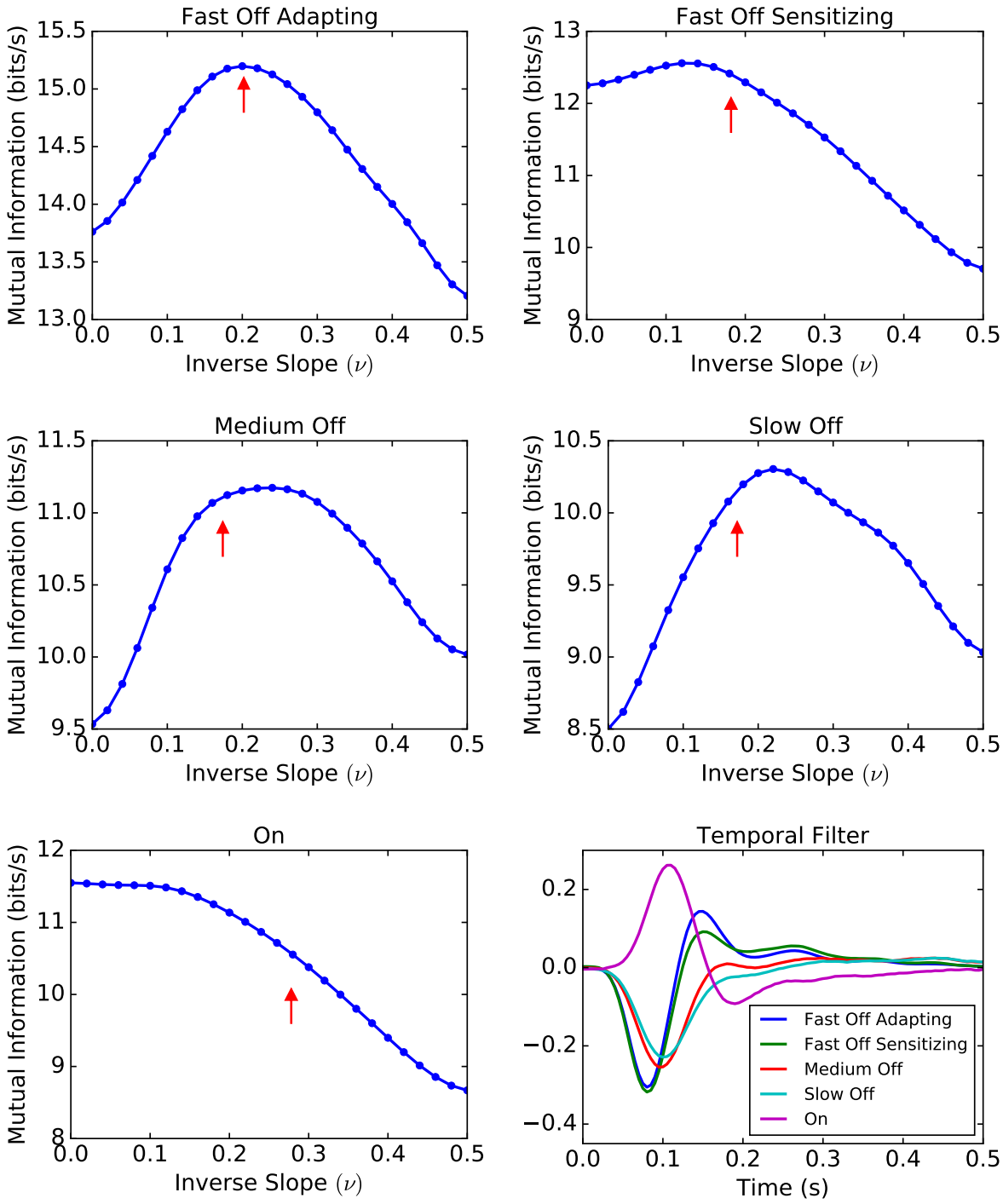

Figure 4.3: Optimal nonlinearity of different neuron types compared with experimental data (red arrows). Measured temporal filters and firing rate are used, as shown in the lower right subfigure.

μ are mostly in the region where information transmit rate is near optimal, consistent with our prediction.

The above predictions are purely based on maximizing information transmission without considering energy consumption. Neuron, in principle, should work in an energy efficient way so that it captures most information with least amount of energy. Therefore, information transmission is often maximized subject to energy constraint [54, 10], leading to more sparse firing activity. Assuming spiking consumes most energy in a neuron, we can characterize the energy consumption of a neuron by its mean firing rate

$$\bar{r} = R \int_{-\infty}^{\infty} p(r=1|s)p(s)ds. \tag{4.13}$$

In our calculation, $p(s)$ is a zero mean unit variance Gaussian distribution, $p(r=1|s)$ is a sigmoid function, and the mean firing rate primarily depends on firing threshold μ. Keeping mean firing rate as a constant, the optimal nonlinearity moves towards higher ν and μ from the optimal value without firing rate constraint, as shown in Fig. 4.4. The optimal nonlinearity should not move to the opposite direction because in that way it transmit less information with more energy consumption, which is a complete loss. Indeed, for all cell types, measured ν and μ are larger than or close to the predicted optimal value without energy constraint, consistent with our theory. Moreover, for OFF cell types, measured ν and μ are close to predicted optimal value when we keep mean firing rate at certain level (Fig. 4.4).

4.3.5 Optimal nonlinearity depends on maximum firing rate

As shown above, the optimal nonlinearity depends on stimulus power spectrum $C(\omega)$. Larger ν is preferable when stimulus power is more concentrated on lower frequencies. On the other hand, the system is scale invariant in time so that nothing changes if both the stimulus dynamics and neuron spiking are speeded up or slowed down for a constant factor. Thus, the

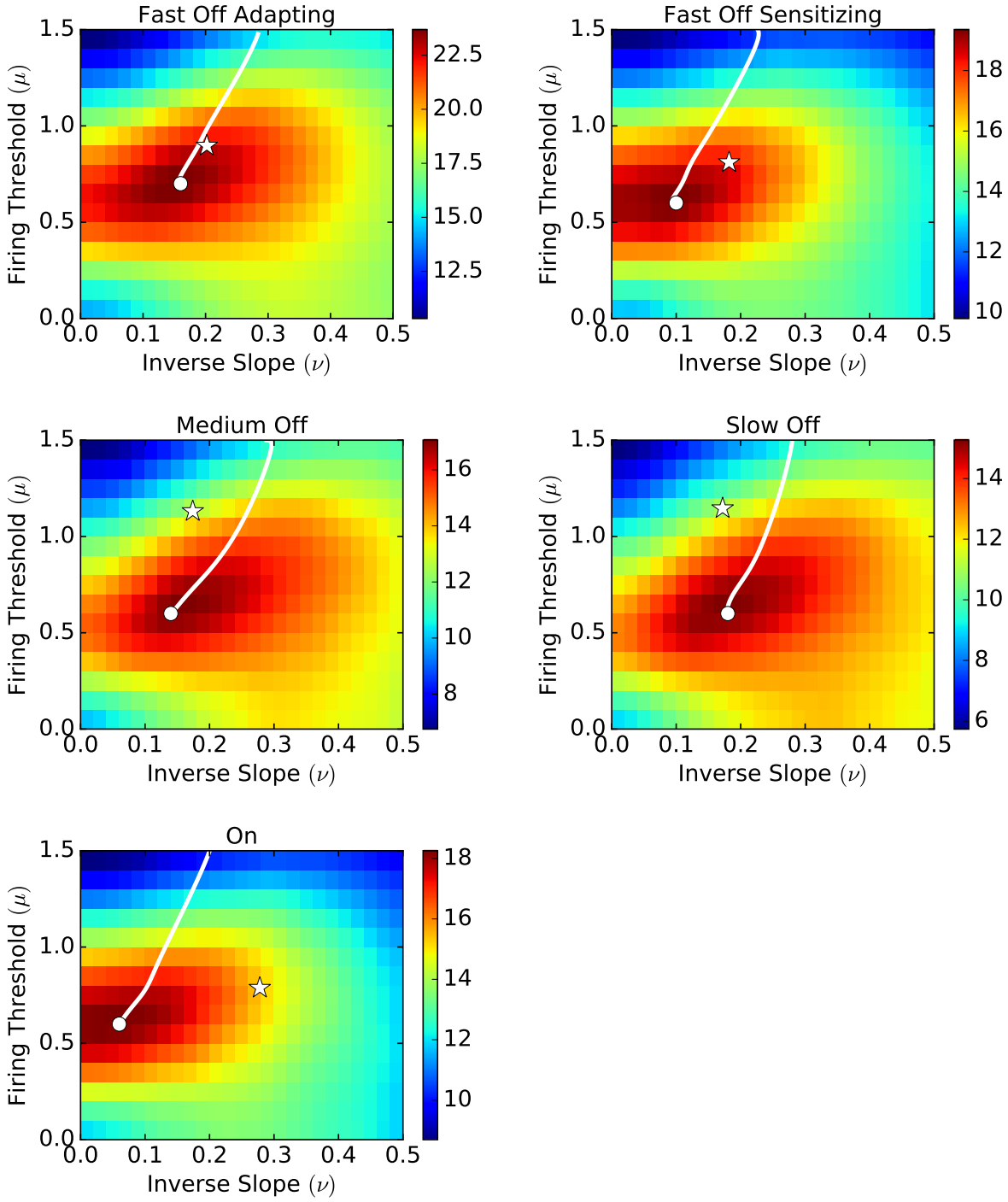

Figure 4.4: Optimal nonlinearity of different neuron types compared with experimental data (white stars). Colorbar shows mutual information (bits/s) between signal and neuron pair response. White circle indicates optimal value without any constraints, while white line shows the trace of optimal values when total firing rate is constrained. Measured temporal filters and firing rate are used, as shown in the lower right subfigure.

optimal nonlinearity also depends on the maximal firing rate R through a scaled version of power spectrum

$$C_R(\omega) = C(\omega R). \tag{4.14}$$

At the one end, for any stimulus power spectrum $C(\omega)$ that is continuous in ω and has finite value, $C_R(\omega)$ becomes the spectrum of white noise as $R \to 0$ considering the fact that high enough frequencies will be cut off due to sampling limit. This corresponds to the limit where neuron fires very slowly so that the interspike interval is much longer than the correlation time scale, so that nearby spikes are independent. Thus, the response entropy $H(r)$ does not depend on nonlinearity as long as firing rate is constant. The optimal nonlinearity should minimize conditional entropy $H(r|s)$, which leads to step function ($\nu = 0$).

At the other end, if stimulus has finite variance, $\int_0^\infty C(\omega)d\omega$, $C_R(\omega)$ becomes a delta function as $R \to \infty$. This corresponds to the limit where neuron fires infinite times before the stimulu changes. The variance of the spike count vanishes compared to its mean, so the neuron is effectively a deterministic rate neuron with zero noise. Therefore, the optimal nonlinearity should be the one that maximizes response entropy $H(r)$, which leads to the cumulative distribution of stimulus, as has been studied [4]. Since the cumulative distribution of unit variance Gaussian can be well approximated by sigmoid function with $\nu \approx 0.6$ [53], neuron with finite firing rate should not have higher ν than 0.6 if no firing rate (energy) constraints are considered. Indeed, none of the neuron types we measured have ν higher than 0.6, consistent with our prediction.

Neurons in biological system fall between these two limits as they have finite maximal firing rate R. From the analysis above, optimal ν depends on $C_R(\omega)$, but should be smaller than the high rate limit 0.6. Neuron with higher R has the chance to fire more spikes before stimulus changes (significantly), and thus will prefer larger ν. We confirm this by computing the optimal ν with different maximal firing rate R while keeping stimulus power spectrum $C(\omega)$, as shown in Fig. 4.5.

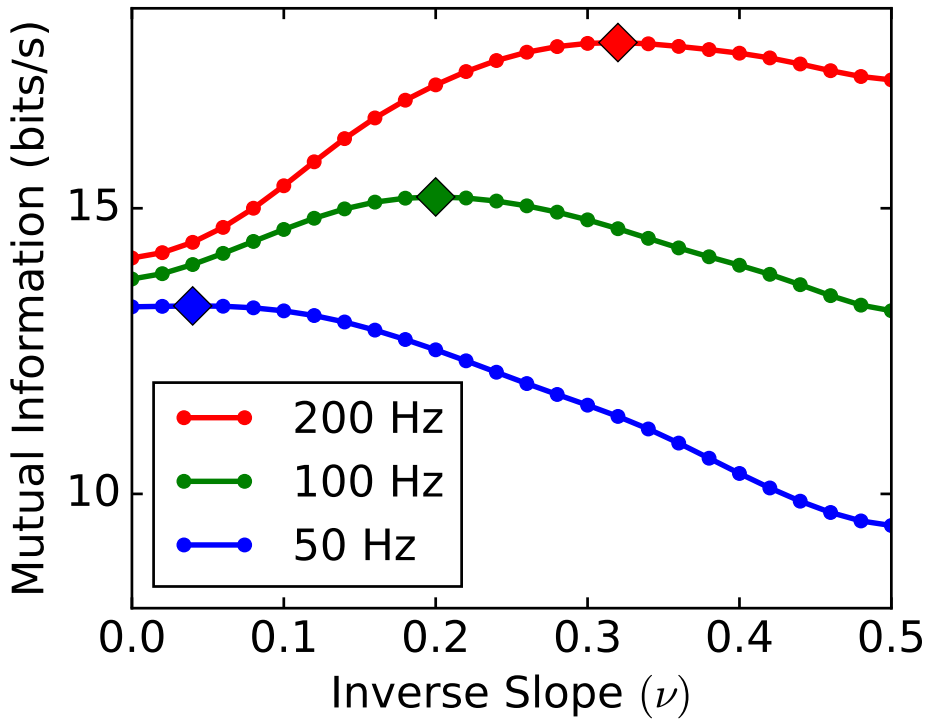

Figure 4.5: Optimal ν increases with maximum firing rate R. Diamond marker on each curve marks the optimal ν for that maximum firing rate. $1/\omega^2$ signal with temporal filter from fast OFF adapting cells are used to compute optimal nonlinearity.

4.4 Nonzero ν is a choice rather than noise

Nonzero ν, as observed in real neurons [17, 48], is sometimes called noise in previous studies [10] because: first, it was previously thought to reduce information transmission rate; and second, sigmoid nonlinearity with nonzero ν is equivalent to step function nonlinearity ($\nu = 0$) after adding a Gaussian input noise (see Methods). By treating ν as noise, we imply that neuron tries to reduce ν to zero to increase information transmission rate, but there is an uncontrollable source adding noise to signal before it goes through nonlinearity so that neuron cannot achieve zero ν. We show that in biological system where neuron receives temporal correlated signal, optimal ν that transmits most information significantly deviates from zero. Neuron should adapt its nonlinearity to match stimulus statistics. Actually, it has been observed that neuron actively adapts its ν to match stimulus contrast [17], showing that nonzero ν is a choice, rather than noise.

4.5 Noise can enhance information transmission

As shown in Methods, adding input noise is approximately equivalent to increasing ν. Thus, when ν is smaller than optimal, adding input noise to stimulus before nonlinearity can actually enhance information transmission rate. Such effect is also called stochastic resonance [25, 26, 27, 28]. However, unlike previous studies where stochastic resonance is only restricted to limited cases such as neurons receive subthreshold signal [29, 30], multiple neurons have the same nonlinearity [31] or slope [32], only temporal independent stimuli are considered [10], we show that stochastic resonance is much more common. In generally, stochastic resonance is possible as long as the optimal ν is nonzero, which is usually true when neuron receives temporal correlated stimulus.

Chapter 4, in part, is currently being prepared for submission for publication of the material. Zhang, Yilun; Sharpee, Tatyana. The dissertation author was the primary investigator and author of this material.

Chapter 5

Optimal Connectivity Rate for Information Transmission in Olfactory System

5.1 Early olfaction as compressive sensing

Although it is still debated how many different odorants humans can perceive, the most commonly cited number is on the order of 10^4 [55, 56, 57], much greater than the 500 olfactory receptor neuron (ORNs) types. Many other species, including both vertebrates and insects, have the same order of magnitude of ORN types or even fewer (around 1000 in mice, 50 in Drosophila). The order of magnitude difference between the number of odorants and ORN types implies that humans as well as other species rely on compressed representations, potentially following the principles of compressed sensing [58, 59, 60, 61].

In the compressed sensing framework [58], sparse high dimensional signals can be accurately reconstructed using a small number of measurements provided that the input signals are sparse. Natural odors are sparse in the sense that they are dominated by a few molecular compounds [62, 63, 64]. The relevance of compressed sensing algorithms to olfactory coding is reinforced by the anatomical organization of the olfactory system. High dimensional odor signals

are compressed into a low-dimensional representation in terms of the activity of a relatively small number of glomeruli in the olfactory bulb, in the case of vertebrates, or the antennal lobe in the case of invertebrates. The standard compressed sensing algorithm performs signal reconstruction as a constrained ℓ_1 minimization [58]. Such optimization can be solved through neural dynamics [60, 59], but the resulting reconstructions were considerably less fault tolerant than observed experimentally. For example, mice olfactory discrimination remains essentially intact when half of glomeruli are disabled [65] whereas theoretical reconstructions fail at this level of signal interference [59]. Furthermore, signal reconstruction based on dynamical optimization by construction requires more time for signal recognition compared to feedforward reconstruction schemes. Here we describe a feedforward reconstruction scheme based on compressed sensing ideas that is both fault tolerant and matches the main features of the organization of the olfactory system. The results demonstrate that a purely feedforward network is capable of robustly compressing/decompressing binary signal without dynamical optimization. We analytically compute the optimal connectivity rate of the network, and showed that it is mathematically equivalent to the optimal sparsity of a random projection with threshold nonlinearity.

5.2 A compressed sensing model of the olfactory system

We begin by reviewing the main results from compressed sensing literature as they pertain to olfactory coding. The odor signal s^0 can be described as a binary vector of length N where each element is either 1 or 0 depending upon whether a given molecular compound is present or not in the odor. We refer to the number K of nonzero components in the odor as the odor sparsity. The main premise of compressed sensing is that a sparse signal s^0 can be compressed into a vector $x = As^0$ of length $M < N$ and then recovered with high reconstruction quality provided $K \ll N$. The encoding matrix A has dimensions $M \times N$; its matrix elements can be chosen randomly. With this setup, the original signal s^0 can be recovered exactly from the convex ℓ_1 optimization

problem [58]

$$\hat{s} = \arg\min ||s||_1 \quad \text{subject to } x = As^0. \tag{5.1}$$

Although the ℓ_1 minimization problem can be solved in polynomial time, it is not straightforward to implement such optimization algorithms in a neural circuit. One solution involves a two-layer neural network that perform similar ℓ_1 minimization through neural dynamics [60]. However, this imposes certain requirements on the structure of recurrent connections in the second layer together with a static nonlinear activation function. Another alternative implementation relies on ℓ_2 minimization instead of ℓ_1. In this case, the reconstruction is obtained simply as $\hat{s} = (A^T A)^{-1} A^T x$ where the $^{-1}$ represents a pseudo-inverse relation. However, such an approach does not produce exact signal reconstruction [61] and would predict much larger errors than observed in olfactory experiments.

5.3 Robust feedforward reconstruction of sparse odors

We now propose a model for the olfactory system, which can compress and robustly recover sparse binary signal with high probability, without using any dynamical optimization. The solution is based on a nonlinear binary encoding model instead of the linear encoding model used in the conventional compressed sensing approach. Specifically, the compressed vector x has the form of a threshold function $x_i = \mathcal{H}(x_i^l - \theta_c)$ where $x^l = As^0$ and $\mathcal{H}$ is the Heaviside step function with $\mathcal{H}(0) = 1$. We assume that the measurement matrix (affinity matrix) A is a $M \times N$ random binary matrix where each element is chosen independently to be either 1 or 0 with equal probability p and $1 - p$, respectively. It is worth mentioning that while we use a random connectivity matrix in our model, we do not assume that this matrix differs across individuals. Rather, the randomness is meant to characterize how well the system works in the absence of specificity between odorants and glomeruli identity. By extending the definition of $\mathcal{H}$ to vectors,

the measurement vector x can be compactly written as

$$x = \mathcal{H}(As^0 - \theta_c), \tag{5.2}$$

where $\theta_c = 1$, reflects that all measurements larger than 1 are set to 1 so that x is binary. This corresponds to a binary model of glomeruli activity described by the binary vector x. The threshold value of $\theta_c = 1$ corresponds to a logical OR operation, so that glomerulus k will be activated if any of the odor components that are associated with inputs to this glomerulus are activated.

To reconstruct the original signal, the glomeruli activity x are projected to another layer of neurons (neurons in the olfactory cortex of vertebrates or Kenyon cells in the mushroom body of insects) which has the same dimension as the original signal s^0. The activity of neurons in this layer is denoted by vector $\hat{s}$ which has the same dimensionality N as the original signal s^0. The reconstructed signal can be computed as

$$\hat{s} = \mathcal{H}(W^T x - \theta_r), \tag{5.3}$$

where θ_r is the activation threshold for neurons in the reconstruction layer. The reconstruction matrix W equals the measurement matrix A normalized to 1 by column, i.e. $W_{ki} = A_{ki}/\sum_k A_{ki}$. With this normalization, the reconstruction threshold $\theta_r = 1$ corresponds to logical AND operation. That is, odor component i will be detected as present if all glomeruli that feed signals to node i in the reconstruction layer are activated. Below we will present most of the results for $\theta_r = 1$ and then analyze how the reconstruction quality and recovery robustness depend on this threshold. We will also determine the optimal connectivity ratio from the compression to the reconstruction layer that maximizes the fidelity of reconstructions.

5.4 Maximal information transmission

Our feedforward model can be thought of as an information transmission channel that compresses, transmits, and decompresses a sparse binary signal. To find the optimal network configuration, we seek to maximize mutual information between the input and output of the channel as has been done to characterize performance in the visual and other sensory systems. The mutual information between s^0 and $\hat{s}$ is given by

$$I(s^0,\hat{s}) = \sum_{s^0}\sum_{\hat{s}} P(\hat{s}|s^0)P(s^0)\log_2\frac{P(\hat{s}|s^0)}{P(\hat{s})}. \tag{5.4}$$

For a given signal sparsity K, the conditional probability $P(\hat{s}|s^0)$ of the reconstructed signal $\hat{s}$ given the original signal s^0 can be computed as:

$$P(\hat{s}|s^0) = p_{\text{false}}^{N_{err}}(1-p_{\text{false}})^{(N-K-N_{err})}, \tag{5.5}$$

where $p_{\text{false}} \equiv P(\hat{s}_i = 1 \mid s_i^0 = 0)$ is the probability of false detection for an odor component and $N_{err} = ||\hat{s}||_0 - K$ is the number of false detection events for the odor s^0. We note that for $\theta_r = 1$, the probability to miss an odor component is zero provided the corresponding ORN activates at least one of the glomeruli. In this regime, the information is fully determined by the false detection rate p_{false}, and as we show below decreases proportionally with p_{false}.

Assuming a uniform prior over individual odor components $P(s^0) = 1/\binom{N}{K}$, one can also compute the probability distribution of reconstructed signals:

$$P(\hat{s}) = \sum_{s^0} P(\hat{s}|s^0)P(s^0) = \frac{\binom{K+N_{err}}{K}}{\binom{N}{K}} p_{\text{false}}^{N_{err}}(1-p_{\text{false}})^{(N-K-N_{err})} \tag{5.6}$$

Putting together Eqs. (5.4)-(5.6), the mutual information can be written as

$$I(s^0,\hat{s}) = \log_2 \binom{N}{K} - \sum_{N_{err}=0}^{N-K} \binom{N-K}{N_{err}} p_{\text{false}}^{N_{err}} (1-p_{\text{false}})^{(N-K-N_{err})} \log_2 \binom{K+N_{err}}{K}.$$

When $(N-K)p_{\text{false}} \ll 1$, the summation above can be well approximated by its leading nonzero term

$$\begin{aligned} &\sum_{N_{err}=0}^{N-K} \binom{N-K}{N_{err}} p_{\text{false}}^{N_{err}} (1-p_{\text{false}})^{(N-K-N_{err})} \log_2 \binom{K+N_{err}}{K} \\ &\approx (N-K)p_{\text{false}} \log_2(K+1), \end{aligned} \tag{5.7}$$

so that the expression for the mutual information becomes:

$$I(s^0,\hat{s}) \approx \log_2 \binom{N}{K} - (N-K)p_{\text{false}} \log_2(K+1). \tag{5.8}$$

Thus, for given N and K, maximizing $I(s^0,\hat{s})$ can be approximated by minimizing the probability of false detection p_{false}.

5.5 Optimal connectivity rate

The false detection rate that appears in Eq. 5.8 can be computed as

$$\begin{aligned} p_{\text{false}} &\equiv P(\hat{s}_i = 1 \mid s_i^0 = 0) \\ &= \sum_{k=1}^{M} P(\hat{s}_i = 1 \mid ||T_i||_0 = k) P(||T_i||_0 = k \mid s_i^0 = 0) \\ &= \sum_{k=1}^{M} \left[1-(1-p)^K\right]^k \frac{\binom{M}{k} p^k (1-p)^{M-k}}{1-(1-p)^M} \\ &= \frac{1}{1-(1-p)^M} \sum_{k=0}^{M} \left[1-(1-p)^K\right]^k \binom{M}{k} p^k (1-p)^{M-k} - \frac{(1-p)^M}{1-(1-p)^M} \\ &= \frac{\left[1-p(1-p)^K\right]^M - (1-p)^M}{1-(1-p)^M}, \end{aligned} \tag{5.9}$$

where $T_i \equiv \{x_k \in x | A_{ki} = 1\}$, and p is the average connectivity rate from the compression to the reconstruction layer. In the last line above we use the binomial expansion. Because we are interested in the regime where M is large, we have $(1-p)^M \ll \left[1-p(1-p)^K\right]^M \ll 1$ as long as p is not too small. Thus, Eq. 5.9 can be approximated with great accuracy by the following simple equation:

$$p_{\text{false}} = \left[1-p(1-p)^K\right]^M. \tag{5.10}$$

As shown in the inset of Fig. 5.1B, Eq. 5.10 provides an accurate approximation when the connectivity p is not too sparse. Since our main interest is near the optimal connectivity rate (see below) where Eq. 5.10 is very accurate, we will use Eq. 5.10 unless specified.

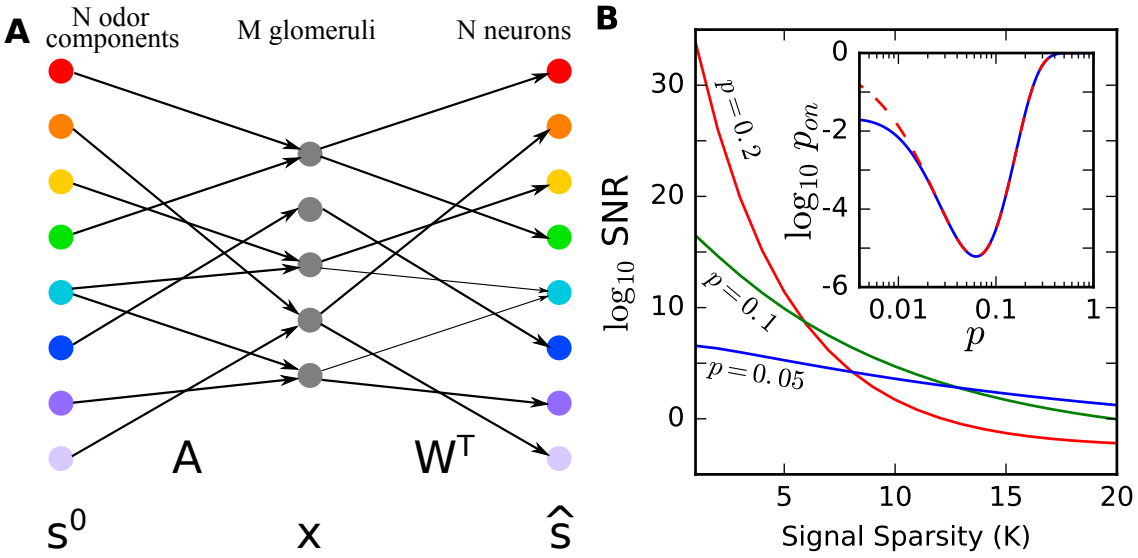

Figure 5.1: (A) Illustration of the model structure. An odor is represented by a sparse binary vector s^0 of its mono-molecular components. This signal is compressed into the activities of M glomeruli represented by a binary vector x through a binary measurement matrix A. The signal is then recovered as the activities of N neurons in the mushroom body or olfactory cortex represented by a binary vector $\hat{s}$ through another matrix W^T. (B) The signal-to-noise ratio (SNR) as a function of signal sparsity K where $N = 10000$ and $M = 500$. For a given K, there is a optimal connectivity rate $p = p_m$ that maximizes SNR. At the same time, even for a system optimized to a given K, decreasing K still increases SNR. Inset: false detection rate p_{false} as a function of average connectivity p; $M = 500$ and $K = 15$ are chosen for this illustration. Solid line is exact formula, while dashed line is the approximation using Eq. 5.10. We can see that Eq. 5.10 is a very good approximation to the exact formula when p is not too small.

As expected, the false detection rate p_{false} decreases as the number of glomeruli M increases and as the signal sparseness K decreases. Importantly, for a given M and K, there is an optimal p, which we refer to as p_m, that minimizes p_{false}, as shown in Fig. 5.1B. Taking $\partial p_{\text{false}}/\partial p = 0$ leads to

$$p_m = \frac{1}{K+1}. \tag{5.11}$$

It is worth noticing that the optimal connectivity p_m is independent of the number of glomeruli M, and depends only on the signal sparseness K. Thus, optimal connectivity depends exclusively on the level of sparseness of signals in the environment and can be determined prior to any measurements on neural circuits.

For an optimal connectivity $p = p_m$, the probability of fault activation decreases exponentially as M increases and thus can be very small. This indicates that the proposed feedforward compression-reconstruction scheme from Fig. 5.1A can achieve exact recovery with high probability.

To test the reconstruction quality, we compute the signal-to-noise-ratio (SNR) of the recovered signal. Since all nonzero components in the original will be recovered, the only source of errors in the reconstructed signal are due to false detection rates. Therefore, we can define the SNR of recovered signal as

$$\text{SNR} = \frac{||s^0||_0}{<||\hat{s}||_0> - ||s^0||_0} = \frac{K}{(N-K)p_{\text{false}}}, \tag{5.12}$$

as shown in Fig. 5.2A-C, where $<\cdot>$ denotes the expectation value. We can see from Fig. 5.2B that the SNR increases exponentially with M. For our case where $K \ll N$, we can achieve a high SNR for a number of glomeruli M much smaller than the number of odor components N or, equivalently, the number of third-order neurons.

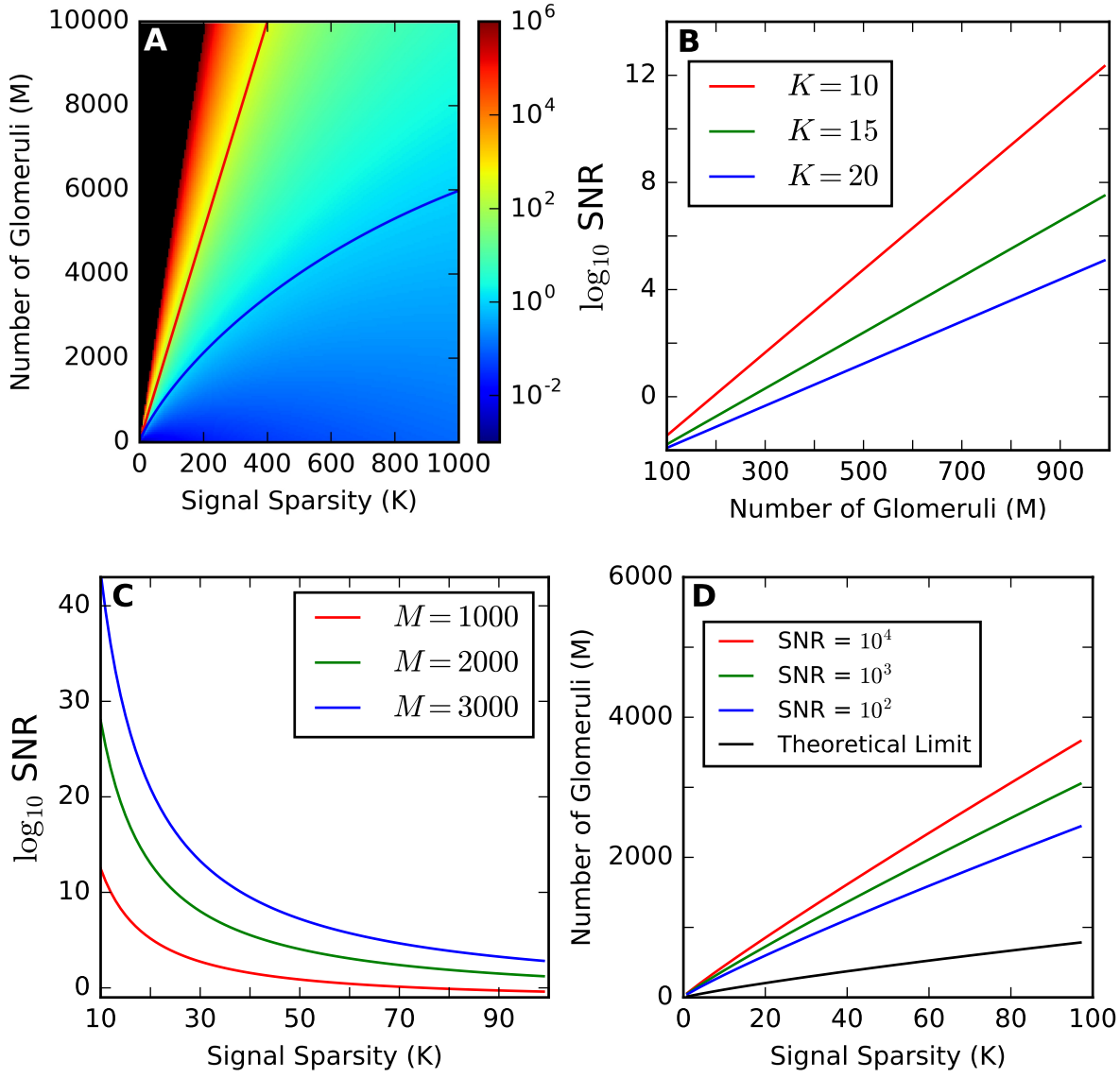

Figure 5.2: Signal-To-Noise-Ratio (SNR) of the recovered signal in our model. $N = 10000$ is used. (A) SNR as a function of K and M. Black is shown for SNR $> 10^6$. The blue line shows SNR $= 1$, and the red line shows SNR $= K$, i.e. one error occurs on average. (B) Optimal SNR as a function of M. (C) Optimal SNR as a function of K. (D) Number of glomeruli required to reach threshold SNR when optimal connectivity rate is used.

5.6 Equivalence to optimal random projection without reconstruction

Although p_m is defined as the optimal connectivity rate that maximizes the mutual information between original signal s^0 and recovered signal $\hat{s}$, we now show that p_m also maximizes the mutual information between signal s^0 and compressed signal x.

To show this, we consider two different signal s_1^0 and s_2^0 that lead to the same reconstruction $\hat{s}_1 = \hat{s}_2 = s$. Due to the symmetry of the connectivity, when signal $s^0 = s$, reconstructed signal is also s. So s produce the same reconstruction as s_1^0 and s_2^0. Since only false positive occurs in the model, the nonzero set S of s is a superset of the nonzero set S_1 of s_1^0. Consider their difference $DS_1 \equiv \{i|s_i = 1, s_{1i}^0 = 0\}$, for $\forall i \in DS_1$, $T_i \subseteq \cup_{j\in S_1} T_j$. So $\cup_{j\in S_1} T_j = \cup_{j\in S} T_j$, i.e. $x_1 = x$. Similarly, we have $x_2 = x$ and $x_1 = x_2$. Thus, for any two difference signal that lead to the same reconstructed signal, their compressed signal x are identical. As a result, the map from compressed signal x to reconstructed signal $\hat{s}$ is reversible, and the mutual information between original signal and compressed signal is the same as the mutual information between original signal and reconstructed signal $I(s^0, x) = I(s^0, \hat{s})$. Therefore, p_m, which maximizes $I(s^0, \hat{s})$, also maximizes $I(s^0, x)$.

The derivative above generalizes our results from a specific model with certain type of reconstruction to a general random projection with threshold nonlinearity. Our results show that in general, the optimal connectivity sparsity is inverse propotional to the input sparsity.

5.7 Compression rate and sparsity

A key characteristic of a compression algorithm is the compression ratio $\alpha \equiv M/N$. In previous compressed sensing frameworks, the critical compression ratio α_c above which the signal can be perfectly recovered was shown to only depend on the relative signal sparsity $f \equiv K/N$. As

$f \to 0$, $\alpha_c(f) \sim -f \log f$ [66]. To compute the critical compression ratio for our reconstruction algorithm, we note that from Eq. 5.12, $\log p_{\text{false}} = \log f - \log(1-f) - \log \text{SNR}$. In the strong compression limit where $f \equiv K/N$ is small, this yields

$$\log p_{\text{false}} \approx \log f - \log \text{SNR}. \tag{5.13}$$

On the other hand, for the optimal connectivity rate p_m and large K, $\log p_{\text{false}}$ can also be simplified using Eq. 5.10 as follows:

$$\begin{aligned} \log p_{\text{false}} &= M \log \left[1 - \frac{1}{K+1} \left(1 + \frac{1}{K} \right)^{-K} \right] \\ &\approx M \log \left(1 - \frac{1}{K+1} e^{-1} \right) \approx -\frac{M}{eK} = -\frac{\alpha_{\text{SNR}}}{ef}. \end{aligned} \tag{5.14}$$

where α_{SNR} is defined as the compression rate to achieve a certain SNR. Combining Eq. 5.13 and Eq. 5.14, in the limit of strong compression where $f \to 0$, the critical compression ratio behaves as $\alpha_{\text{SNR}} \sim -f \log f$. We note that care should be taken when the SNR becomes comparable to or larger than N because $1/f = N/K \leq N$, so that $\log \text{SNR}$ cannot be neglected when $f \to 0$.

The obtained critical compression rate can be compared to its theoretical limit. The latter corresponds to the minimal number of bits required to encode a sparse signal:

$$M_{min} = \left\lceil \log_2 \binom{N}{K} \right\rceil, \tag{5.15}$$

where $\lceil x \rceil$ is the smallest integer not less than x. When N and K are large but $f \equiv K/N$ is small, using Stirling's approximation, we obtain that

$$\begin{aligned} M_{min} \times \log 2 &\approx N \log N - K \log K - (N-K) \log(N-K) \\ &\approx K \log N - K \log K + K = K - K \log f, \end{aligned} \tag{5.16}$$

This yields that the theoretically possible compression ratio α_{min} in the strong compression limit of $f \to 0$ as

$$\alpha_{min} \to f \log_2 e/f, \tag{5.17}$$

which also yields $\alpha_{min} \sim -f \log f$ as $f \to 0$.

Notice that although both α_{SNR} and α_{min} behave as $-f \log f$ for $f \to 0$, they have different proportionality coefficients. To be more specific, $\alpha_{\text{SNR}} \sim ef \log 1/f$ while $\alpha_{min} \sim (\log 2)^{-1} f \log 1/f$. As a result, $\alpha_{\text{SNR}}/\alpha_{min} \to e \log 2 \approx 1.88$ as $f \to 0$. Thus, the number of glomeruli needed in our model is about twice the theoretical limit but is achieved here with an extremely simple feedforward encoding model.

As shown in Fig. 5.2-D, the number of required glomeruli increases sub-linearly with K, and logarithmically with SNR. In practice, with only a few times more glomeruli than the theoretical limit, a very high SNR can be achieved.

5.8 Robustness and fault tolerance

Advances in experimental techniques provide opportunities to test our theory under the circumstances of extreme genetic manipulations. For example, following a genetic manipulation that caused most olfactory receptor neurons to express a single odorant receptor M71, the M71 ligand acetophenone activates half of the glomeruli. Despite this drastic manipulation, mice can still readily detect other odors in the presence of acetophenone, while their discrimination performance is only moderately compromised [65]. This result is consistent with our model. Assume there are M glomeruli in our model and half of them are always turned on (corrupted). Such a system is equivalent to a model with only $M/2$ glomeruli, since the anomalously activated glomeruli will not affect signal recovery. Thus, the odor signal can still be recovered, but the SNR is decreased, which is in agreement with the experimental result. As a comparison, in previous compressed sensing framework, one can only allow a small percentage of corrupted glomeruli

even when $M > N$ [58].

In another set of experimental studies, part of the glomeruli in mice are removed or disabled [67, 68, 69]. It is shown that the ability to discriminate odors and simple odor mixtures is not impaired even when most of the glomeruli are removed or disabled. This seemingly surprising finding is also consistent with our model. From previous results, one can see that decreasing M will only lead to larger noise in the recovered odor signal but not to a failure of the system if the activation threshold for neurons in the reconstruction layer can be properly adapted to the new M. Assume the mice need SNR $> \nu$ to discriminate odors. When K is small, the minimal M needed for discrimination is

$$M_{low} = \frac{\log \frac{K}{N\nu}}{\log[1-p(1-p)^K]}. \tag{5.18}$$

From experiment data, $p \approx 0.05$ (although this is a very rough estimation, see [70, 65, 71, 72]). One can check that the equation above is insensitive to variations in K and $N\nu$ over a broad range. If we assume $K < 10$ (as in the experiments) and $N\nu$ is within the range of $10^4 \sim 10^5$, then M_{low} is roughly between 200 and 300, or around 20% of the glomeruli, which is in good agreement with the data in those experiments.

On the other hand, our model can tolerate negative gloleruli noise (false negative) by changing its recovery threshold θ_r. Although we use $\theta_r = 1$ in our results for analytical solution, it is very likely that real biological systems would use a lower threshold θ_r. With $\theta_r < 1$, the SNR is somewhat lower, as shown in Fig. 5.3, yet the system is more robust to noise in the reconstruction stage since the activation of a third-order neuron doesn't require all of its connected gloleruli to be active and it also leaves room for odor generalization and pattern completion [73]. Indeed, when the threshold at the reconstruction stage is less than 1, the reconstruction can tolerate some incompleteness in the glomeruli activation patterns. Real biological systems likely have the ability to adaptively change the activation threshold in order to balance the needs of high quality reconstruction and pattern completion.

Our model is shown to be very robust and fault tolerant, and this robustness is achieved

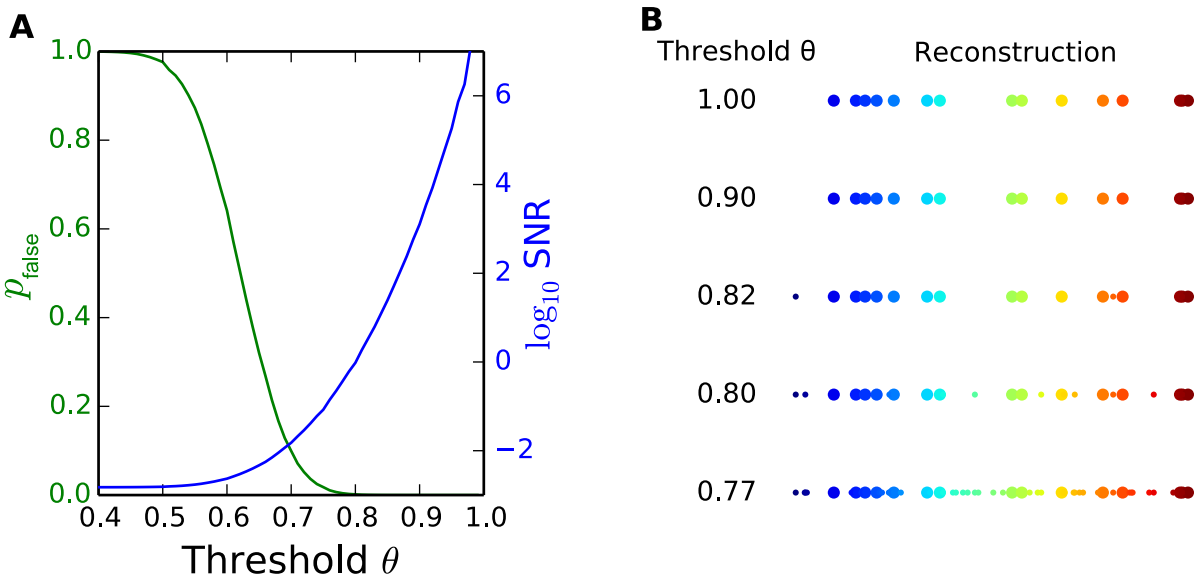

Figure 5.3: Demonstration of the accuracy-robustness trade-off. $N = 10000$, $K = 15$, $M = 1000$ and the optimal connectivity rate are used. (A) p_{false} and SNR for different activation thresholds at the reconstruction stage. With lower recovery thresholds, the robustness of the system to recovery noise increases, while the false detection rate increases, and the SNR of recovered signal decreases. (B) An example of the recovered signal with different recovery thresholds. True signal is shown in big colored dots, while the reconstruction error is represented by small colored dots. As we lower the threshold, the recovered signal becomes noisier.

with accuracy. As one can see, each glomerulus in the model only contains part of the information about the original signal. Because the measurement matrix A is random, no single glomerulus or cluster contains more or unique information, so any subset of the glomeruli could recover the original signal. The more glomeruli there are, the better recovery quality (SNR) can be achieved. Thus, removing or disabling part of the glomeruli will not change the system qualitatively, but will make the recovered signal more noisy, up to a point where noise becomes comparable to the true signal at which point the reconstruction fails. For a real biological system, it is reasonable to assume that the recovered signal has very high SNR, which also means high redundancy, as is observed experimentally.

5.9 Predicted optimal connectivity rate compared with experimental data

From our analysis we observed that for a given level of signal sparseness K, there is an optimal connectivity rate p_m that maximizes SNR as well as the mutual information. Assuming that the biological system is adapted to a given value of odor sparseness in its environmental niche, one can essentially make predictions on the connectivity rate of matrix A. This is followed by another prediction that the percentage of glomeruli activated by a single odorant should be close to the percentage of glomeruli that could activate a neuron in olfactory cortex or a Kenyon cell, and this number should be similar among species which operate in similar olfactory environments. The latter prediction should be easier to test, since the number of coexisting odorants in the environment is hard to measure.

Fortunately, previous experiments have gathered sufficient data to test our prediction indirectly. It has been shown that in Drosophila, 9% of the glomeruli have a strong response to an odorant [74], while the connectivity rate between glomeruli and Kenyon Cells is 6.5% [75] to 12.5% [76]. (The latter number is obtained based on the average number of claws per Kenyon

cell measured in [76]) These estimates are consistent with model predictions. Furthermore, in the locust, a typical projection neuron responds to about half of the odorants [77], while the connectivity rate between projection neurons and Kenyon Cell is also around 50% [78], which is also consistent with our prediction.

We can see that the connectivity rate is very different between species. Such differences can be unified in our model as the adaptation to different environmental niches. The locust has an anomalously high connectivity rate (50%), which in our model implies that its olfactory system is adapted to extreme odor sparseness tuned to odors with primarily a single component ($p_m = 0.5$ when $K = 1$). Similarly, Drosophila is adapted to sense odors composed of a mixture of about 10 odor components, while mice are tuned to detect a mixture of about 20 mono-molecular odors. In general, our model predicts that species with sparse connectivity will behave better in environments with complex odor mixtures, while species with dense connectivity have better performance in detecting simple odor mixtures.

5.10 Structural and functional evidence

In addition to the predictions above, further experimental evidence supports the structure of our model, in particular the approximate logical OR/AND operations associated with the compression/reconstruction stages, respectively. For example, it has been observed experimentally that Kenyon Cells in Drosophila receive convergent input from different glomeruli and require several inputs to be co-active to spike [79]. This is consistent with our threshold activation function which at the reconstruction stage uses a logical AND operation.

Functionally, experiments have shown that locust Kenyon cells are individually much better than projection neurons from glomeruli at detecting a single odorant; Kenyon cells that respond to an odorant also often respond to odor mixtures containing it [80]. This observation agrees with our assumption that each Kenyon cell only responses to one odorant and it will

respond when an odor mixture contains that odorant.

5.11 Stereotyped versus non-stereotyped connectivity

Since the affinity matrix A is determined genetically, all the connections in our model are predetermined before birth. There is some debate about such stereotypy versus random connectivity, and a compressed sensing model of olfaction based on random connections from glomeruli to mushroom body has been proposed[81]. Yet, our model supports both stereotyped and non-stereotyped projection from glomeruli to the mushroom body/olfactory cortex because the model is invariant under the exchange of neurons within the same layer. In order to verify such predetermination, one needs to obtain a detailed connectivity map from glomeruli to the mushroom body/olfactory cortex for different individuals, which is experimentally very challenging. An indirect approach to verify the predetermined connectivity hypothesis could be through an examination of innate behaviors that should depend primarily on predetermined connections. If one could relate innate behaviors to projections between glomeruli and the mushroom body/olfactory cortex, it would then provide additional supporting evidence for the genetically predetermined structural connectivity of the feedforward model.

5.12 Effective feedforward model for non-feedforward structure

The feedforward structure of our model is an effective approximation to the more complicated structure of biological olfactory system where recurrent and feedforward-feedback connections exist. For example, it has been observed that inhibitory interneurons modulate neuronal responses in the olfactory bulb[82, 83]. In linear dynamic systems, such feedforward-feedback structure could be mathematically modeled as a pure feedforward system with different

effective feedforward connectivity. Suppose that we add a layer of interneurons z in Fig. 5.1 that is connected to the glomeruli layer x by feedforward-feedback connectivity B. Then the linear dynamics of the system are $\dot{x} = -x + As^0 - B^T z$ and $\dot{z} = -z + Bx$, where we assume B is feedforward excitatory and feedback inhibitory. The steady state solution is $x = (I + B^T B)^{-1} As^0$, which is the same for a pure feedforward system, except that connectivity A is replaced by $(I + B^T B)^{-1} A$. This analysis is not exact if the activation function is nonlinear. In general, the feedforward-feedback system in steady state with a nonlinear activation function does not have an equivalent feedforward system, but one can still write the linear perturbation when neurons receive only weak inputs, which allows a feedforward approximation. Such a feedforward approximation is supported by experimental observations that the representations of odor mixtures in mouse glomeruli can be explained well by the summation of the glomeruli responses to their components [84].

One advantange of the effective feedforward model is that it enables an adaptive affinity matrix even with pre-determined connectivity. In the feedforward-feedback architecture mentioned above, the effective affinity matrix is $(I + B^T B)^{-1} A$, where A is the pre-determined affinity matrix encoded in the genes, while B could be a learned matrix adapted to the environment. From this perspective, the existence of interneurons in both insects and vertebrates[85, 86], as well as adult neurogenesis in the olfactory bulb of mammals[87], could play the role of adjusting the effective affinity matrix for the purpose of adaptation.

5.13 Comparison with ℓ_1 minimization algorithm

We compare the performance of our feedforward architecture with the often-used LASSO ℓ_1 minimization algorithm[88] provided by the Python scikit-learn library

$$\min_{\hat{s}} \frac{1}{2M} ||A\hat{s} - x||_2^2 + \beta ||\hat{s}||_1, \tag{5.19}$$

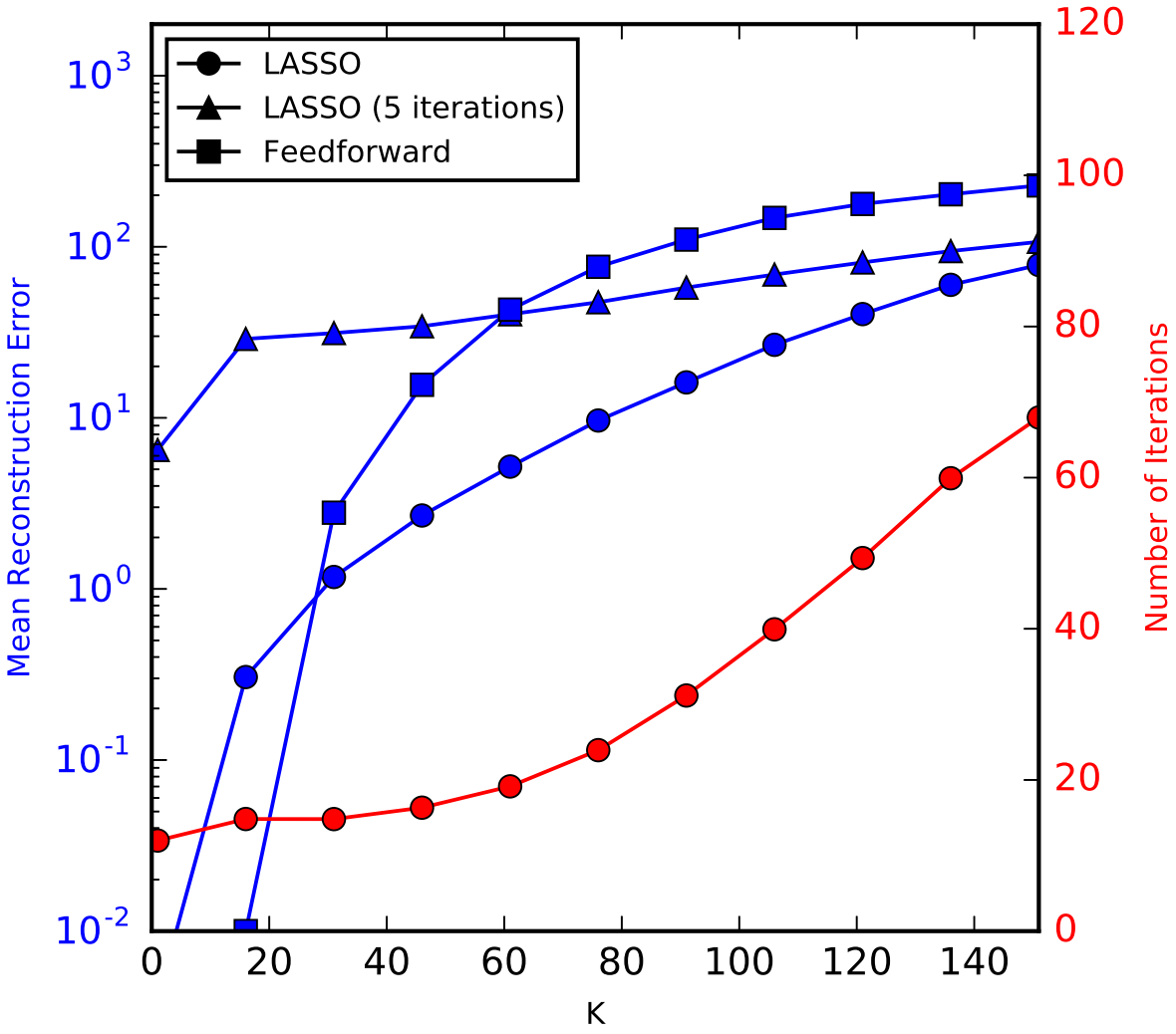

Figure 5.4: Comparison of the performance of feedforward architecture with that of LASSO. For this example, we chose $N = 1000$ and $M = 500$. Linear measurement is used for LASSO. Feedforward architecture performs well when the signal is very sparse, while LASSO has lower reconstruction error as K increases, at the price of increasingly more iterations. On the other hand, if we constrain the number of iterations, LASSO still performs better when K is large, but significantly worse with very sparse signals.

where $N = 1000$, $M = 500$, $\beta = 0.001$ are used. Linear measurement $x = As^0$ is used for LASSO. For each K, we conduct 100 experiments with different random measurement matrices and signals, and compute the average of the reconstruction errors $||\hat{s} - s^0||_1$ as well as the number of iterations used in LASSO. We also compute the mean reconstruction error when only 5 iterations are used in LASSO as a comparison. The results are shown in Fig. 5.4. As shown in the figure, the feedforward architecture has a lower reconstruction error when the signal is very sparse, while LASSO has a lower reconstruction error than the feedforward architecture when K becomes larger. However, the number of iterations also increases as the signal becomes denser. If we restrict the number of iterations to 5 in the LASSO (equivalent to setting a maximum response time), LASSO performs much worse when the signal is very sparse. But as K increases, it still has a lower reconstruction error than the feedforward architecture.

5.14 Performance with non-sparse signal

One drawback of this feedforward architecture is that it may not be able to achieve both compression and high-quality reconstruction simultaneously when the signal is not sparse. Unlike the ℓ_1 minimization method where the number of measurements required to reconstruct the signal will never exceed signal length N ($N/2$ for binary signal)[89, 90], the feedforward architecture may need more measurements than the signal length to accurately reconstruct the signal. This can be seen by restoring the term in Eq. 5.13 that we have previously neglected assuming that f is small

$$\log p_{\text{false}} = \log f - \log(1-f) - \log \text{SNR}. \tag{5.20}$$

Combining this with Eq. 5.14 that remains the same when f is not small, we obtain:

$$\alpha_{\text{SNR}} = ef \log \text{SNR} + ef \log(f^{-1} - 1), \tag{5.21}$$

which could be larger than 1 when f is not small. Thus, the feedforward computation may require number of measurements that are larger than the input dimensionality to achieve reliable reconstruction.

From another perspective, we can compute the upper bound on the reconstruction SNR that can be achieved for a given compression level. From Eq. 5.21 and $\alpha_{\text{SNR}} < 1$ we get

$$\log \text{SNR} < \frac{1}{ef} - \log(f^{-1} - 1), \tag{5.22}$$

which only depends on signal sparsity. For example, if $f = 0.1$, then SNR < 4.4, and the reconstructed signal will not be accurate.

5.15 Extension to continuous variables and other activation functions

Although our analysis above is based on a binary signal / measurement matrix / glomeruli activity and threshold activation function, our results can be extended to positive real-valued signal / measurement matrix / glomeruli activity and any monotonically increasing activation function. Consider the case where the signal s^0 and the element of measurement matrix A_{ij} could take any positive value rather than just 0 and 1. Denoting $x^l = As^0$, and letting the activation function g be any monotonically increasing function, the output at the glomerulus stage can be written as $x_i = g(x_i^l)$. Now, signal reconstruction can proceed based on the evaluation of a minimum function (rather than the logical AND function that was used in the case of binary inputs and binary measurement matrices). Indeed, when the ith component of the reconstructed signal $\hat{s}_i$ is computed as the smallest value $\{g^{-1}(x_j)/A_{ji}\}$ across the set of its inputs (i.e. where $A_{ji} \neq 0$), then our analysis remains valid. The only modification is that now the distribution of the signal and the measurement matrix elements are both required to compute the noise magnitude.

This procedure ensures that the recovered components are still recovered exactly, while corrupted components are still corrupted. As a practical aside, we note that the minimum function can be implemented by short-term synaptic plasticity, see Neural Implementation of Minimum Function Using Short-Term Plasticity.

5.16 Neural Implementation of Minimum Function Using Short-Term Plasticity

The following simulation shows that the Minimum function (output firing rate equals the smallest input firing rate) can be implemented in neural circuits with Short-Term Plasticity (STP).

We use conductance-based Leaky-Integrate-and-Fire (LIF) neuron model with exponentially decaying post-synaptic conductance. The dynamics of the system are given by [91]

$$\tau_m \frac{dv(t)}{dt} = v_{\text{rest}} - v(t) + g_{\text{ex}}(t)[v_{\text{rev}} - v(t)] + I(t)R \tag{5.23}$$

and

$$\tau_{\text{ex}} \frac{dg_{\text{ex}}(t)}{dt} = -g_{\text{ex}}(t) + w(t) g_{\text{peak}} \delta(t - t_{\text{spk}}), \tag{5.24}$$

where t_{spk} is the time when the input spike is received. Inhibitory postsynaptic potential is omitted since we only consider excitatory neurons. When $v > v_{\text{thres}}$, the neuron fires a spike and v is reset to v_{reset}. We use the following parameters: $v_{\text{thres}} = -50$ mV, $v_{\text{rest}} = v_{\text{reset}} = -65$ mV, $v_{\text{rev}} = 0$, $\tau_m = 15$ ms, $\tau_{\text{ex}} = 5$ ms, $g_{\text{peak}} = 0.015$, $I = 0$.

Synaptic weight $w(t)$ is described by Short-Term Plasticity. The dynamics are given by [92]

$$w(t) = r(t)u(t)W, \tag{5.25}$$

$$u(t_{n+1}) = (1 - U)u(t_n) \exp\left(\frac{t_n - t_{n+1}}{\tau_{\text{facil}}}\right) + U, \tag{5.26}$$

and

$$r(t_{n+1}) = r(t_n)[1-u(t_{n+1})]\exp\left(\frac{t_n - t_{n+1}}{\tau_{\text{rec}}}\right) + 1 - \exp\left(\frac{t_n - t_{n+1}}{\tau_{\text{rec}}}\right), \tag{5.27}$$

where t_n is the time of the n-th spike received, $U = 0.6$, $\tau_{\text{facil}} = 5$ ms, $\tau_{\text{rec}} = 100$ ms and W is the synaptic strength.

In the simulation, a LIF neuron receives 6 input spike trains with 6 STP synapses. Each spike train is generated with a fixed firing rate. A Gaussian noise with $\sigma = 1$ ms is added to each spike. The smallest input firing rate f_{min} is chosen uniformly from 0 to 100 Hz. The other 5 input firing rates are then chosen uniformly from f_{min} to 200 Hz. Initially $v = v_{\text{rest}}$, $g_{\text{ex}} = 0$, and for each synapse $u = 0$, $r = 1$. Each simulation runs for 2 s, with 1000 trials for each parameter. The results are shown in Fig. 5.5 and Fig. 5.6.

As shown in Fig. 5.5, the neuronal output firing rate can be well approximated by its minimal input firing rate when the minimal input firing rate is not high, while the median/mean/max of its input firing rates fail to do that due to the large variance in y-axis. In Fig. 5.6, we compare the results for different synaptic strength W. As we can see, such approximation is valid for a range of W. Larger W leads to better approximation when the minimal firing rate is high.

Chapter 5, in full, is a reprint of the material as it appears in PLoS computational biology. Zhang, Yilun; Sharpee, Tatyana, 2016. The dissertation author was the primary investigator and author of this paper.

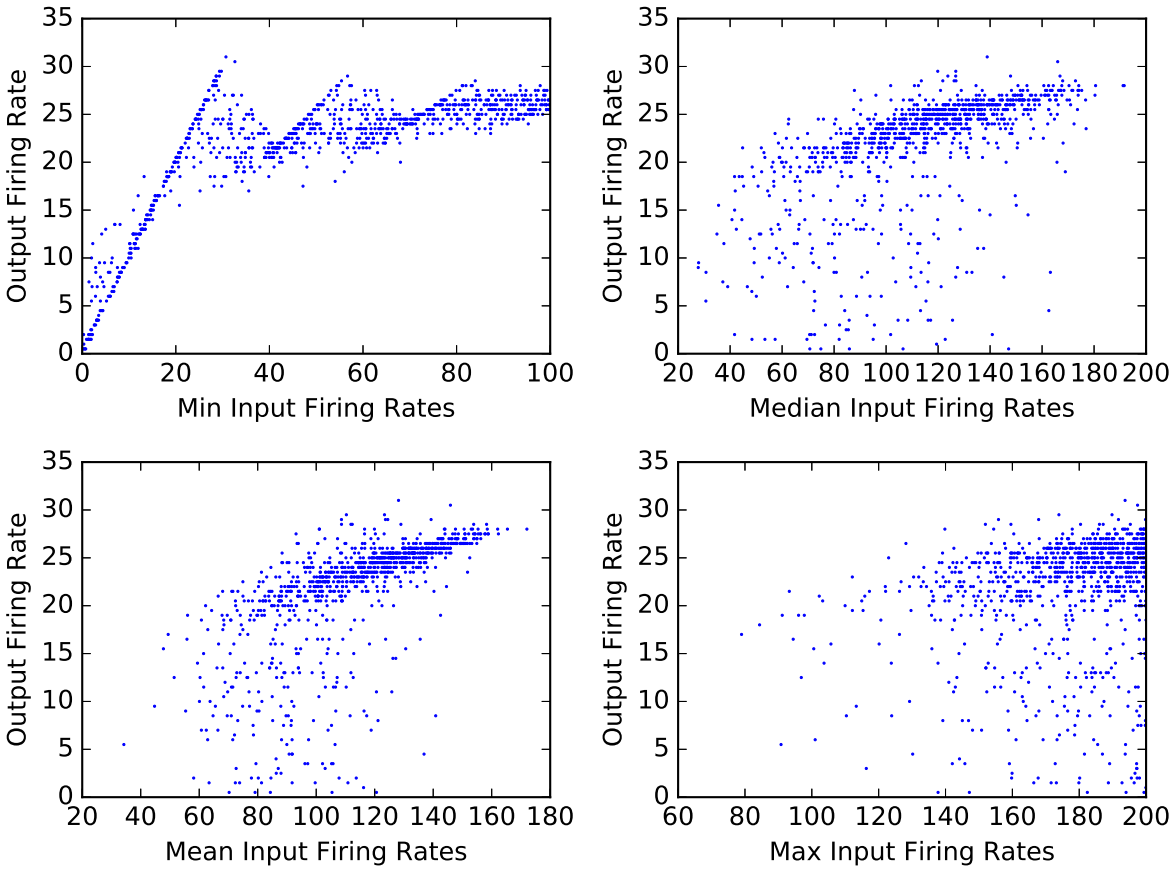

Figure 5.5: Comparison between the firing rate of the output neuron and the min/median/mean/max function applied to the firing rate of its inputs. $W = 74$ is used. As shown in the figure, provided the minimal input firing rate is not too large, the output firing rate can be well approximated by the minimal input firing rate. As a comparison, median/mean/max functions applied to the input firing rates are not well approximated by short-term plasticity, because the variance along the y-axis is large.

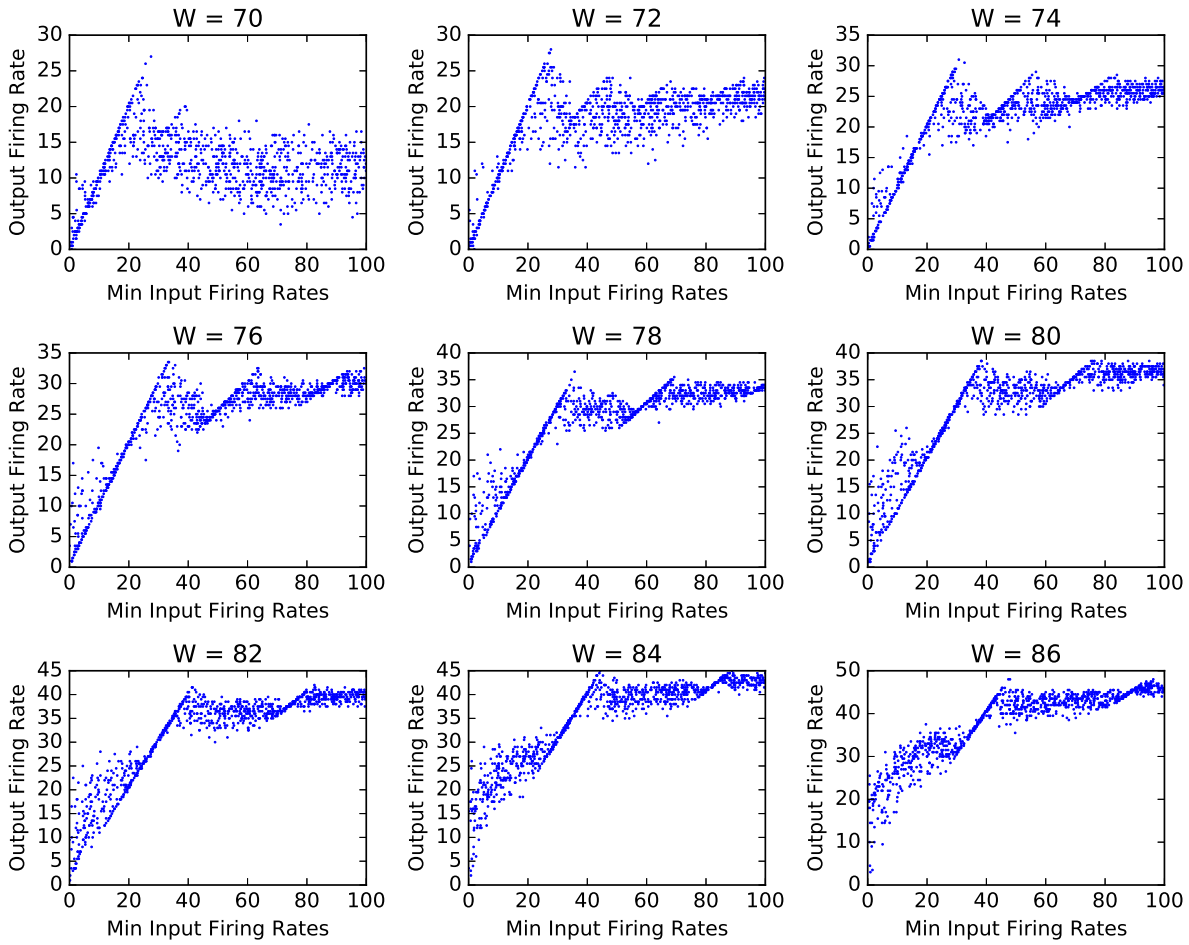

Figure 5.6: Comparison between the firing rate of the output neuron and the smallest input firing rate for different synaptic strength W. We can see that there is a range of W where the neuron output firing rate can be well approximated by its minimal input firing rate provided the latter one is not too large.

Bibliography

[1] Horace B Barlow. Possible principles underlying the transformation of sensory messages. *Sensory communication*, pages 217–234, 1961.

[2] F. Attneave. Some informational aspects of visual perception. *Psychol Rev*, 61:183–193, 1954.

[3] Ralph Linsker. Self-organization in a perceptual network. *Computer*, 21(3):105–117, 1988.

[4] Simon Laughlin. A simple coding procedure enhances a neuron's information capacity. *Zeitschrift für Naturforschung c*, 36(9-10):910–912, 1981.

[5] Joseph J Atick and A Norman Redlich. Towards a theory of early visual processing. *Neural Computation*, 2(3):308–320, 1990.

[6] Joseph J Atick and A Norman Redlich. What does the retina know about natural scenes? *Neural computation*, 4(2):196–210, 1992.

[7] Z. Li and J. J. Atick. Efficient stereo coding in the multiscale representation. *Network*, 5:157–174, 1994.

[8] Y. Dan, J. J. Atick, and R. C. Reid. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *J. Neurosci*, 16:3351–3362, 1996.

[9] A. J. Bell and T. J. Sejnowski. The "independent components" of natural scenes are edge filters. *Vision Res.*, 23:3327–3338, 1997.

[10] David B Kastner, Stephen A Baccus, and Tatyana O Sharpee. Critical and maximally informative encoding between neural populations in the retina. *Proceedings of the National Academy of Sciences*, 112(8):2533–2538, 2015.

[11] Claude E Shannon and Warren Weaver. The mathematical theory of information. 1949.

[12] D Gabor. Theory of communication. *Institution of Electrical Engineers*, 93 (Part III):429–457, 1946.

[13] Yan Karklin and Eero P Simoncelli. Efficient coding of natural images with a population of noisy linear-nonlinear neurons. *Advances in neural information processing systems*, 24:999, 2011.

[14] Julijana Gjorgjieva, Haim Sompolinsky, and Markus Meister. Benefits of pathway splitting in sensory coding. *The Journal of Neuroscience*, 34(36):12127–12144, 2014.

[15] Alexander P Nikitin, Nigel G Stocks, Robert P Morse, and Mark D McDonnell. Neural population coding is optimized by discrete tuning curves. *Physical review letters*, 103(13):138101, 2009.

[16] H. Wassle. Parallel processing in the mammalian retina. *Nat. Rev. Neurosci.*, 5:747–757, 2004.

[17] David B Kastner and Stephen A Baccus. Coordinated dynamic encoding in the retina using opposing forms of plasticity. *Nature neuroscience*, 14(10):1317–1322, 2011.

[18] Vijay Balasubramanian and Peter Sterling. Receptive fields and functional architecture in the retina. *The Journal of physiology*, 587(12):2753–2767, 2009.

[19] Nicolas Brunel and J-P Nadal. Mutual information, fisher information, and population coding. *Neural Computation*, 10(7):1731–1757, 1998.

[20] Xue-Xin Wei and Alan A Stocker. Mutual information, fisher information, and efficient coding. *Neural computation*, 28(2):305–326, 2016.

[21] Wentao Huang and Kechen Zhang. Information-theoretic bounds and approximations in neural population coding. *Neural computation*, (Early Access):1–60, 2018.

[22] David Barber and Felix Agakov. The im algorithm: a variational approach to information maximization. In *Proceedings of the 16th International Conference on Neural Information Processing Systems*, pages 201–208. MIT Press, 2003.

[23] Felix Agakov and David Barber. Variational information maximization for neural coding. In *Neural Information Processing*, pages 543–548. Springer, 2004.

[24] Shakir Mohamed and Danilo Jimenez Rezende. Variational information maximisation for intrinsically motivated reinforcement learning. In *Advances in neural information processing systems*, pages 2125–2133, 2015.

[25] John K Douglass, Lon Wilkens, Eleni Pantazelou, and Frank Moss. Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance. *Nature*, 365(6444):337–340, 1993.

[26] JJ Collins, Carson C Chow, and Thomas T Imhoff. Stochastic resonance without tuning. *Nature*, 376(6537):236–238, 1995.

[27] Luca Gammaitoni. Stochastic resonance and the dithering effect in threshold physical systems. *Physical Review E*, 52(5):4691, 1995.

[28] Peter Jung. Stochastic resonance and optimal design of threshold detectors. *Physics Letters A*, 207(1-2):93–104, 1995.

[29] Bart Kosko and Sanya Mitaim. Stochastic resonance in noisy threshold neurons. *Neural networks*, 16(5):755–761, 2003.

[30] Sanya Mitaim and Bart Kosko. Adaptive stochastic resonance for noisy threshold neurons based on mutual information. In *Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on*, volume 2, pages 1980–1985. IEEE, 2002.

[31] NG Stocks. Information transmission in parallel threshold arrays: Suprathreshold stochastic resonance. *Physical Review E*, 63(4):041114, 2001.

[32] Mark D McDonnell, Nigel G Stocks, Charles EM Pearce, and Derek Abbott. Optimal information transmission in nonlinear arrays through suprathreshold stochastic resonance. *Physics Letters A*, 352(3):183–189, 2006.

[33] P. Ala-Laurila, M. Greschner, E. J. Chichilnisky, and F. Rieke. Cone photoreceptor contributions to noise and correlations in the retinal output. *Nat Neurosci*, 14(10):1309–1316, 2011.

[34] F. Rieke and D. A. Baylor. Origin and functional impact of dark noise in retinal cones. *Neuron*, 26:181–186, 2000.

[35] B. G. Borghuis, P. Sterling, and R. G. Smith. Loss of sensitivity in an analog neural circuit. *J Neurosci.*, 29:3045–3058, 2009.

[36] T. M. Cover and J. A. Thomas. *Information theory*. John Wiley & Sons, INC., New York, 1991.

[37] John Berkowitz and Tatyana Sharpee. Decoding neural responses with minimal information loss. *bioRxiv*, page 273854, 2018.

[38] Amadeus Dettner, Sabrina Münzberg, and Tatjana Tchumatchenko. Temporal pairwise spike correlations fully capture single-neuron information. *Nature communications*, 7:13805, 2016.

[39] Braden AW Brinkman, Alison I Weber, Fred Rieke, and Eric Shea-Brown. How do efficient coding strategies depend on origins of noise in neural circuits? *PLOS Comput Biol*, 12(10):e1005150, 2016.

[40] B. G. Borghuis, C. P. Ratliff, R. G. Smith, P. Sterling, and Vijay Balasubramanian. Design of a neuronal array. *J Neurosci.*, 28:3178–3189, 2008.

[41] S. H. Devries and D. A. Baylor. Mosaic arrangement of ganglion cell receptive fields in rabbit retina. *J Neurophysiol.*, 78:2048–2060, 1997.

[42] Y. S. Liu, C. F. Stevens, and T. O. Sharpee. Predictable irregularities in retinal receptive fields. *PNAS*, 106:16499–16504, 2009.

[43] G. D. Field and E. J. Chichilnisky. Information processing in the primate retina: Circuitry and coding. *Ann. Rev. Neurosci.*, 30:1–30, 2007.

[44] G. D. Field, A. Sher, J. L. Gauthier, M. Greschner, J. Shlens, A. M. Litke, and E. J. Chichilnisky. Spatial properties and functional organization of small bistratified ganglion cells in primate retina. *J Neurosci.*, 27:11261–13272, 2007.

[45] David J Field. Relations between the statistics of natural images and the response properties of cortical cells. *JOSA A*, 4(12):2379–2394, 1987.

[46] D. W. Dong and J. J. Atick. Statistics of natural time-varying images. *Network: Comput. Neural Syst.*, 6:345–358, 1995.

[47] L. C. Sincich, J. C. Horton, and T. O. Sharpee. Preserving information in neural transmission. *J. Neurosci.*, 29:6207–6216, 2009.

[48] Kareem A Zaghloul, Kwabena Boahen, and Jonathan B Demb. Different circuits for on and off retinal ganglion cells cause different contrast sensitivities. *The Journal of Neuroscience*, 23(7):2645–2654, 2003.

[49] A. C. Aho, K. Donner, C. Hyden, L. O. Larsen, and T. Reuter. Low retinal noise in animals with low body temperatyure allows high visual sensitivity. *Nature*, 334:348–350, 1988.

[50] Greg D Field and Fred Rieke. Mechanisms regulating variability of the single photon responses of mammalian rod photoreceptors. *Neuron*, 35(4):733–747, 2002.

[51] Claude E Shannon. Communication theory of secrecy systems. *Bell Labs Technical Journal*, 28(4):656–715, 1949.

[52] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical review E*, 69(6):066138, 2004.

[53] Paul Maragakis, Felix Ritort, Carlos Bustamante, Martin Karplus, and Gavin E Crooks. Bayesian estimates of free energies from nonequilibrium work data in the presence of instrument noise. *The Journal of chemical physics*, 129(2):07B609, 2008.

[54] Vijay Balasubramanian, Don Kimber, and Michael J Berry II. Metabolically efficient information processing. *Neural Computation*, 13(4):799–815, 2001.

[55] Avery N Gilbert. *What the nose knows: the science of scent in everyday life*. Crown Publishers New York, 2008.

[56] Thomas Hummel and Antje Welge-Lüssen. *Taste and smell: An update*, volume 63. Karger Medical and Scientific Publishers, 2006.

[57] Eric R Kandel, James H Schwartz, Thomas M Jessell, Steven A Siegelbaum, and AJ Hudspeth. Principles of neural science, 2014.

[58] Emmanuel J Candes and Terence Tao. Decoding by linear programming. *Information Theory, IEEE Transactions on*, 51(12):4203–4215, 2005.

[59] Sina Tootoonian and Mate Lengyel. A dual algorithm for olfactory computation in the locust brain. In *Advances in Neural Information Processing Systems*, pages 2276–2284, 2014.

[60] Christopher J Rozell, Don H Johnson, Richard G Baraniuk, and Bruno A Olshausen. Sparse coding via thresholding and local competition in neural circuits. *Neural computation*, 20(10):2526–2563, 2008.

[61] Surya Ganguli and Haim Sompolinsky. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annual review of neuroscience*, 35:485–508, 2012.

[62] Céline Jouquand, Craig Chandler, Anne Plotto, and Kevin Goodner. A sensory and chemical analysis of fresh strawberries over harvest dates and seasons reveals factors that affect eating quality. *Journal of the American Society for Horticultural Science*, 133(6):859–867, 2008.

[63] Marie Bengtsson, Anna-Carin Bäckman, Ilme Liblikas, Martha I Ramirez, Anna-Karin Borg-Karlson, Lena Ansebo, Peter Anderson, Jan Löfqvist, and Peter Witzgall. Plant odor analysis of apple: antennal response of codling moth females to apple volatiles during phenological development. *Journal of agricultural and food chemistry*, 49(8):3736–3741, 2001.

[64] Miriam Frida Karlsson, Goran Birgersson, Alba Marina Cotes Prado, Felipe Bosa, Marie Bengtsson, and Peter Witzgall. Plant odor analysis of potato: response of guatemalan moth to above-and belowground potato volatiles. *Journal of agricultural and food chemistry*, 57(13):5903–5909, 2009.

[65] Alexander Fleischmann, Benjamin M Shykind, Dara L Sosulski, Kevin M Franks, Meredith E Glinka, Dan Feng Mei, Yonghua Sun, Jennifer Kirkland, Monica Mendelsohn, Mark W Albers, and Richard Axel. Mice with a monoclonal nose: perturbations in an olfactory map impair odor discrimination. *Neuron*, 60(6):1068–1081, 2008.

[66] Surya Ganguli and Haim Sompolinsky. Statistical mechanics of compressed sensing. *Physical review letters*, 104(18):188701, 2010.

[67] Kathleen M Guthrie, Jacquelyn Holmes, and Michael Leon. Focal 2-dg uptake persists following olfactory bulb lesions. *Brain research bulletin*, 38(2):129–134, 1995.

[68] X-CM Lu and BM Slotnick. Olfaction in rats with extensive lesions of the olfactory bulbs: implications for odor coding. *Neuroscience*, 84(3):849–866, 1998.

[69] Kathleen McBride, Burton Slotnick, and Frank L Margolis. Does intranasal application of zinc sulfate produce anosmia in the mouse? an olfactometric and anatomical study. *Chemical senses*, 28(8):659–670, 2003.

[70] Antoniu L Fantana, Edward R Soucy, and Markus Meister. Rat olfactory bulb mitral cells receive sparse glomerular inputs. *Neuron*, 59(5):802–814, 2008.

[71] Naoshige Uchida, Yuji K Takahashi, Manabu Tanifuji, and Kensaku Mori. Odor maps in the mammalian olfactory bulb: domain organization and odorant structural features. *Nature neuroscience*, 3(10):1035–1043, 2000.

[72] Matt Wachowiak and Lawrence B Cohen. Representation of odorants by receptor neuron input to the mouse olfactory bulb. *Neuron*, 32(4):723–735, 2001.

[73] Adil Ghani Khan, K Parthasarathy, and Upinder Singh Bhalla. Odor representations in the mammalian olfactory bulb. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 2(5):603–611, 2010.

[74] Elissa A Hallem and John R Carlson. Coding of odors by a receptor repertoire. *Cell*, 125(1):143–160, 2006.

[75] Sophie JC Caron, Vanessa Ruta, LF Abbott, and Richard Axel. Random convergence of olfactory inputs in the drosophila mushroom body. *Nature*, 497(7447):113–117, 2013.

[76] Nancy J Butcher, Anja B Friedrich, Zhiyuan Lu, Hiromu Tanimoto, and Ian A Meinertzhagen. Different classes of input and output neurons reveal new features in microglomeruli of the adult drosophila mushroom body calyx. *Journal of Comparative Neurology*, 520(10):2185–2201, 2012.

[77] Javier Perez-Orive, Ofer Mazor, Glenn C Turner, Stijn Cassenaer, Rachel I Wilson, and Gilles Laurent. Oscillations and sparsening of odor representations in the mushroom body. *Science*, 297(5580):359–365, 2002.

[78] Ron A Jortner, S Sarah Farivar, and Gilles Laurent. A simple connectivity scheme for sparse coding in an olfactory system. *The Journal of neuroscience*, 27(7):1659–1669, 2007.

[79] Eyal Gruntman and Glenn C Turner. Integration of the olfactory code across dendritic claws of single mushroom body neurons. *Nature neuroscience*, 16(12):1821–1829, 2013.

[80] Kai Shen, Sina Tootoonian, and Gilles Laurent. Encoding of mixtures in a simple olfactory system. *Neuron*, 80(5):1246–1262, 2013.

[81] Charles F Stevens. What the flys nose tells the flys brain. *Proceedings of the National Academy of Sciences*, 112(30):9460–9465, 2015.

[82] Shawn R Olsen and Rachel I Wilson. Lateral presynaptic inhibition mediates gain control in an olfactory circuit. *Nature*, 452(7190):956–960, 2008.

[83] Pierre-Marie Lledo, Armen Saghatelyan, and Morgane Lemasson. Inhibitory interneurons in the olfactory bulb: from development to function. *The Neuroscientist*, 10(4):292–303, 2004.

[84] Stephen D Shea and Lawrence C Katz. Representation of natural stimuli in the rodent main olfactory bulb. *Neuron*, 50(6):937–949, 2006.

[85] Ya-Hui Chou, Maria L Spletter, Emre Yaksi, Jonathan CS Leong, Rachel I Wilson, and Liqun Luo. Diversity and wiring variability of olfactory local interneurons in the drosophila antennal lobe. *Nature neuroscience*, 13(4):439–449, 2010.

[86] Gordon M Shepherd, Wei R Chen, David Willhite, Michele Migliore, and Charles A Greer. The olfactory granule cell: from classical enigma to central role in olfactory processing. *Brain research reviews*, 55(2):373–382, 2007.

[87] Françoise Lazarini and Pierre-Marie Lledo. Is adult neurogenesis essential for olfaction? *Trends in neurosciences*, 34(1):20–30, 2011.

[88] Emmanuel J Candè and Michael B Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30, 2008.

[89] David Donoho and Galen Reeves. The sensitivity of compressed sensing performance to relaxation of sparsity. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 2211–2215. IEEE, 2012.

[90] Mihailo Stojnic. Recovery thresholds for l1 optimization in binary compressed sensing. In *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pages 1593–1597. IEEE, 2010.

[91] Anthony N Burkitt. A review of the integrate-and-fire neuron model: I. homogeneous synaptic input. *Biological cybernetics*, 95(1):1–19, 2006.

[92] Henry Markram, Yun Wang, and Misha Tsodyks. Differential signaling via the same axon of neocortical pyramidal neurons. *Proceedings of the National Academy of Sciences*, 95(9):5323–5328, 1998.