The optimal odor-receptor interaction network is sparse in olfactory systems: Compressed sensing by nonlinear neurons with a finite dynamic range

Shanshan Qin[1],* Qianyi Li[2],* Chao Tang[1,3,4],† and Yuhai Tu[5]‡
[1] *Center for Quantitative Biology, Peking University, Beijing, 100871, China*
[2] *Yuanpei College, Peking University, Beijing, 100871, China*
[3] *School of Physics, Peking University, Beijing 100871, China*
[4] *Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China*
[5] *IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, USA*
(Dated: March 10, 2019)

There are numerous different odorant molecules in nature but only a relatively small number of olfactory receptor neurons (ORNs) in brains. This "compressed sensing" challenge is compounded by the constraint that ORNs are nonlinear sensors with a finite dynamic range. Here, we investigate possible optimal olfactory coding strategies by maximizing mutual information between odor mixtures and ORNs' responses with respect to the bipartite odor-receptor interaction network (ORIN) characterized by sensitivities between all odorant-ORN pairs. We find that the optimal ORIN is sparse – a finite fraction of sensitives are zero, and the nonzero sensitivities follow a broad distribution that depends on the odor statistics. We show that the optimal ORIN enhances performances of downstream learning tasks (reconstruction and classification). For ORNs with a finite basal activity, we find that having a basal-activity-dependent fraction of inhibitory odor-receptor interactions increases the coding capacity. All our theoretical findings are consistent with existing experiments and predictions are made to further test our theory. The optimal coding model provides a unifying framework to understand the peripheral olfactory systems across different organisms.

I. INTRODUCTION

Animals rely on their olfactory systems to detect, discriminate and interpret external odor stimuli to guide their behavior. Natural odors are typically mixtures of different odorant molecules whose concentrations can vary over several orders of magnitude [1–3]. Remarkably, animals can distinguish a large number of odorants and their mixtures by using a relatively small number of odor receptors (ORs) [4, 5]. For example, humans have only $\sim$ 300 ORs [6, 7], and the often cited number of odors that can be distinguished is $\sim$ 10000 [8], the real number may be even larger [9] (see also [10] and [11]). Humans can also distinguish odor mixtures with up to 30 different compounds [12]. In comparison, the highly olfactory lifestyle and exquisite olfactory learning ability of the fly is afforded by only $\sim$ 50 ORs [4, 13]. The olfactory system achieves such remarkable ability through a combinatorial code in which each odorant is sensed by multiple receptors and each receptor can be activated by many odorants [14–16]. In both mammals and insects, odorants bind to receptors in the cilia or dendrites of olfactory receptor neurons (ORNs), each of which expresses only one type of receptor. ORNs that express the same receptors then converge onto the same glomerulus in olfactory bulb (mammals) or antennal lobe (insects), whose activity patterns contain the information about external odor stimuli [13, 17–19]. A key question that we want to address here is how ORNs best represent external olfactory information that can be interpreted by the brain to guide an animal's behavior [4, 13, 20].

It has long been hypothesized that the input-output response functions of sensory neurons are "selected" by statistics of the stimuli in the organism's natural environment to transmit a maximum amount of information about its environment, generally known as the efficient coding hypothesis [21, 22] or the related InfoMax principle [23, 24]. For instance, the contrast-response function of interneurons in the fly's compound eye can be well approximated by the cumulative probability distribution of contrast in the natural environment [22]. The receptive fields of neurons in early visual pathway are thought to exploit statistics of natural scenes [25–29]. Similar result has also been observed in the auditory system [30]. In all these cases, to achieve maximum information transmission, an "ideal" neuron should transform the input distribution into a uniform output distribution [22, 31] and a population of neurons should decorrelate their responses [28, 32, 33].

However, unlike light or sound, which can be characterized by a single quantity such as wavelength or frequency, there are a huge number of odorants each with its own unique molecular structure and different physiochemical properties [34, 35]. The high dimensionality of the odor space thus poses a severe challenge for the olfactory system to code olfactory signals. Fortunately, typical olfactory stimuli are sparse with only a few types of odorant molecules in an odor mixture [1, 2, 36]. The sparsity of the odor mixture immediately reminds us of the powerful compressed sensing (CS) theory developed in computer sceince and signal processing community [37, 38]. The CS theory shows that sparse high dimensional signals can be encoded by a small number of sensors (measurements)

* These two authors contributed equally
† email address: tangc@pku.edu.cn; Corresponding author
‡ email address: yuhai@us.ibm.com; Corresponding author

through random projections; and the highly compressed signal can be reconstructed (decoded) with high fidelity by using an L_1-minimization algorithm [37–39]. However, conventional CS theory assumes the sensors to have a linear response function with essentially an infinite dynamic range [40]. In contrast, ORN response is highly nonlinear [41, 42], with a typical dynamic range less than 2 orders of magnitude, which is far less than the typical concentration range of odorants [42].

The use of the CS theory has recently been explored in olfactory systems. For example, Zhang and Sharpee proposed a fast reconstruction algorithm in a simplified setup with binary ORNs and binary odor mixtures without concentration information [43]. In another work, Krishnamurphy *et al.* studied how the overall "hour-glass" (compression followed by decompression) structure of the olfactory circuit can facilitate olfactory association and learning, again with the assumption that ORN responses to odor mixtures are linear [44]. Following ideas in CS theory, Singh *et al* recently proposed a fast olfactory decoding algorithm that might be implemented in the downstream olfactory system [45]. However, All these studies primarily focus on the downstream decoding and learning of the compressed signals by assuming a linear neuron response function. The question of how ORNs with a nonlinear input-output response function can best compress the sparse high dimensional odor information remains unanswered.

Another related study is the recent work by Zwicker *et al.* [46] where the authors investigated the maximum entropy coding scheme for the olfactory system by using a simplified binary response function, where an odor only induces a response when its concentration is above a threshold that is inversely proportional to the receptor sensitivity to the odor. They found two conditions for the binary ORNs to maximize the information transmission. The first condition is that each ORN on average responses to half of the odors, i.e., half of the odors have a concentration that is higher than the corresponding threshold; and the other condition is that the responses from different ORNs need to be uncorrelated. These results were obtained by studying the average activities of the ORNs and their correlations. However, due to the limitation of the binary input-output response function and the specific prior for the sensitivity distribution used in [46], the optimal coding strategy for neurons with realistic physiological properties remains unclear.

To address these important open questions, here we study the optimal coding scheme by using a realistic ORN input-output response function where the ORN output depends on the odor concentration continuously in a nonlinear (sigmoidal with odor concentration on a logarithmic scale) form characterized by its sensitivity, or equivalently the inverse of the half-maximum response concentration. By optimizing the input-output mutual information in the full sensitivity matrix space without any prior and following general insights from compressed sensing for sparse odor mixtures, we systematically study the statistical properties of the optimal sensitivity matrix and their dependence on odor statistics.

We found that the optimal ORN sensitivity matrix is sparse, i.e., each receptor only responds to a fraction of the odorants in its environment and its sensitivity to the other odorants is *zero*. The sparsity itself is a robust (universal) feature of the optimal sensitivity matrix, and the value of the optimal sparsity depends on statistics of the odor mixture and the number of ORNs. By using a simple mean-field theory, we show that the sensitivity matrix sparsity is caused by the competition (trade-off) between enhancing multiple odor detection and avoiding odor-odor interference. Next, we demonstrate the advantages of the maximum entropy (information) coding scheme in two downstream "decoding" (learning) tasks: reconstruction and classification of odor stimuli. Finally, we generalize our theory to neurons with a finite basal activity, where we found that the optimal coding strategy is to allow co-existence of both odor-evoked inhibition and activation with the fraction of inhibitory interactions depending on the basal activity. Comparisons with existing data in different organisms are consistent with our theory, which provides a unified framework to understand olfactory coding. Possible predictions based on our theory and future directions to incorporate more biological complexities such as odor-odor and receptor-receptor correlations in our model are also discussed.

II. RESULTS

We first describe the mathematical setup of the problem before presenting the results. An odor mixture can be represented as a vector $\boldsymbol{c} = (c_1, \ldots, c_N)$, where c_j is the concentration of odorant (ligand) $j(= 1, 2, ..., N)$ and N is the number of all possible odorants in the environment. A typical odor mixture is sparse with only $n(\ll N)$ odorant molecules that have non-zero concentrations. As illustrated in Fig. 1(a) (dotted box), the odor mixture signal $\boldsymbol{c}$ is sensed by M sensors. The encoding process, which maps $\boldsymbol{c}$ to the ORN response vector $\boldsymbol{r} = (r_1, r_2, ..., r_M)$, is determined by the bipartite odorant-ORN interaction network characterized by the sensitivity matrix $\boldsymbol{W}$, whose elements are denoted as W_{ij}, the sensitivity of i-th sensor (ORN) to j-th odorant for all odorant-ORN pairs with $j = 1, 2, ..., N$ and $i = 1, 2, ..., M$. For simplicity, we used a simple competitive binding model [47, 48] [and see Supplemental Material (SM)], in which the normalized response of ORN i $(= 1, 2, \cdots, M)$ to odor $\boldsymbol{c}$ can be described by a nonlinear function

$$r_i = F_i(\boldsymbol{c}, \boldsymbol{W}) + \eta_i = \frac{\sum_j^N W_{ij} c_j}{1 + \sum_j^N W_{ij} c_j} + \eta_i, \qquad (1)$$

where F_i is the response function which is taken to be the form in Eq. 1, and η_i represents the noise term. For convenience, we assume η_i is a Gaussian noise with zero

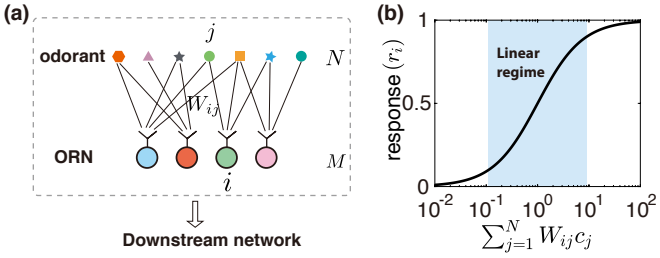

FIG. 1. Schematics of peripheral odor coding. (a) Illustration of interaction between odorants and odor receptors. There are N possible odorant molecules and M ORNs. The interaction between the i'th ORN and the j'th odorant is characterized by the sensitivity W_{ij}. The odorant-ORN interactions in the peripheral olfactory system (the dotted box) are characterized by the $(M \times N)$ sensitivity matrix $\boldsymbol{W}$. The odor information collected by the ORNs are passed on to downstream network for further processing. (b) Typical ORN-odorant dose-response curve according to Eq. (1). The range of linear response (in the log scale) is highlighted by the shaded area.

mean and standard deviation σ_0. Other forms of the nonlinear response function and noise can be used without affecting the general conclusions of our paper.

As illustrated in Fig. 1(b), the input-output response curve is highly nonlinear (Sigmoidal) resulting in a finite response range for each sensor, which is less than the range of concentration for a typical odorant molecule. Therefore, to encode the full concentration range of an odorant molecule, an odorant needs to interact with multiple sensors with different sensitivities. On the other hand, given the fact that $M < N$, each sensor has to sense multiple odorant molecules. These two considerations lead to the many-to-many odor-receptor interaction network characterized by the sensitivity matrix $\boldsymbol{W} = \{W_{ij} |\ i = 1, 2, ..., M;\ j = 1, 2, ..., N\}$.

Eq. (1) maps the external odor stimulus $\boldsymbol{c}$ to the internal neuronal activity $\boldsymbol{r} = (r_1, r_2, \cdots, r_M)$. The downstream olfactory circuits then use this response pattern to evaluate (decode) odor information (both their identities and concentrations) in order to guide the animal's behaviors. The quality of encoding odor information by the periphery ORNs directly sets the upper limit of how well the brain can decode the odor information [49]. In this paper, we focus on discovering the key statistical properties of the sensitivity matrix $\boldsymbol{W}$ that allows biologically realistic ORNs to best represent the external odor information.

A given odor environment can be generally described by a probability distribution $P_{\text{env}}(\boldsymbol{c})$. To convey maximum information about external odor stimuli in their response patterns to the brain, ORs/ORNs can adjust their sensitivity matrix $\boldsymbol{W}$ to "match" odor statistics $P_{\text{env}}(\boldsymbol{c})$. Without any assumption on what information the brain may need, the mutual information $I(\boldsymbol{c}, \boldsymbol{r})$ between stimuli and response pattern of ORNs sets the limit on how much odor information are received by the peripheral ORNs and thus serves as a good "target function" to be maximized [23, 24, 46, 50–52]. I is defined as

$$I(\boldsymbol{c}, \boldsymbol{r}) = H(\boldsymbol{r}) - H(\boldsymbol{r}|\boldsymbol{c}) = -\int P_r(\boldsymbol{r}) \log P_r(\boldsymbol{r}) d\boldsymbol{r} + \int P(\boldsymbol{r}|\boldsymbol{c}) \log P(\boldsymbol{r}|\boldsymbol{c}) d\boldsymbol{c}, \tag{2}$$

where $H(\boldsymbol{r})$ and $H(\boldsymbol{r}|\boldsymbol{c})$ are the entropy of output distribution $P_r(\boldsymbol{r})$ and conditional distribution $P(\boldsymbol{r}|\boldsymbol{c})$. In the limit of small noise, the second term is independent of $\boldsymbol{W}$ and negligible, hence, we will use $H(\boldsymbol{r})$ as our target function for optimization.

$H(\boldsymbol{r})$ depends on $\boldsymbol{W}$ and $P_{\text{env}}(\boldsymbol{c})$ because $P_r(\boldsymbol{r})$ depends on $\boldsymbol{W}$ and $P_{\text{env}}(\boldsymbol{c})$:

$$P_r(\boldsymbol{r}) = \int (2\pi\sigma_0^2)^{-\frac{M}{2}} \exp[-\sum_{i=1}^{M} \frac{(r_i - F_i(\boldsymbol{c}, \boldsymbol{W}))^2}{2\sigma_0^2}] P_{\text{env}}(\boldsymbol{c}) d\boldsymbol{c}. \tag{3}$$

In this paper, we study the optimal coding strategy by maximizing the mutual information I or equivalently the differential entropy H with respect to the sensitivity matrix $\boldsymbol{W}$ for different odor mixture statistics $P_{\text{env}}(\boldsymbol{c})$ and different numbers of ORNs. The mutual information as given in Eq. (2) can only be computed analytically for very simple cases. For more general cases, we used the covariance matrix adaptation - evolution strategy (CMA-ES) algorithm to find the optimal sensitivity matrix [53, 54] (see SM for technical details).

A. The optimal sensitivity matrix is sparse

Odor concentration varies widely in natural environment [1, 55]. To capture this property, we studied the case where the odorant concentrations in an odor mixture follows a log-normal distribution with variance σ_c^2. Other broad distributions such as power law distributions are also studied without changing the general conclusions. For simplicity, we consider the case where odorants appear independently in the mixture; more realistic consideration such as correlation among odorants will be discussed later in the Discussion section.

For given odor statistics (characterized by N, n and σ_c), and a given number of nonlinear sensors M, we can compute and optimize the input-output mutual information $I(\boldsymbol{W}|N, n, \sigma_c; M)$ with respect to all the $M \times N$ elements in the sensitivity matrix $\boldsymbol{W}$. We found that the optimal sensitivity matrix $\boldsymbol{W}$ is "sparse": only a fraction (ρ_w) of its elements have non-zero values [sensitive, shown as the colored elements in Fig. 2(a)], and the rest are insensitive [the black elements in Fig. 2(a)], with essentially zero values of W_{ij}. The sparsity in the optimal $\boldsymbol{W}$ was not found in the previous theoretical study [46] mainly due to the oversimplified binary ORN response function used there. From the histogram of $\ln(W_{ij})$ shown in Fig. 2(b), it is clear that elements in the optimal sensitivity matrix fall into two distinctive populations: the insensitive population that has practically

zero sensitivity [note the log scale used in Fig. 2(b)], and a sensitive population with a finite sensitivity. For the cases when the odor concentration follows a log-normal distribution, the distribution of the sensitive (nonzero) elements $P_s(w)$ can be fitted well with a log-normal distribution as shown in Fig. 2(b).

Our main finding here, i.e., sparsity in the odor-receptor sensitivity matrix, is supported by existing experimental measurements. As shown in Fig. 2(c), the sparsity ρ_w is estimated to be ~ 0.4 for fly larva [42] and ~ 0.1 for mouse [56]. Additionally, the broad distribution of the non-zero sensitivity observed in our model also agree qualitatively with those estimated from experiments [Fig. 2(c), two right panels], which are slightly skewed log-normal distributions.

Besides the distribution of the individual sensitivity matrix elements, we also calculated the row (sensor)-wise and column (odorant)-wise rank-order correlation coefficients (Kendall's tau, τ) and compared them with those from the same matrix but with its elements shuffled randomly. As shown in the Supplemental Material (SM), we found that both the rows and columns (Fig. S1 in SM) in the optimal matrix have a higher level of orthogonality (and thus independence) than that from random matrices. This orthogonality in the optimal $\boldsymbol{W}$ matrix leads to a higher input-output mutual information than those from the shuffled matrices [see Fig. S2(a) in SM] and a nearly uniform distribution of ORN activity for different odor mixtures [see Fig. S2(b)-(d) in SM].

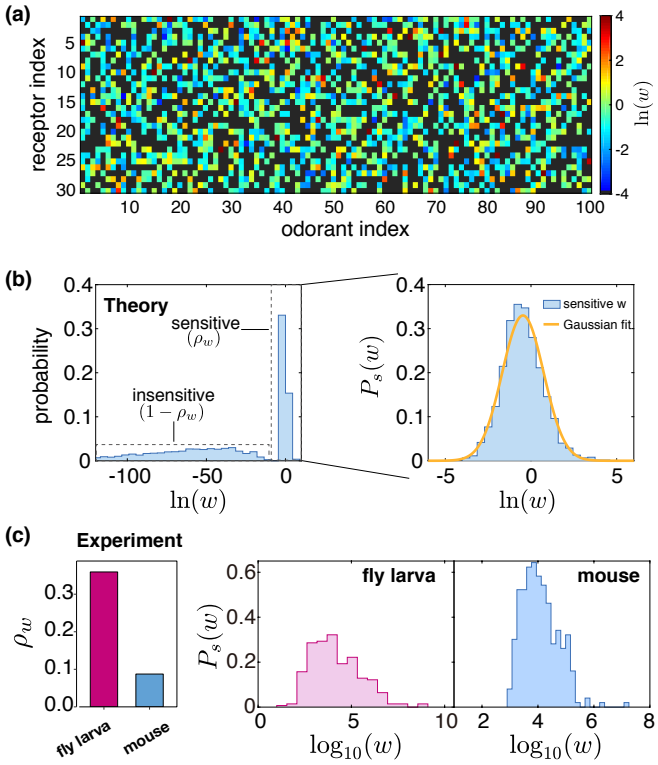

FIG. 2. Statistics of the optimal sensitivity matrix elements from theory and comparison with experiments. (a) Heatmap of a typical optimal sensitivity matrix from our model. Color indicates the value of $\ln(W_{ij})$, black indicates the "inactive" or negligible interactions. (b) Histogram of all the W_{ij} values from our model. It shows a bimodal distribution: an insensitive part with near zero W_{ij} and a sensitive part with nonzero W_{ij}. The distribution of the sensitive elements can be fitted by a log-normal distribution. (c) Experimental data from fly larva and mouse. Left: the fraction of sensitive odorant-receptor interactions ρ_w estimated in experiments for fly larva [42] and mouse [56]. Right: the histogram of sensitive W_{ij}, $P_s(w)$, for fly larva and mouse. Model parameters are $N = 100$, $M = 30$, $n = 2$, $\sigma_c = 2$, $\mu = 0$.

B. The optimal sparsity depends on odor statistics and the number of sensors

The statistics of the optimal sensitivity matrix elements are characterized by the sparsity ρ_w defined as the fraction of *non-zero* elements in $\boldsymbol{W}$, and the distribution of the sensitive (nonzero) elements, $P_s(w)$, which is further characterized by its mean (μ_w) and standard deviation (σ_w). Note that the sparsity parameter ρ_w is defined in such a way that a smaller value of ρ_w corresponds to a sparser sensitivity matrix. We investigated systematically how ρ_w, μ_w, and σ_w depend on statistical properties of the odor mixture characterized by N, n, and σ_c, as well as M, the total number of sensors (ORNs).

We found that as the odor concentration becomes broader with increasing σ_c, ρ_w increases [Fig. 3(a)]. This is expected as more receptors with different sensitivities are required to sense a broad range of input concentrations. When we increased the odor mixture sparsity n or the total number of possible odors N, the optimal sensitivity matrix sparsity ρ_w decreases [Fig. 3(b)&(c)]. In general, as the mapping from odor space to ORN space becomes more "compressed" with larger values of n and/or N, the optimal strategy is to have each receptor respond to a smaller fraction of odorants to avoid saturation. Finally, we gradually increased the number of receptors M with fixed values of N, n, and σ_c. We found that ρ_w decreases, i.e., the sensitivity matrix becomes more sparse as the number of sensors M increases [Fig. 3(d)]. This somewhat counter-intuitive result can be understood as the system has more sensors to encode signals, each sensor can respond to a smaller number of odors to avoid interference. For all the cases we studied, when the odor concentrations follow a log-normal distribution, then the distribution of the non-zero sensitivities in the optimal sensitivity matrix follows roughly a log-normal distribution with its mean μ_w and standard deviation σ_w depending on the odor statistics (σ_c, n, N) and the number of ORNs M (see Fig. S3 in SM).

To verify whether sparsity is a general (robust) feature in the optimal sensitivity matrix, we studied the cases when the odor concentration follows different distributions, such as a symmetrized power-law distribution, $P_{\rm env}(c) \propto \exp(-\beta|\ln c|)$[see Fig. S4(a) in the SM for the

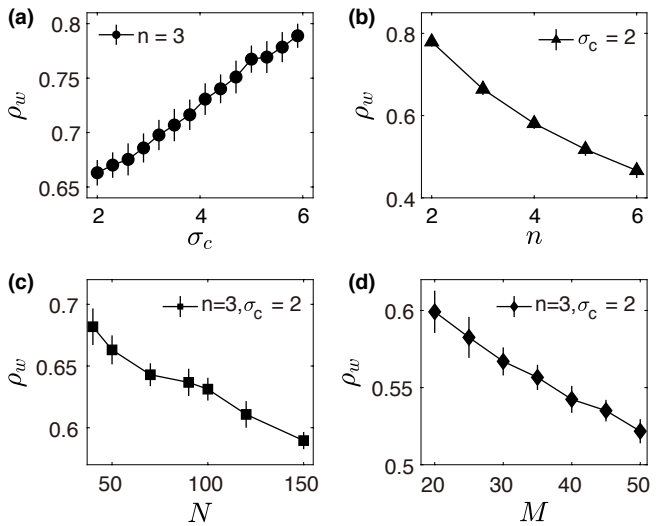

FIG. 3. Dependence of ρ_w on the width of odor log concentration σ_c (a), the input sparsity n (b), the number of total odorants N (d), and the number of receptors M (d). In (a-b), $N = 50, M = 13$, in (c), $M = 13$, and in (d), $N = 50$. Error bars are standard deviation of 40 times simulation.

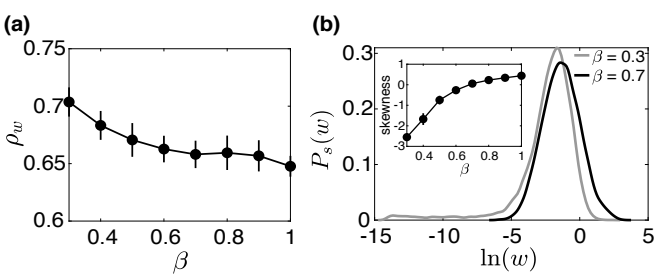

FIG. 4. The optimal sensitivity matrix for the symmetric power-law odor concentration distribution $P_{\rm env}(c) \propto \exp[-\beta|\ln c|]$. (a) The sparsity ρ_w versus the power-law exponent β. (b) The distribution of the non-zero sensitivities $P_s(w)$ for $\beta = 0.3, 0.7$. Inset shows the dependence of the skewness of the distribution on β. Parameters: $N = 50, M = 13, n = 3, \sigma_0 == 10^{-3}$.

comparison with log-normal distribution], with different exponent β. For all values of β studied, there is always a finite sparsity $\rho_w < 1$ in the optimal sensitivity matrix. As shown in Fig. 4(a), ρ_w decreases slightly when β increases and the odor concentration distribution becomes narrower, which is consistent with the previous cases when the odor concentration distribution is log-normal [Fig. 3(a)]. However, as shown in Fig. 4(b), the distribution of the sensitive elements, $P_s(w)$, does not follow exactly a log-normal distribution [see Fig. S4 (b) in SM]. In fact, $P_s(w)$ is asymmetric in the $\ln(w)$ space with a skewness that depends on β as shown in the inset of Fig. 4(b).

Taken together, our results suggest that sparsity in the sensitivity matrix is a robust feature for nonlinear compressed sensing problems. This theoretical finding is supported by and explains existing experiments in olfactory systems [42, 56]. Our study also showed that the non-zero sensitivities follow a broad distribution whose exact shape, mean, and variance depend on odor statistics and total number of ORNs.

C. The origin of sparsity in the optimal sensitivity matrix

Given the constraint that the number of sensors is much smaller than the possible number of odorants, i.e., $M \ll N$, each sensor needs to respond (sense) to multiple types of odorant molecule so that all odorant molecules can be sensed by at least one sensor. However, in an odor mixture with a few types of odorant molecules, two or more odorants in the mixture can bind with the same sensor and interfere with each other, e.g., by saturating the nonlinear sensor. The probability of interference increases with the sparsity of the sensitivity matrix. This tradeoff between sensing multiple odorants and the possible interference determines the sparsity in the optimal sensitivity matrix. We demonstrate this tradeoff and its effect more rigorously by developing a mean-field theory (MFT) as described below.

We begin with the simplest case where many receptors sense only one odorant ($N = 1, M \gg 1$) where obviously there is no interference. As first proposed by Laughlin [22], the optimal coding scheme is for the M receptors to distribute their sensitivities according to the input concentration distribution so that the output distribution is uniform. For the case when the distribution of the odorant concentration is log-normal with a standard deviation σ_c, the optimal sensitivity distribution $P_1(w)$ that maximizes $H(\boldsymbol{r})$ is also approximately a log-normal distribution:

$$P_1(w) \approx \frac{1}{w\sqrt{2\pi\sigma_w^2}} \exp[-\frac{(\ln w - \mu_w)^2}{2\sigma_w^2}], \qquad (4)$$

where the mean $\mu_w = 0$ and the variance σ_w^2 increases with the variance (σ_c^2) of logarithmic concentration distribution. More importantly, we show analytically that in general the coding capacity I_1 increases logarithmically with the number of receptors M when $M \gg 1$ (see SM for details), which is verified by simulation results as shown in Fig. 5(a):

$$I_1(M) = \frac{1}{2}\log(M) + \text{const.} \qquad (5)$$

This means that sparsity $\rho_w = 1$, i.e., all sensitivities should be nonzero because there is no interference when only one type of odorant molecule ($n = 1$) is present in the mixture. However, it is important to note that the maximum mutual information only increases weakly (logarithmically) for large M.

We next consider the case where two odorants are sensed by multiple receptors ($N = 2, M \gg 1$). Let's de-

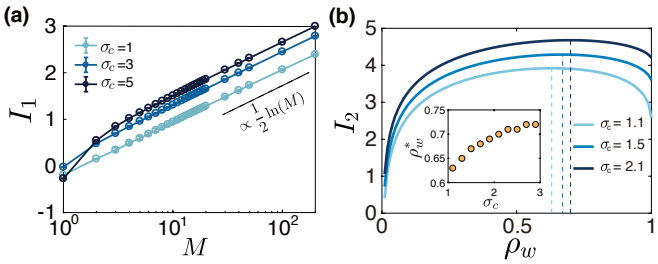

FIG. 5. The tradeoff between increasing single odorant information and interference among multiple odorants. (a) The differential entropy with one odorant, I_1, versus the number of receptors M for different width (σ_c) of odor log-concentration distribution. I_1 increases monotonically with M but it only grows logarithmically with M for large M (dashed line). (b) Differential entropy I_2 for the case with two odorants in the mixture with their concentrations following the same log-normal distribution with width σ_c. I_2 depends non-monotonically on the fraction of sensitive receptors $\rho_w(= m/M)$ with a maximum (marked by the dashed lines) at ρ_w^* that depends on σ_c, which is shown in the inset.

note the number of receptors that respond to each odorant as m ($m \leq M$) and the sparsity $\rho_w = m/M$. If each odorant is sensed by a disjoint set of receptors, the total differential entropy will simply double the amount for a single odorant: $I_2(m) = 2I_1(m)$. However there is a finite probability $p = m/M = \rho_w$, that a given receptor in one set will also respond to the other odorant. Therefore, on average there are $m \times p = m^2/M$ receptors whose output is "corrupted" due to interference between two different odorants in a given mixture. We can write down the differential entropy as

$$I_2(m) = 2I_1(m) - \frac{m^2}{M}\Delta I, \tag{6}$$

where $I_1(m)$ is the maximum differential entropy for one odor [Eq. (5)] and ΔI is the marginal loss of information (entropy loss), which can be approximated by $\Delta I \approx \alpha(I_1(m+1) - I_1(m)) \approx \alpha \partial I_1(m)/\partial m$ where $\alpha \leq 1$ is the average fraction of information loss for a "corrupted" sensor. We can then obtain the optimal value of m by maximizing $I_2(m)$ with respect to m. For $m \ll M$, the interference effect is small, so $I_2(m) \approx 2I_1(m)$, which increases with m logarithmically according to Eq. (5). As m increases, the interference effect given by the second term on the RHS of Eq. (6) increases with m, which is faster than the slow logarithmic growth of $2I_1(m)$. This leads to a peak of $I_2(m)$ at an optimal value of $m = m^* < M$ or a sparsity of the sensitivity matrix $\rho_w = m^*/M < 1$ [Fig. 5(b)].

In the MFT, we can compute the olfactory coding and interference by ignoring the weak rank-order correlation in the optimal sensitivity matrix and assuming the distributions for the optimal sensitivity matrix elements are *i.i.d.*. In particular, we used the following approximation for the distribution of the sensitivity matrix $\boldsymbol{W}$:

$$P(\boldsymbol{W}) = \prod_{j=1}^{N}\prod_{i=1}^{M}[\rho_w \times P_s(W_{ij}) + (1-\rho_w) \times \delta(W_{ij})], \tag{7}$$

where ρ_w is the matrix sparsity, and $P_s(W_{ij})$ is a smooth distribution function, which is approximated here as a log-normal distribution with mean μ_w and standard deviation σ_w as given in Eq. (4). The mean differential entropy of ORN response pattern

$$\langle H \rangle_{\boldsymbol{W}}(\rho_w, \mu_w, \sigma_w) \equiv \int H(\boldsymbol{W})P(\boldsymbol{W})d\boldsymbol{W},$$

which is averaged over the distribution of the sensitivity matrix $\boldsymbol{W}$, can be maximized with respect to the parameters ρ_w, μ_w, and σ_w (see SM for details). The resulting optimal parameters agree with our direct numerical simulations qualitatively with a sparsity $\rho_w < 1$ that increases with the width of the input distribution σ_c (see Fig. S5 in SM).

D. The optimal sparse sensitivity matrix enhances downstream decoding performance

The response patterns of ORNs form the internal representation of external odor stimuli that the higher (downstream) regions of the brain can use to infer the odor information for controlling the organism's behavior. In previous sections, we focused on understanding the statistical properties of the optimal sensitivity matrix W that maximize mutual information between odor input and ORN output. Here in this section, we test whether the optimal sensitivity matrix can enhance the downstream decoding performance by examining two specific learning tasks: classification and reconstruction.

Task I: classification. The goal of the classification task is to infer the category of odor mixture such as the odor valence by training with similar odor stimuli. Classification is believed to be carried out by the *Drosophila* olfactory circuit, which we describe briefly here. After odor signals are sensed by ~ 50 ORNs, they are relayed by the projection neurons (PNs) in antennal lobes to a much larger number of Kenyon cells (KCs) in the mushroom body (MB), as illustrated in Fig. 6(a). Each of the ~ 2000 KCs in MB on average receives input from ~ 7 randomly selected PNs [57]. A single GABAergic neuron (APL) at each side of the brain can be activated by the KCs and inhibits KCs globally [58]. Such random expansion and global inhibition enable sparse and decorrelated representation of odor information in MB [57, 59, 60]. The large number of KCs in MB then converge to a few (only 34) mushroom body output neurons (MBONs) [61], which project to the other brain regions and drives attractive or repulsive behavior [62]. Olfactory learning is mainly mediated by the dopaminergic neurons (DANs) which controls the synaptic weights between KCs and MBONs [63].

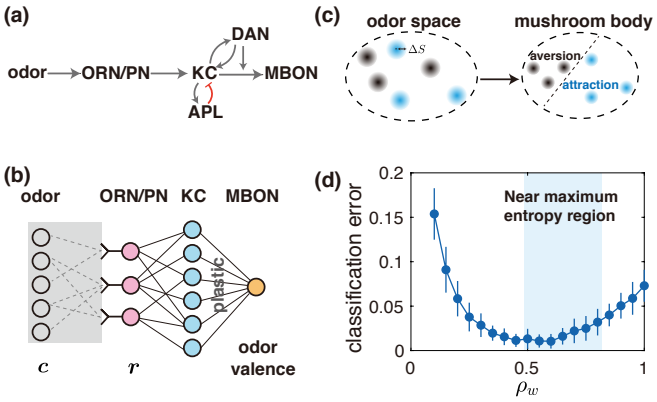

FIG. 6. Maximum entropy coding facilitates olfactory learning and classification. (a) Schematics of the neural circuitry for olfactory learning in fly, see text for detailed description. (b) A simplified model of the fly olfactory system shown in (a) for learning the valence of odor stimuli, where the effect of dopaminergic neurons (DAN) is replaced by simple plastic weights from KC to MBON. (c) Odors are organized as clusters of size ΔS and are randomly assigned with an odor valence. The decoding network receives the response pattern of ORNs and classifies them into the right categories. 100 clusters were drawn and each cluster contains 50 variations, resulting in 5000 odor stimuli among which 80% were used as training data and the rest were used as testing data. (d) Classification performance with respect to the sparsity of sensitivity matrix. Best performance appears at around $\rho_w = 0.6$, within the 95% maximum entropy region. Parameters: $N = 100, M = 10, n = 3, \sigma_c = 2, \sigma_0 = 0.05, \Delta S = 0.1$, 500 KC units and two odor categories. Error bars are standard deviation from 40 simulations.

To mimic the properties of MB, our model "classifier" network, as illustrated in in Fig. 6(b), contains a high dimensional mixed-layer (KCs). For simplicity, we consider a single readout neuron. Each KC unit in the mixed-layer pools the ORNs with a fixed random, sparse matrix. Only the synaptic weights from the KCs to the readout neuron are plastic. To consider the variability of natural odors, we assumed that odor stimuli fall into clusters whose centers represent corresponding typical odor stimuli. Members in a given cluster are variations to the centroid [64]. The radius of a cluster ΔS characterizes the variability of a specific odor mixture. Centroids were drawn from $P_{\text{env}}(\boldsymbol{c})$ with randomly assigned labels (attractive or aversive). Members inside each cluster were generated by adding noise of size ΔS, which results in clouds of points in the odor space [Fig. 6(c)] with each cloud having a randomly assigned label (see SM for details).

The synaptic weights from the KCs to the readout neuron are trained by using a simple linear discriminant analysis (LDA) method although other linear classification algorithms such as support vector machine (SVM) would also work. After training, the performance of the "classifier" is quantified by the accuracy of classification on the testing dataset.

To test effects of different coding schemes on the classification performance, we vary the distribution of the sensitivity matrix elements by changing the sparsity ρ_w without changing the distribution of the non-zero sensitivity matrix elements (e.g., the log-normal distribution with fixed mean and variance). The output of the coding process $\boldsymbol{r}(\boldsymbol{c}, \boldsymbol{W})$ serves as the input for the "classifier" network and the classifier error is computed for different values of ρ_w. As shown in Fig. 6(d), we find that the best performance is achieved near $\rho_w = 0.6$, which belongs to the range of ρ_w with large mutual information between odor input and the ORN/PN response [shaded region in Fig. 6(d)]. Changing parameters such as M, n and number of categories give similar results (see Fig. S6 in SM). In line with recent studies which show that sparse high-dimensional representation facilitates downstream classification [64, 65], our results suggest that maximum entropy coding at the ORNs/PNs level may enhance classification by retaining maximum odor mixture information in a form that can be decoded by the KCs through random expansion.

Task II: reconstruction. This goal of the reconstruction task is to infer (decode) both the composition and the exact concentrations of all odorant components in an odor mixture from the sensor responses. This more stringent task is motivated directly by the original compressed sensing problem in computer science, its relevance to the olfactory systems will be discussed later in the Discussion section.

As illustrated in Fig. 7(a), the output of the coding process $\boldsymbol{r}(\boldsymbol{c}, \boldsymbol{W})$ serves as the input for the downstream reconstruction network. Here, we used a generic feed-forward artificial neural network (ANN) with a few (1-5) hidden layers and a output layer that has the same dimension N as the odor space. We trained the ANN with a training set of sparse odor mixtures drawn from the odor distribution $P_{\text{env}}(\boldsymbol{c})$, and tested its performance by using new odor mixtures randomly drawn from the same odor distribution. Denote the reconstructed odor vector as $\hat{\boldsymbol{c}}$ and a binary vector $\boldsymbol{\xi}$ associated with $\boldsymbol{c}$, i.e., $\xi_i = 1$ if $c_i \neq 0$, otherwise, $\xi_i = 0$. Due to the sparse nature of odor mixture and the wide concentration range, the reconstruction error $\mathcal{L}$ is defined as the sum of "identity error" $\mathcal{L}_1$ and "intensity error" $\mathcal{L}_2$:

$$\mathcal{L} \equiv \mathcal{L}_1 + \mathcal{L}_2 = -\frac{1}{N}\sum_i [\xi_i \log(z_i) + (1-\xi_i)\log(1-z_i)] + \frac{1}{N}\sum_i \xi_i [\ln(\hat{c}_i) - \ln(c_i)]^2, \tag{8}$$

where $z_i = 1/(1+\exp(-W_i^c \hat{c}_i - b_i^c)$ with W_i^c and b_i^c determined from training (supervised learning).

The reconstruction error depends on the coding matrix $\boldsymbol{W}$, in particular its sparsity ρ_w, as shown in Fig. 7(b). Pair-wise comparisons of non-zero concentrations in the original and reconstructed odor mixtures for three different coding regimes are shown in Fig. 7(c)

(see Fig. S7 in SM for a direct comparison of the whole reconstructed and original odor vectors).The best performance is achieved around $\rho_w = 0.6$, within the region where sparse $\boldsymbol{W}$ enable nearly maximum entropy coding (shaded region), this property is insensitive to the number of hidden layers in the reconstruction network (see Fig. S8 in SM). Our results show that the optimal entropy code provides an efficient representation of the high-dimensional sparse data so that the downstream machine learning algorithms can achieve high reconstruction accuracy.

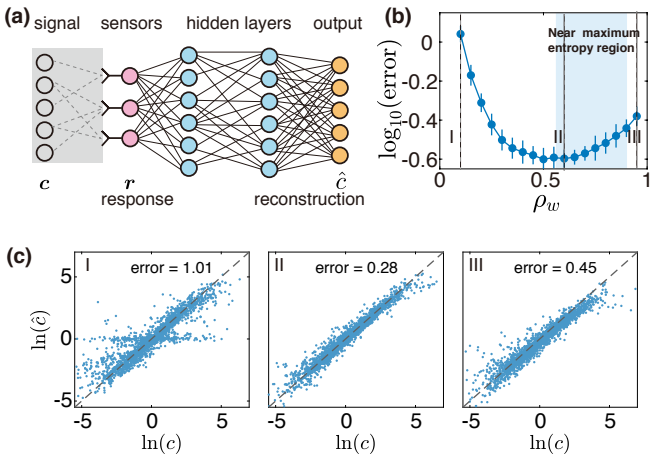

FIG. 7. Maximum entropy coding facilitates signal reconstruction. (a) Schematic of the reconstruction task. The decoding feed-forward network with multiple hidden layers (blue nodes) receives the response $\boldsymbol{r}$ of the peripheral sensors (red nodes) as 'input' and produces the output $\hat{\boldsymbol{c}}$ to reconstruct (infer) the original signal $\boldsymbol{c}$. (b) The minimum reconstruction error on the testing set is achieved around sparsity level $\rho_w = 0.6$ of $\boldsymbol{W}$, within the 95% maximum entropy region (shaded area).(c) Comparison of the original c_i and reconstructed $\hat{c}_i$ for three different sparsity of $\boldsymbol{W}$: I. $\rho_w = 0.1$, II. $\rho_w = 0.6$, III. $\rho_w = 0.95$, which are labeled by the dashed lines in panel (b). Parameters used are: $N = 100$,$M = 20$,$n = 2$,$\sigma_c = 2$, $\sigma_0 = 0.05$, two hidden layers, each with 100 units. Error bars in (b) are standard deviation from 40 times simulation.

E. Optimal coding strategy for ORNs with a finite basal activity

So far, we have only considered the case where the neuron activity is zero in the absence of stimulus and odorants only activate the ORs/ORNs. It has been widely observed that some ORNs show substantial spontaneous activities, and some odorants can act as inhibitors to suppress the activities of neurons they bind to [41, 66, 67], as shown in Fig. 8(a). The presence of an inhibitory odorant can shift a receptor's dose-response curve to an excitatory odorant, thereby diminishing the sensitivity of the receptor to excitatory odorants[68]. It is then natural to ask what is the optimal design of the sensitivity matrix to maximize coding capacity if odorants can be either excitatory or inhibitory. To answer this question, we used a two-state model to characterize both odor-evoked excitation and inhibition [68]. Now, the interaction between the odorant j and ORN i has two possibilities – it can be either excitatory with a sensitivity W^A_{ij} or inhibitory with a sensitivity W^I_{ij}. The normalized response of i-th ORN to odor mixture $\boldsymbol{c}$ is

$$r_i = \left[1 + \gamma \frac{1 + \sum_{q=1}^{n_I} W^I_{iq} c_q}{1 + \sum_{p=1}^{n_A} W^A_{ip} c_p}\right]^{-1} + \eta_i, \qquad (9)$$

where γ determines the basal activity by $r_0 = 1/(1+\gamma)$, n_A and n_I are the number of excitatory and inhibitory odorants to the i-th receptor, and η_i is a small Gaussian white noise.

Our simulations show that with a finite spontaneous activity, the receptor array achieves maximum entropy coding by assigning a certain number of inhibitory interactions in the sensitivity matrix [Fig. 8(b)]. The strength (sensitivity) of both the excitatory and inhibitory elements follow (approximately) log-normal distributions [Fig. 8(c)]. The fraction of inhibitory interaction (ρ_i) in the optimal W is roughly proportional to the spontaneous activity of ORN r_0, with only a slight deviation when $r_0 \to 0$ and $r_0 \to 1$ [Fig. 8(d, upper panel)]. Interestingly, as $r_0 \to 0$, ρ_i approaches a finite value that is related to the fraction of zero sensitivity elements $(1-\rho_w)$ we studied in the previous sections for ORNs without a spontaneous activity. As the basal activity increases, the coding capacity increases rapidly at first and quickly plateaus around $r_0 = 0.3$ [Fig. 8(d, lower panel)]. The increase of coding capacity can be understood intuitively by considering that the effective dynamic range of receptors increases in the presence of inhibition. Odor-evoked inhibition enables receptors to work bi-directionally and avoid saturation when responding to many odorants simultaneously.

To verify our theoretical results, we have analyzed the statistics of the sensitivities for the excitatory and inhibitory interactions obtained from the experimental data in fly by Hallem and Carlson [41] as well as in mosquito by Carey *et al* [69]. As shown in Fig. 8(e) for the fly data, both the excitatory and inhibitory sensitivities follow log-normal distributions, which are consistent with our model results shown in Fig. 8(c). The mosquito data shows very similar results (see Fig. S9 in SM). Our theory also showed that the fraction of inhibitory interaction ρ_i increases with the basal activity r_0 as shown in Fig. 8(d, upper panel). We have tested this theoretical result from the experimental data. As shown in Fig. 8(f), the number of inhibitory odor-ORN interaction for an ORN shows a strong positive correlation with its basal activity for both fly and mosquito, which is in agreement with our theoretical prediction. Finally, we note that the relative basal activity $\langle r_0 \rangle$ from the experimnetal data [41] is smaller than 0.16 (see SM for detailed analysis), where the differential entropy rises sharply with

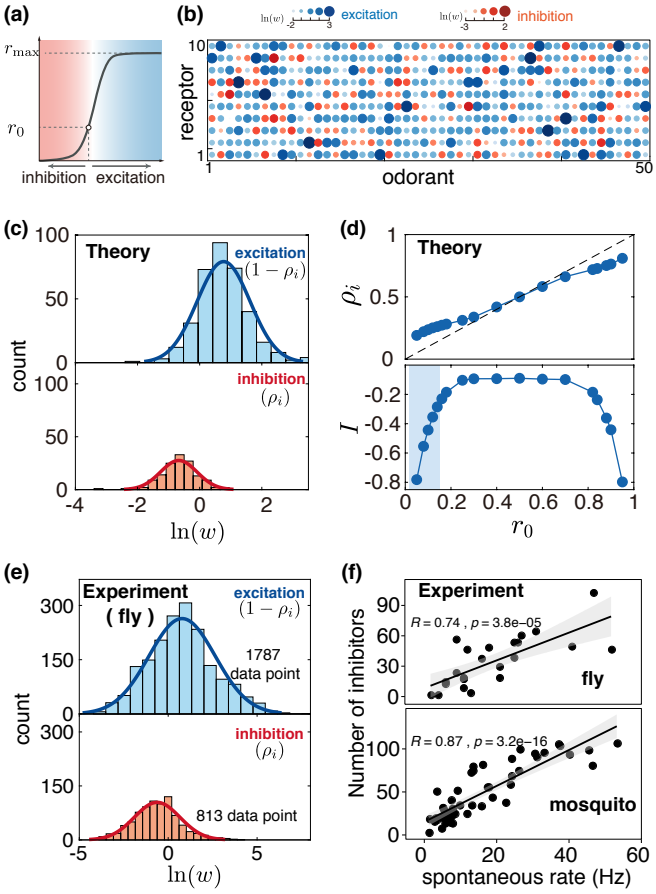

FIG. 8. The optimal sensitivity matrix for ORNs with a finite basal activity and comparison with experiments. (a) Schematic of ORN response to excitatory (blue region) and inhibitory odorants (red region). Note that the neuron has a finite acitivity r_0 in the absence of any stimulus. (b) Heatmap of a typical optimal W from our model, with the size of the elements indicating the strength of excitatory (blue) and inhibitory (red) interactions. (c) Both the excitatory and inhibitory interactions in optimal W can be well approximated by log-normal distributions (solids lines). (d) The fraction of inhibitory interaction ρ_i increases the basal activity r_0 nearly linearly (upper panel). The differential entropy I also increases with r_0 (lower panel). The shaded region shows the range of r_0 corresponding to the fraction of inhibitory interaction estimated from experiments [41], which coincide with the range of r_0 where the differential entropy increases sharply with r_0. (e) The distributions of the estimated relative excitatory and inhibitory receptor-odor sensitivities from experimental data for the fly [41]. Both distributions can be well fitted by log-normal distributions. (f) The correlation between the number of odorants that inhibit an ORN with the ORN's spontaneous activity obtained from experimental data on fly [41] (upper panel) and mosquito [69] (lower panel). Each point corresponds to an ORN, the line is the linear fit and shaded region is the 0.95 confidence interval. Model parameters used are $N = 50, M = 10, n = 2, \sigma_c = 2$, with 40 repeated simulations. In (d), error bars are small, comparable to the size of the symbols. In (b) and (c), $r_0 = 0.18$.

r_0 as highlighted by the shaded region in Fig. 8(d, lower panel). Although an even higher spontaneous activity towards $r_0 = 0.5$ can further increase the coding capacity, the gain is diminishing, while the metabolic cost increases drastically in maintaining the spontaneous activity [70]. Thus, an optimal basal activity would be expected in the shaded region of Fig. 8(d) due to the tradeoff between coding capacity and energy cost.

SUMMARY AND DISCUSSIONS

In this paper, we studied how a relatively small number of nonlinear sensors (ORNs) with a limited dynamic range can optimize the transmission of high dimensional but sparse information in the environment. We found that the optimal sensitivity matrix elements follow a bi-modal distribution. For neurons without a basal activity, the sensitivity matrix is sparse – a neuron only responds to a fraction $\rho_w (< 1)$ of odorants with its sensitivities (to these odorants) following a broad distribution and it is insensitive to the rest $(1 - \rho_w)$ fraction of the odorants. This sparsity in the odor-ORN sensitivity matrix is a direct consequence of the finite dynamic range of the realistic nonlinear ORNs, which are different from the linear sensors in the conventional compressed sensing problem. For neurons with a finite basal activity r_0, the optimal sensitivity distribution is also bi-modal with a fraction ρ_i of the odor-neuron interaction inhibitory and the rest $(1-\rho_i)$ fraction of the odor-neuron interaction excitatory and ρ_i increases with r_0. Details of the odor-receptor sensitivity distribution depend on the odor mixture statistics and the sensor characteristics, but the bi-modal distribution is robust. By investing the effects of different coding schemes on the downstream decoding/learning tasks, we showed that the maximum entropy code (representation) of the external signal enhances the performance of downstream reconstruction and classification tasks.

Connection to experiments and testable predictions. Our primary finding - the sparsity in the odor-receptor sensitivity matrix $\boldsymbol{W}$ - seems to be consistent with existing experimental measurements of receptor-odor sensitivity matrices in different organisms [Fig. 2(c)]. Although the natural odor environment varies for different organisms, it is interesting to see that the broad distribution of the non-zero sensitivity observed in our model is consistent with the sensitivity matrices estimated from experiments in fly larva, mouse, adult fly, and mosquito [Fig. 2(c) and Fig. 8(e)&(f)]. The optimal coding strategy, if exists, would be the result of evolution. Thus, our theory may be tested by comparing olfactory systems in different species. In particular, our theory predicts that the sparsity parameter ρ_w decreases with the number of ORNs M [Fig. 3(d)], which can be tested by measuring the sparsity in the odor receptor sensitivity matrices in different organisms.

The relatively high level of spontaneous activity in ORNs has long been thought to only play a role in the

formation of topographic map during development [71]. A recent study shows that odor-evoked inhibition can code olfactory information that drives the behavior of the fly [68]. Our results provide a quantitative explanation for the advantage of having certain level of spontaneous basal activity and odor-evoked inhibitions in odor coding. For neurons with a finite basal activity, our theory predicts that the fraction of odorants that inhibit the neuron increases with the basal activity of the neuron. The data from adult fly and mosquito are consistent with this prediction [Fig. 8(f)]. However, powerful high throughput techniques such as Calcium imaging, which only indirectly measure the odor-ORN interaction, seem to be incapable of detecting inhibitory interaction [42]. Therefore, more large scale direct measurements using electrophysiological methods such as those done for *Drosophila* [41, 72] and mosquito [69] should be carried out to test our predictions in different organisms.

By considering how the coding capacity of ORNs changes with basal activity [Fig. 8(d)] and the associated extra energy cost [70], one can hypothesize the existence of an "optimal" r_0. Our result suggests that as the number of sensors increases, the benefit of having basal activity diminishes, hence, the "optimal" r_0 should decrease as the number of sensory neurons increases. Indeed, this is consistent with the fact that *E. coli* has 5 chemoreceptors [73] which work bi-directionally with a high basal activity $r_0 \approx 1/3 - 1/2$ [74], and r_0 in mouse is smaller than that in fly [67]. Of course, more experiments across different organisms with different numbers of sensory neurons are needed to test this hypothesis.

Future Directions. In this study, we assumed that odor information is contained in the instantaneous spiking rate of ORNs, and did not consider adaptation dynamics. Although adaptation plays an important role in all sensory systems [75], it happens in a relatively slower time scale than the time required for animals to detect and respond to odor stimuli [76, 77]. In general, sensory adaptation shifts the response function of the sensory neuron according to the background stimulus concentration and it leads to a larger but still finite effective dynamic range without changing the qualitative characteristics of the input-output response curve [75, 78]. Therefore, even though ORN level adaptation can further increase coding capacity at a slightly longer time scale as shown recently by Kadakia and Emonet [79], we do not expect it to qualitatively affect the optimal coding strategy found here. It remains an interesting question to understand how neuronal dynamics such as adaptation can be used for coding time-dependent odor signals.

We have used reconstruction and classification as two learning tasks to demonstrate the advantage of having maximum entropy coding at the ORN level. While the classification task has clear biological relevance, it is unclear to what extent animals need to infer the concentrations of individual odorants in an odor mixture. The perception of odors has been thought as synthesis, i.e., odorant mixture is perceived as a unit odor [16]. Nevertheless, the performance of the reconstruction task indicates that most of the information about the odor mixture including the identities and concentrations of individual odorants in a sparse mixture can potentially be extracted from the activity pattern of ORNs, which is consistent with the experimental finding that mice after training can detect a target odorant in odor mixtures with up to 16 different odorants [80]. In this work, we focused only on the optimal coding strategy for the peripheral ORNs. In the fly olfactory system, odorants that elicit very similar ORN response patterns can be represented by very distinct patterns of KCs [41, 60]. It remains an interesting open question whether and how the architecture of the ORN/PN to KC network optimizes the odor information transmission to enhance precision of downstream learning and decision-making.

In conventional compressed sensing theory with linear sensors, a random measurement matrix enables accurate reconstruction of sparse high dimensional input signals [37, 40]. By using prior information about the input, a better sensory matrix can be designed [81, 82]. In many cases, the optimal matrix maximizes the entropy of compressed representation [83]. Unlike the linear CS problem where the measurement matrix is known and can be used directly for reconstructing the sparse input signal by using the L_1-minimization algorithm, reconstruction in the nonlinear CS problem studied here has to be done by learning without prior knowledge of the sensitivity matrix. Despite this difference, our results suggest that with nonlinear sensors, the sparse optimal sensory matrix that maximizes information transmission enables better learning and more accurate reconstruction. This general observation and the limit of reconstruction in nonlinear CS should be examined with more rigorous analysis and larger numerical simulations.

Finally, in our study, we considered the simplest case where odorants appear independently in odor mixtures. However, even in this simplest case, we have found weak but statistically relevant "orthogonal" structure in the optimal sensitivity matrix [Fig. 2(c) and Fig. S1 in SM]. In naturally occurring odor mixtures, co-occurrence of odorants in different odor sources are common. For example, odorants that are products in the same biochemical reaction pathway, i.e., fermentation, are likely to appear together [2, 84]. Although odorant-evoked ORN response patterns are not simply determined by the molecular structure, some very similar odorants do trigger similar ORN response patterns [41]. On the other hand, ORNs and their responses to different odorants can be correlated due to structural similarities in their receptor proteins. It would be interesting to explore how such correlations among ORNs and odorant molecules as well as co-occurrences among different odorants in odor mixtures can affect the optimal coding strategy at the olfactory periphery in a future study.

ACKNOWLEDGMENTS

We thank Xiaojing Yang, Guangwei Si, Jingxiang Shen, Louis Tao, and Roger Traub for helpful discussions and comments. The work was supported by the Chinese Ministry of Science and Technology (Grant No. 2015CB910300) and the National Natural Science Foundation of China (Grant No. 91430217). The work by YT is partially supported by a NIH grant (R01-GM081747).

[1] Robert A Raguso, "Wake up and smell the roses: the ecology and evolution of floral scent," Annual Review of Ecology, Evolution, and Systematics **39**, 549–569 (2008).

[2] Andreas Dunkel, Martin Steinhaus, Matthias Kotthoff, Bettina Nowak, Dietmar Krautwurst, Peter Schieberle, and Thomas Hofmann, "Nature's chemical signatures in human olfaction: a foodborne perspective for future biotechnology," Angewandte Chemie International Edition **53**, 7124–7143 (2014).

[3] Ivo Beyaert and Monika Hilker, "Plant odour plumes as mediators of plant–insect interactions," Biological Reviews **89**, 68–81 (2014).

[4] Chih-Ying Su, Karen Menuz, and John R Carlson, "Olfactory perception: receptors, cells, and circuits," Cell **139**, 45–59 (2009).

[5] Kazushige Touhara and Leslie B Vosshall, "Sensing odorants and pheromones with chemosensory receptors," Annual review of physiology **71**, 307–332 (2009).

[6] Gustavo Glusman, Itai Yanai, Irit Rubin, and Doron Lancet, "The complete human olfactory subgenome," Genome research **11**, 685–702 (2001).

[7] Christophe Verbeurgt, Françoise Wilkin, Maxime Tarabichi, Françoise Gregoire, Jacques E Dumont, and Pierre Chatelain, "Profiling of olfactory receptor gene expression in whole human olfactory mucosa," PloS one **9**, e96333 (2014).

[8] Cornelia I Bargmann, "Comparative chemosensation from receptors to ecology," Nature **444**, 295 (2006).

[9] John P. McGann, "Poor human olfaction is a 19th-century myth," Science **356**, eaam7263 (2017).

[10] Caroline Bushdid, Marcelo O Magnasco, Leslie B Vosshall, and Andreas Keller, "Humans can discriminate more than 1 trillion olfactory stimuli," Science **343**, 1370–1372 (2014).

[11] Richard C Gerkin and Jason B Castro, "The number of olfactory stimuli that humans can discriminate is still unknown," Elife **4**, e08127 (2015).

[12] Tali Weiss, Kobi Snitz, Adi Yablonka, Rehan M Khan, Danyel Gafsou, Elad Schneidman, and Noam Sobel, "Perceptual convergence of multi-component mixtures in olfaction implies an olfactory white," Proceedings of the National Academy of Sciences **109**, 19959–19964 (2012).

[13] Nicolas Y Masse, Glenn C Turner, and Gregory SXE Jefferis, "Olfactory information processing in drosophila," Current Biology **19**, R700–R713 (2009).

[14] Bettina Malnic, Junzo Hirono, Takaaki Sato, and Linda B Buck, "Combinatorial receptor codes for odors," Cell **96**, 713–723 (1999).

[15] JJ Hopfield, "Odor space and olfactory processing: collective algorithms and neural implementation," Proceedings of the National Academy of Sciences **96**, 12506–12511 (1999).

[16] Gilles Laurent, "A systems perspective on early olfactory coding," Science **286**, 723–728 (1999).

[17] Venkatesh N Murthy, "Olfactory maps in the brain," Annual review of neuroscience **34**, 233–258 (2011).

[18] Veit Grabe, Amelie Baschwitz, Hany KM Dweck, Sofia Lavista-Llanos, Bill S Hansson, and Silke Sachse, "Elucidating the neuronal architecture of olfactory glomeruli in the drosophila antennal lobe," Cell reports **16**, 3401–3413 (2016).

[19] Olivier Gschwend, Nixon M Abraham, Samuel Lagier, Frédéric Begnaud, Ivan Rodriguez, and Alan Carleton, "Neuronal pattern separation in the olfactory bulb improves odor discrimination learning," Nature neuroscience **18**, 1474 (2015).

[20] Ronald L Davis, "Olfactory memory formation in drosophila: from molecular to systems neuroscience," Annu. Rev. Neurosci. **28**, 275–302 (2005).

[21] Horace B Barlow, "Possible principles underlying the transformations of sensory messages," (1961).

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

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

[24] Joseph J Atick, "Could information theory provide an ecological theory of sensory processing?" Network: Computation in neural systems **3**, 213–251 (1992).

[25] David J Field, "Relations between the statistics of natural images and the response properties of cortical cells," Journal of the Optical Society of America A **4**, 2379–2394 (1987).

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

[27] Yang Dan, Joseph J Atick, and R Clay Reid, "Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory," Journal of Neuroscience **16**, 3351–3362 (1996).

[28] Anthony J Bell and Terrence J Sejnowski, "The "independent components" of natural scenes are edge filters," Vision research **37**, 3327–3338 (1997).

[29] Daniel L Ruderman and William Bialek, "Statistics of natural images: Scaling in the woods," in *Advances in neural information processing systems* (1994) pp. 551–558.

[30] Michael S Lewicki, "Efficient coding of natural sounds," Nature neuroscience **5**, 356 (2002).

[31] Naama Brenner, William Bialek, and Rob de Ruyter Van Steveninck, "Adaptive rescaling maximizes information transmission," Neuron **26**, 695–702 (2000).

[32] Odelia Schwartz and Eero P Simoncelli, "Natural signal statistics and sensory gain control," Nature neuroscience **4**, 819 (2001).

[33] Xaq Pitkow and Markus Meister, "Decorrelation and efficient coding by retinal ganglion cells," Nature neuroscience **15**, 628 (2012).

[34] Rafi Haddad, Rehan Khan, Yuji K Takahashi, Kensaku Mori, David Harel, and Noam Sobel, "A metric for odorant comparison," Nature methods **5**, 425 (2008).

[35] Andreas Keller, Richard C Gerkin, Yuanfang Guan, Amit Dhurandhar, Gabor Turu, Bence Szalai, Joel D Mainland, Yusuke Ihara, Chung Wen Yu, Russ Wolfinger, *et al.*, "Predicting human olfactory perception from chemical features of odor molecules," Science **355**, 820–826 (2017).

[36] Hany KM Dweck, Shimaa AM Ebrahim, Tom Retzke, Veit Grabe, Jerrit Weißflog, Ales Svatoš, Bill S Hansson, and Markus Knaden, "The olfactory logic behind fruit odor preferences in larval and adult drosophila," Cell reports **23**, 2524–2531 (2018).

[37] Emmanuel J Candes and Terence Tao, "Decoding by linear programming," IEEE transactions on information theory **51**, 4203–4215 (2005).

[38] Emmanuel J Candes, Justin K Romberg, and Terence Tao, "Stable signal recovery from incomplete and inaccurate measurements," Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences **59**, 1207–1223 (2006).

[39] 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).

[40] Emmanuel J Candès and Michael B Wakin, "An introduction to compressive sampling," IEEE signal processing magazine **25**, 21–30 (2008).

[41] Elissa a. Hallem and John R. Carlson, "Coding of Odors by a Receptor Repertoire," Cell **125**, 143–160 (2006).

[42] Guangwei Si, Jessleen K Kanwal, Yu Hu, Christopher J Tabone, Jacob Baron, Matthew Berck, Gaetan Vignoud, and Aravinthan DT Samuel, "Structured odorant response patterns across a complete olfactory receptor neuron population," Neuron (2019).

[43] Yilun Zhang and Tatyana O Sharpee, "A robust feedforward model of the olfactory system," PLoS computational biology **12**, e1004850 (2016).

[44] Kamesh Krishnamurthy, Ann M Hermundstad, Thierry Mora, Aleksandra M Walczak, and Vijay Balasubramanian, "Disorder and the neural representation of complex odors: smelling in the real world," arXiv preprint arXiv:1707.01962 (2017).

[45] Vijay Singh, Martin Tchernookov, and Vijay Balasubramanian, "What the odor is not: Estimation by elimination," Bulletin of the American Physical Society (2019).

[46] David Zwicker, Arvind Murugan, and Michael P Brenner, "Receptor arrays optimized for natural odor statistics," Proceedings of the National Academy of Sciences **113**, 5570–5575 (2016).

[47] Vijay Singh, Nicolle R Murphy, Vijay Balasubramanian, and Joel D Mainland, "A competitive binding model predicts nonlinear responses of olfactory receptors to complex mixtures," arXiv preprint arXiv:1805.00563 (2018).

[48] Gautam Reddy, Joseph D Zak, Massimo Vergassola, and Venkatesh N Murthy, "Antagonism in olfactory receptor neurons and its implications for the perception of odor mixtures," eLife **7**, e34958 (2018).

[49] Thomas M Cover and Joy A Thomas, *Elements of information theory* (John Wiley & Sons, 2012).

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

[51] Chris R Sims, "Efficient coding explains the universal law of generalization in human perception," Science **360**, 652–656 (2018).

[52] Gašper Tkačik and William Bialek, "Information processing in living systems," Annual Review of Condensed Matter Physics **7**, 89–117 (2016).

[53] Nikolaus Hansen and Andreas Ostermeier, "Completely derandomized self-adaptation in evolution strategies," Evolutionary computation **9**, 159–195 (2001).

[54] Nikolaus Hansen, "The cma evolution strategy: a comparing review," in *Towards a new evolutionary computation* (Springer, 2006) pp. 75–102.

[55] Antonio Celani, Emmanuel Villermaux, and Massimo Vergassola, "Odor landscapes in turbulent environments," Physical Review X **4**, 041015 (2014).

[56] Harumi Saito, Qiuyi Chi, Hanyi Zhuang, Hiroaki Matsunami, and Joel D Mainland, "Odor coding by a mammalian receptor repertoire," Sci. Signal. **2**, ra9–ra9 (2009).

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

[58] Andrew C Lin, Alexei M Bygrave, Alix De Calignon, Tzumin Lee, and Gero Miesenböck, "Sparse, decorrelated odor coding in the mushroom body enhances learned odor discrimination," Nature neuroscience **17**, 559–568 (2014).

[59] Glenn C Turner, Maxim Bazhenov, and Gilles Laurent, "Olfactory representations by drosophila mushroom body neurons," Journal of neurophysiology **99**, 734–746 (2008).

[60] Robert AA Campbell, Kyle S Honegger, Hongtao Qin, Wanhe Li, Ebru Demir, and Glenn C Turner, "Imaging a population code for odor identity in the drosophila mushroom body," Journal of Neuroscience **33**, 10568–10581 (2013).

[61] Yoshinori Aso, Daisuke Hattori, Yang Yu, Rebecca M Johnston, Nirmala A Iyer, Teri-TB Ngo, Heather Dionne, LF Abbott, Richard Axel, Hiromu Tanimoto, *et al.*, "The neuronal architecture of the mushroom body provides a logic for associative learning," Elife **3**, e04577 (2014).

[62] David Owald, Johannes Felsenberg, Clifford B Talbot, Gaurav Das, Emmanuel Perisse, Wolf Huetteroth, and Scott Waddell, "Activity of defined mushroom body output neurons underlies learned olfactory behavior in drosophila," Neuron **86**, 417–427 (2015).

[63] Paola Cognigni, Johannes Felsenberg, and Scott Waddell, "Do the right thing: neural network mechanisms of memory formation, expression and update in drosophila," Current opinion in neurobiology **49**, 51–58 (2018).

[64] Baktash Babadi and Haim Sompolinsky, "Sparseness and expansion in sensory representations," Neuron **83**, 1213–1226 (2014).

[65] Ashok Litwin-Kumar, Kameron Decker Harris, Richard Axel, Haim Sompolinsky, and LF Abbott, "Optimal degrees of synaptic connectivity," Neuron **93**, 1153–1164 (2017).

[66] Joby Joseph, Felice A Dunn, and Mark Stopfer, "Spontaneous olfactory receptor neuron activity determines follower cell response properties," Journal of Neuroscience **32**, 2900–2910 (2012).

[67] Timothy Connelly, Agnes Savigner, and Minghong Ma,

"Spontaneous and sensory-evoked activity in mouse olfactory sensory neurons with defined odorant receptors," Journal of neurophysiology **110**, 55–62 (2013).

[68] Li-Hui Cao, Dong Yang, Wei Wu, Xiankun Zeng, Bi-Yang Jing, Meng-Tong Li, Shanshan Qin, Chao Tang, Yuhai Tu, and Dong-Gen Luo, "Odor-evoked inhibition of olfactory sensory neurons drives olfactory perception in drosophila," Nature communications **8**, 1357 (2017).

[69] Allison F Carey, Guirong Wang, Chih-Ying Su, Laurence J Zwiebel, and John R Carlson, "Odorant reception in the malaria mosquito anopheles gambiae," Nature **464**, 66 (2010).

[70] Jeremy E Niven and Simon B Laughlin, "Energy limitation as a selective pressure on the evolution of sensory systems," Journal of Experimental Biology **211**, 1792–1804 (2008).

[71] Ai Nakashima, Haruki Takeuchi, Takeshi Imai, Harumi Saito, Hiroshi Kiyonari, Takaya Abe, Min Chen, Lee S Weinstein, C Ron Yu, Daniel R Storm, *et al.*, "Agonist-independent gpcr activity regulates anterior-posterior targeting of olfactory sensory neurons," Cell **154**, 1314–1325 (2013).

[72] Scott A Kreher, Dennis Mathew, Junhyong Kim, and John R Carlson, "Translation of sensory input into behavioral output via an olfactory system," Neuron **59**, 110–124 (2008).

[73] T.W. Grebe and J. Stock, "Bacterial chemotaxis: The five sensors of a bacterium," Current Biology **8**, R154–R157 (1998).

[74] T.S. Shimizu, Y. Tu, and H.C. Berg, "A modular gradient-sensing network for chemotaxis in escherichia coli revealed by responses to time-varying stimuli," Molecular systems biology **6** (2010).

[75] Barry Wark, Brian Nils Lundstrom, and Adrienne Fairhall, "Sensory adaptation," Current opinion in neurobiology **17**, 423–429 (2007).

[76] Alexander Egea-Weiss, Christoph J Kleineidam, Paul Szyszka, *et al.*, "High precision of spike timing across olfactory receptor neurons allows rapid odor coding in drosophila," iScience **4**, 76–83 (2018).

[77] Quentin Gaudry, Elizabeth J Hong, Jamey Kain, Benjamin L de Bivort, and Rachel I Wilson, "Asymmetric neurotransmitter release enables rapid odour lateralization in drosophila," Nature **493**, 424–428 (2013).

[78] Carlotta Martelli, John R Carlson, and Thierry Emonet, "Intensity invariant dynamics and odor-specific latencies in olfactory receptor neuron response," Journal of Neuroscience **33**, 6285–6297 (2013).

[79] Nirag Kadakia and Thierry Emonet, "Front-end weber-fechner gain control enhances the fidelity of combinatorial odor coding," bioRxiv , 475103 (2018).

[80] Dan Rokni, Vivian Hemmelder, Vikrant Kapoor, and Venkatesh N Murthy, "An olfactory cocktail party: figure-ground segregation of odorants in rodents," Nature neuroscience **17**, 1225 (2014).

[81] Yair Weiss, Hyun Sung Chang, and William T Freeman, "Learning compressed sensing," in *Snowbird Learning Workshop, Allerton, CA* (Citeseer, 2007).

[82] David L Donoho, Adel Javanmard, and Andrea Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," IEEE transactions on information theory **59**, 7434–7464 (2013).

[83] Amit Ashok, Liang-Chih Huang, and Mark A Neifeld, "Information optimal compressive sensing: static measurement design," JOSA A **30**, 831–853 (2013).

[84] Suzan Mansourian and Marcus C Stensmyr, "The chemical ecology of the fly," Current opinion in neurobiology **34**, 95–102 (2015).

Supplemental Material: The optimal odor-receptor sensitivity matrix is sparse in olfactory systems: Compressed sensing by nonlinear neurons with a finite dynamic range

Shanshan Qin, Qianyi Li, Chao Tang and Yuhai Tu

S1. ORN RESPONSE

A. Odor-receptor interaction and OR sensitivities

We model the odor-evoked ORN response using an equilibrium binding model, where odor receptors (ORs) have two states, ON and OFF. The activity of ORN is then determined by the fraction of ORs that are in the ON state. First, consider the case that the i-th receptor R_i is activated by the binding of an j-th odorant with concentration c_j, which can be described as a chemical reaction $R_i + c_j \underset{k_-}{\overset{k_+}{\rightleftharpoons}} R_i c_j$, where $R_i c_j$ is the receptor-odorant complex. At equilibrium, the activity of ORN is proportional to the fraction of odorant-bound receptor, which reads

$$r_i = r_{\max} \frac{c_j/K_m^{ij}}{1 + c_j/K_m^{ij}} = r_{\max} \frac{c_j W_{ij}}{1 + c_j W_{ij}}, \tag{S1}$$

where $r_{\max}$ is the maximum neuronal activity, i.e., the maximum spiking rate, and we have normalized it to be 1 for simplicity. $K_m^{ij} \equiv k_-/k_+$ is the dissociation constant between the i-th receptor and the j-th odorant, $W_{ij} \equiv 1/K_m^{ij}$ is the sensitivity constant.

Next, consider the multiple odorant molecules scenario. The response of an ORN to them can be characterized by a simple competitive binding model[1, 2], thus odorants compete for the only binding site of the receptor. The response of R_i to the possible N odorants can be described by the following reactions:

$$\begin{aligned} c_1 + R_i &\underset{k_{1-}}{\overset{k_{1+}}{\rightleftharpoons}} R_i c_1, \\ &\dots \\ c_j + R_i &\underset{k_{j-}}{\overset{k_{j+}}{\rightleftharpoons}} R_i c_j, \\ &\dots \\ c_N + R_i &\underset{k_{N-}}{\overset{k_{N+}}{\rightleftharpoons}} R_i c_N. \end{aligned}$$

At equilibrium, the activity of i-th ORN is determined by the fraction of R_i that are in the bound state

$$\theta \equiv \frac{\sum_{j=1}^N R_i c_j}{R_{tot}} = \frac{\sum_{j=1}^N c_j/K_m^{ij}}{1 + \sum_{j=1}^N c_j/K_m^{ij}}, \tag{S2}$$

where R_{tot} is the total number of receptors, $K_m^{ij} \equiv k_{j-}/k_{j+}$ is the dissociation constant between i-th OR and j-th odorant. Hence, the response of i-th ORN becomes

$$r_i = r_{\max}\theta = r_{\max} \frac{\sum_{j=1}^N c_j W_{ij}}{1 + \sum_{j=1}^N c_j W_{ij}}, \tag{S3}$$

where $W_{ij} \equiv 1/K_m^{ij}$.

B. Inhibitory response

Model for both odor-evoked excitation and inhibition have been explained in detail in [3]. Briefly, each OR has two states, ON and OFF, each of which can be further divided into two subtypes: free and odorant-bound, resulting in four different microstates. The free energy difference $\Delta\mathcal{F}$ between the ON and OFF state of an OR is modulated by the odorants binding to the receptor

$$\begin{aligned} \Delta\mathcal{F} &= \mathcal{F}_{i,I} - \mathcal{F}_{i,A} \\ &= E_{0,i} + \ln(1 + \sum_j c_j W_{ij}^A) - \ln(1 + \sum_j c_j W_{ij}^I), \end{aligned} \tag{S4}$$

where W_{ij}^A and W_{ij}^I are the sensitivity of the ON and OFF state of the i-th OR to the j-th odorant. For simplicity, we have assumed that an excitatory odorant only binds to the ON state ($W_{ij}^I = 0$), while an inhibitory odorant only binds to the OFF state ($W_{ij}^A = 0$). The activity of the i-th ORN is given by

$$r_i = \frac{r_{\max}}{1 + e^{-\Delta\mathcal{F}}} = r_{\max}[1 + \gamma_i \frac{1 + \sum_{j=1}^N c_j W_{ij}^I}{1 + \sum_{j=1}^N c_j W_{ij}^A}]^{-1}, \tag{S5}$$

where γ_i determines the basal activity r_0^i of the i-th OR/ORN through $\gamma_i = r_{\max}/r_0^i - 1$, and we have omitted the Boltzmann factor $k_B T$ in the above equation for simplicity. In our model, basal activity of each receptor has been set to be the same, i.e., $r_0^i = r_0, \gamma_i = \gamma (i = 1, \cdots, M)$.

C. Measured sensitivities of olfactory receptors

Responses of ORs to panels of odorants have been measured for fly larva [4], adult fly [5], mouse [6] and mosquito[7]. For fly larva, cognate odorants have been identified for each of the 18 ORs (cognate odorants could not be found for the other 3 ORs). Then a dose response curve was measured for each odorant-receptor pair. And the sensitivity of an OR to the odorant was estimated by fitting the dose response curve with a Hill function, the resulting sensitivity matrix was reported in [4]. For the mouse receptors, odor mixtures drawn from a panel of 93 odorants with various concentrations were tested against 219 receptors and 169 of them showed some level of response. A further dose-response test resulted in a response matrix for 52 ORs and 63 odorants [6]. The frac-

tion of sensitive OR-odorant interactions were directly calculated from these sensitivity matrices.

For adult fly, the responses of 24 ORs to a panel of 110 odorants have been measured, together with the spontaneous firing rate of each ORN in the absent of stimulus [5]. Since the measurement was carried out using the same concentration of odorants, as a rough estimation, we use Eq. (S5) to fit the response and then estimate the relative sensitivity for both excitatory and inhibitory interactions. The resulting distributions of W^A and W^I are shown in figure 8(e), and the fraction of inhibitory interactions is about 30%. If we take a more stringent criterion and only consider the strong interactions (odorants that elicit > 50Hz response are considered excitatory odorants and odorants that suppress more than 50% of spontaneous activity are inhibitory), then the fraction of inhibition is about 11%. Different ORNs have different spontaneous firing rate, ranging from 1 spike to 52 spikes per second. The maximum spiking rate is about 294 spikes per second. Thus, the basal activity r_0 estimated is in the range of $0.1 \sim 0.18$, as shown in the shaded region in Fig. 8(d).

For the mosquito, the responses of 50 ORs to a panel of 110 odorants have been measured using similar method as in adult fly [7]. Since all the odorants were presented as the same concentration, we can use Eq. (S5) to fit the response and then estimate the relative sensitivities.

S2. MEAN FIELD THEORY

In this section, we describe our mean filed theory that helps understanding the origin of the sparsity of optimal odorant-receptor sensitivity matrix in detail. We first simplify the expression of mutual information in the limit of small Gaussian noise and derive an effective target function ΔH. We then consider the case with only one odorant which has a broad concentration distribution and many receptors ($N = n = 1, M \gg 1$), and next the two odorants scenario ($N = n = 2, M \gg 1$), where we demonstrate the existence of "optimal" sparsity (ρ_w) of sensitivity matrix that enables optimal coding. Finally, we extend the result to more general scenarios, i.e., $N, M \gg 1, n = 2$, where the upper bound of ρ_w can be approximated with the result from the case with $N = n = 2, M \gg 1$.

As shown in Fig. 1(a) in the main text, the "coding" problem is basically how the odor vector $\boldsymbol{c}$ is best mapped to the ORN activity vector $\boldsymbol{r}$ by

$$r_i = F_i(\boldsymbol{c}, \boldsymbol{W}) + \eta_i, \tag{S6}$$

where $F_i(\boldsymbol{c}, \boldsymbol{W})$ is the nonlinear response function for the i-th neuron ($i = 1, 2, \ldots, M$) as introduced in previous section, and η_i is the noise, which may assume certain specific form given its biophysical/biochemical origin. For convenience, here we consider it to be a small Gaussian noise with zero mean and a standard deviation σ_0, and noise among ORNs are independent.

For a given odor environment, characterized by the mixture distribution $P_{\text{env}}(\boldsymbol{c})$, we can derive the distribution of ORNs response

$$P_r(\boldsymbol{r}) = \int (2\pi\sigma_0^2)^{-\frac{M}{2}} \exp[-\sum_{i=1}^{M} \frac{(r_i - F_i(\boldsymbol{c}, \boldsymbol{W}))^2}{2\sigma_0^2}] P_{\text{env}}(\boldsymbol{c}) d\boldsymbol{c}. \tag{S7}$$

The optimal entropy coding strategy is to determine the response functions, which are characterized by the sensitivity matrix $\boldsymbol{W}$, to maximize the mutual information $I(\boldsymbol{c}, \boldsymbol{r})$ [Eq. (2) in main text] between odor environment $P_{\text{env}}(\boldsymbol{c})$ and ORN response pattern $P_r(\boldsymbol{r})$. In the limit of small noise, the conditional entropy $H(\boldsymbol{r}|\boldsymbol{c}) \to 0$, hence $I(\boldsymbol{c}, \boldsymbol{r}) \approx H(\boldsymbol{r})$, the differential entropy in the response space is

$$H(\boldsymbol{r}) = -\int P_r(\boldsymbol{r}) \ln[P_r(\boldsymbol{r})] d\boldsymbol{r}. \tag{S8}$$

We first consider the log-normal distribution of odorant concentration, then $H(\boldsymbol{r})$ can be simplified as

$$\begin{aligned} H(\boldsymbol{r}) &\approx -\int P_{\text{env}}(\boldsymbol{c}) \ln\{\int (2\pi\sigma_0^2)^{-M/2} \exp[-\sum_{i=1}^{M} (F_i(\boldsymbol{c}, \boldsymbol{W}) - F_i(\boldsymbol{c}', \boldsymbol{W}))^2/(2\sigma_0^2)] P_{\text{env}}(\boldsymbol{c}') d\boldsymbol{c}'\} d\boldsymbol{c} \\ &\approx -\int P_{\text{env}}(\boldsymbol{c}) \ln P_{\text{env}}(\boldsymbol{c}) d\boldsymbol{c} + (M - N) \ln(\sqrt{2\pi}\sigma_0) + \frac{1}{2} \int P_{\text{env}}(\boldsymbol{c}) \ln[\det(\chi)] \text{dc} \end{aligned} \tag{S9}$$

where χ is a $N \times N$ Jacobian matrix with its matrix elements given by

$$\chi_{kl} = \sum_{i=1}^{M} \frac{\partial r_i}{\partial c_k} \times \frac{\partial r_i}{\partial c_l}. \tag{S10}$$

Eq.(S9) holds when $N \leq M$. The first two terms in the expression for $H(\boldsymbol{r})$ are independent of $\boldsymbol{W}$ with clear physical meanings. The first term is just the entropy of the odor distribution, and the second term is due to dimension reduction. The third term is the most interesting term. It depends on the sensitivity matrix and it is this term that we need to optimize given the odor distribution $P_{\text{env}}(\boldsymbol{c})$ and the form of the response function

F. We denote the third term as

$$\Delta H \equiv \frac{1}{2}\int P_{\rm env}(\boldsymbol{c})\ln[\det(\chi)]{\rm dc}, \tag{S11}$$

which plays the central rule in the following analysis. Here and after, we will call ΔH as our simplified target function which is to be maximized.

A. One odorant and multiple receptors

Since we are considering the log-normally distributed odorant concentration, it is convenient to work in log scale. We now define $K_{ij} \equiv \ln(K_m^{ij})$ and $\bar{c} = \ln(c)$. For the case $N = n = 1$ and $M \gg 1$, χ becomes a scale, which is

$$\chi = \sum_{i=1}^{M}\Big(\frac{\partial r_i}{\partial \bar{c}}\Big)^2 = \sum_{i=1}^{M} G(\bar{c}-K_i), \tag{S12}$$

where we have defined $G(x) = e^{2x}/(1+e^x)^4$, the square of the derivative of the response function. It has a peak at $x = 0$ with a width ~ 1, i.e., it can be well approximated as the product of the standard Gaussian function and a constant: $G(x) \propto \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$. Eq. (S11) then becomes

$$\Delta H = \frac{1}{2}\int P_{\rm env}(\bar{c})\ln[\sum_{i=1}^{M} G(\bar{c}-K_i)]d\bar{c}, \tag{S13}$$

In the limit of $M \to \infty$, the problem can be solved by assuming a distribution of K, $P_k(K)$, which leads to:

$$\Delta H = \int P_{\rm env}(\bar{c})\ln[M\int G(\bar{c}-K)P_k(K)dK]d\bar{c}. \tag{S14}$$

Now the problem becomes optimizing ΔH with respect to the distribution $P_k(K)$ under the constraint $\int P_k(K)dK = 1$. Using variational method, the problem becomes an optimization of a functional $\mathcal{L}[P_k]$ with Lagrangian multiplier

$$\begin{aligned}\mathcal{L}[P_k] = &\int P_{\rm env}(\bar{c})\ln[M\int G(\bar{c}-K)P_k(K)dK]d\bar{c}\\ &-\lambda\Big(\int P_k(K)dK - 1\Big).\end{aligned} \tag{S15}$$

The necessary condition for the maximum value of $\mathcal{L}$ with respect to $P_k(K)$ is

$$\delta\mathcal{L} = \int\Big(\int \frac{P_{\rm env}(\bar{c})G(\bar{c}-K)}{\int G(\bar{c}-K')P_k(K')dK'}d\bar{c} - \lambda\Big)\delta P_k dK = 0, \tag{S16}$$

which holds for arbitrary variation δP_k. So, we have

$$\begin{aligned}\int \frac{P_{\rm env}(\bar{c})G(\bar{c}-K)}{\int G(\bar{c}-K')P_k(K')dK'}d\bar{c} &= \int f(\bar{c})G(\bar{c}-K)d\bar{c}\\ &= \int f(\bar{c})G(K-\bar{c})d\bar{c} = \lambda,\end{aligned} \tag{S17}$$

where we have defined

$$f(\bar{c}) \equiv \frac{P_{\rm env}(\bar{c})}{\int G(\bar{c}-K')P_k(K')dK'} \tag{S18}$$

and used the fact that $G(x)$ is an even function. To find the condition of $P_k(K)$ that satisfies Eq.S17, take the Fourier transform of Eq.S17, we have

$$\tilde{f}(w)\tilde{G}(w) = \lambda\delta(w). \tag{S19}$$

From the reverse Fourier transformation, we have

$$f(\bar{c}) = \lambda\int\frac{\delta(w)}{\tilde{G}(w)}e^{iw\bar{c}}dw = \frac{\lambda}{\tilde{G}(0)} = \text{const.} \tag{S20}$$

From the definition of $f(\bar{c})$ (Eq. S18), we derive the necessary condition to achieve maximum $\mathcal{L}$

$$\int G(\bar{c}-K)P_k(K)dK = \lambda' P_{\rm env}(\bar{c}), \tag{S21}$$

where $\lambda' = \tilde{G}(0)/\lambda$ is a constant, which can be determined by integrating both sides of the above equation. Substitute $\bar{c}-K$ with x and integrate over x, we have

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(x)P_k(K)dKdx = \lambda'\int_{-\infty}^{\infty} P_{\rm env}(x+K)dx, \tag{S22}$$

the LHS of the above equation is $\int_{-\infty}^{\infty} G(x)dx = 1/6$, while the integration in the RHS is independent of K (recall that $P_{\rm env}(\bar{c})$ is a Gaussian function with respect to $\bar{c}$) and is just 1, thus $\lambda' = 1/6$.

Eq. (S21) can be solved by the deconvolution method. We first approximate $G(x) = \frac{e^{2x}}{(1+e^x)^4}$ by the Gaussian function, i.e., $G(x) \propto \frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})$(they have nearly identical shape). From the Fourier transformation of Eq. (S21), we have

$$\mathcal{F}\{P_k\} \propto \frac{\mathcal{F}\{P_{\rm env}\}}{\mathcal{F}\{G\}}, \tag{S23}$$

where

$$\mathcal{F}\{P_{\rm env}\} = \frac{1}{2\pi}\exp[-\frac{1}{2}w^2\sigma_c^2], \tag{S24}$$

and

$$\mathcal{F}\{G\} \propto \frac{1}{2\pi}\exp[-\frac{1}{2}w^2]. \tag{S25}$$

Thus

$$\mathcal{F}\{P_k\} \propto \exp[-\frac{1}{2}w^2(\sigma_c^2-1)]. \tag{S26}$$

By the inverse Fourier transformation of the above equation, we have

$$P_k(K) \propto \exp[-\frac{K^2}{2(\sigma_c^2-1)}], \tag{S27}$$

which is a Gaussian distribution with a standard deviation $\sigma' = \sqrt{\sigma_c^2 - 1}$. By definition, $W_{ij} = \exp(-K_{ij})$, thus W_{ij} follows a log-normal distribution, which is Eq. (3) in the main text.

As $M \to \infty$, plug Eq. (S27) in Eq. (S14), we derive Eq. (4) in the main text, thus

$$I_1(M) = \Delta H + H(\boldsymbol{c}) = \frac{1}{2}\log(M) + \frac{1}{4}\log(2\pi e \sigma_c^2 \lambda'^2), \tag{S28}$$

where σ_c is the standard deviation of the input and the only term in $I_1(M)$ that depends on M is $\frac{1}{2}\log M$. Thus, when M is sufficiently large, $I_1(M)$ increases logarithmically with M.

B. Two odorants and multiple receptors

In this scenario, we use both explicit computation and phenomenological model to approximate Eq. (S14). In order to compute Eq. (S14) explicitly, we first write down the matrix elements of χ

$$\begin{aligned}
\chi_{11} &= \sum_{i=1}^{M} \Delta_1^{-4} \exp[2(\bar{c}_1 - K_{i1})], \\
\chi_{22} &= \sum_{i=1}^{M} \Delta_2^{-4} \exp[2(\bar{c}_2 - K_{i2})], \\
\chi_{12} = \chi_{21} &= \sum_{i=1}^{M} (\Delta_1\Delta_2)^{-2} \exp(\bar{c}_1 + \bar{c}_2 - K_{i1} - K_{i2}),
\end{aligned} \tag{S29}$$

where $\Delta_i \equiv [1+\exp(\bar{c}_1 - K_{i1}) + \exp(\bar{c}_2 - K_{i2})]$ for $i = 1, 2$. From the χ matrix, we can calculate its determinant $\det(\chi) = \chi_{11}\chi_{22} - \chi_{12}^2$ and thus obtain the expression for ΔH: where $\vec{K}_j = (K_{1j}, K_{2j}, \ldots, K_{Mj})$ is a dimension-M sensitivity vector for j-th odorant in the receptor space. The optimal design can be obtained by requiring $\partial \Delta H / \partial K_{ij} = 0$ for all $j = 1, 2$ and $i = 1, 2, \ldots, M$.

In the limit of $M \to \infty$, the problem can be tackled using a distribution function $P_k(K_1, K_2)$. The matrix elements of χ now depend on this distribution. Using variational method similar to the case for $N = 1$, the final equation for $P_k(K_1, K_2)$ is

$$\int \frac{P_{\text{env}}(\bar{\boldsymbol{c}}) \int G(\bar{\boldsymbol{c}}, \vec{K}, \vec{K}') P_k(\vec{K}') d\vec{K}'}{\int G(\bar{\boldsymbol{c}}, \vec{K}, \vec{K}') P_k(\vec{K}') P_k(\vec{K}) d\vec{K} d\vec{K}'} d\bar{\boldsymbol{c}} = \lambda, \tag{S30}$$

where

$$G(\bar{\boldsymbol{c}}, \vec{K}, \vec{K}') = \frac{e^{2(\bar{c}_1 + \bar{c}_2)}[e^{-2(K_1 + K_2')} - e^{-(K_1 + K_2 + K_1' + K_2')}]}{[(1 + e^{\bar{c}_1 - K_1} + e^{\bar{c}_2 - K_2})(1 + e^{\bar{c}_1 - K_1'} + e^{\bar{c}_2 - K_2'})]^4},$$

and $\vec{K} = (K_1, K_2)$ and $\vec{K}' = (K_1', K_2')$ are the sensitivity vectors, whose components are the sensitivity of a given receptor to one of the two odorants. Multiply both side of Eq. (S30) with $P_k(\vec{K})$ and integrate over $\vec{K}$, we can determine $\lambda = \int P_{\text{env}}(\bar{\boldsymbol{c}}) d\bar{\boldsymbol{c}} = 1$.

To simplify the notation, we now define $v_1 = \exp(-K_1), v_2 = \exp(-K_2)$, i.e., v_1, v_2 are just the sensitivity in linear scale. Now we have an explicit form of ΔH when $N = 2$, we shall first use appropriate approximation to demonstrate the existence of "sparseness" in sensitivity matrix $\boldsymbol{W}$ when $N = 2$, and use numerical calculations to determine how the sparsity of sensitivity matrix and the distribution of its non-zero elements depend on input statistics. Further, we extend the case to $n = 2, N > 2$, and demonstrate that the fraction of sensitive elements at $N = 2$ is an upper bound for scenarios where $n = 2, N > 2$.

For a large number of receptors and two odorants considered here, we can assume that the distributions for the two sensitivities v_1 and v_2 are independent with the joint distribution

$$\begin{aligned}
P(v_1, v_2) =& [\rho \times p_s(v_1) + (1-\rho) \times \delta(v_1)] \\
& \times [\rho \times p_s(v_2) + (1-\rho) \times \delta(v_2)],
\end{aligned} \tag{S31}$$

where $p_s(v)$ is a smooth distribution function that is the same for all odorant-receptor pairs, and ρ is just the fraction of non-zero elements, i.e, the "sparsity" of the matrix.

We can now write down Eq. (S29) explicitly in this particular case,

$$\begin{aligned}
\chi_{11} &= M\rho[\rho\varphi_2(c_1, c_2) + (1-\rho)\varphi_1(c_1)], \\
\chi_{22} &= M\rho[\rho\varphi_2(c_2, c_1) + (1-\rho)\varphi_1(c_2)], \\
\chi_{12} &= \chi_{21} = M\rho^2\varphi_{12}(c_1, c_2),
\end{aligned} \tag{S32}$$

where

$$\begin{aligned}
\varphi_1(c) &= \int \frac{v^2}{(1+vc)^4} p_s(v) dv, \\
\varphi_2(c_1, c_2) &= \int\int \frac{v_1^2}{(1 + v_1c_1 + v_2c_2)^4} p_s(v_1) p_s(v_2) dv_1 dv_2, \\
\varphi_{12}(c_1, c_2) &= \varphi_{12}(c_2, c_1) \\
&= \int\int \frac{v_1 v_2}{(1 + v_1c_1 + v_2c_2)^4} p(v_1) p(v_2) dv_1 dv_2.
\end{aligned}$$

Plug Eq. (S32) into Eq. (S11), we have the expression for the differential entropy

$$\Delta H(\rho, p_s(v)) = \ln(\rho M) + \frac{1}{2}\int P(c_1, c_2) \ln \Sigma(c_1, c_2) dc_1 dc_2, \tag{S33}$$

where $\Sigma(c_1, c_2)$ is defined as:

$$\begin{aligned}
\Sigma(c_1, c_2) =& [\rho\varphi_2(c_1, c_2) + (1-\rho)\varphi_1(c_1)] \\
& \times [\rho\varphi_2(c_2, c_1) + (1-\rho)\varphi_1(c_2)] - \rho^2\varphi_{12}^2(c_1, c_2).
\end{aligned} \tag{S34}$$

To compare with the one odor case, Eq. (S33) can be further written as

$$I_2(\rho) = 2I_1(\rho M) + \frac{1}{2}\int p(c_1,c_2)\ln\{1+\rho[\frac{\varphi_1(c_2,c_1)}{\varphi_1(c_2)} + \frac{\varphi_2(c_1,c_2)}{\varphi_1(c_1)} - 2] + \rho^2[1 + \frac{\varphi_2(c_1,c_2)\varphi_2(c_2,c_1)}{\varphi_1(c_1)\varphi_1(c_2)} - \frac{\varphi_2(c_1,c_2)}{\varphi_1(c_1)} - \frac{\varphi_2(c_2,c_1)}{\varphi_1(c_2)} - \frac{\varphi_{12}^2(c_1,c_2)}{\varphi_1(c_1)\varphi_1(c_2)}]\}dc_1dc_2. \tag{S35}$$

Given a value of ρ, and assuming $p_s(v)$ is the same as the one odorant case, the second term of the above equation can be numerically integrated directly. As shown in Fig. 5(b) in the main text, I_2 achieves maximum value at an optimal $\rho^* < 1$.

C. Analytical approximations of optimal sparsity

With $N > 2$, $M \to \infty$, and $n = 2$, the solution for the optimal sensitivity matrix can be also formulated as a distribution of the weights $P(K)$. Since the distribution of the odorants are i.i.d., it is reasonable to assume a similar symmetric structure for the optimal $P(K)$, suggesting that $P(K)$ should be the same for different odorants, and $P(K_1, K_2)$ should be the same for different odorant pairs.

Since only two odorants appear in any mixtures, the leading term of the constraints on the distribution of K therefore comes from the "pairwise interaction"(receptors that are sensitive to both odorants are likely to be "corrupted"). For simplicity, we ignore the "interference" between different samples. As we have shall see in the next section, ignoring the "interference" term would give us an upper bound of the optimal sparsity level. Therefore we have reduced the problem to two odorants and infinite receptors. So Eq. (S33) can approximate the target function. Given the constraint $\int p_s(v)dv = 1$, the problem becomes optimizing the following function with respect to $\rho, p_s(v)$:

$$\Delta G(\rho, p_s(v), \lambda) = \Delta H - \lambda(\int p_s(v)dv - 1), \tag{S36}$$

where λ is the Lagrange multiplier. Take the derivative with respect to ρ, we have:

$$D_\rho \equiv \frac{1}{2}\int \frac{P(c_1,c_2)}{\Sigma(c_1,c_2)}\{(\varphi_1(c_1) - \varphi_2(c_1,c_2))[\rho\varphi_2(c_1,c_2) + (1-\rho)\varphi_1(c_2)] + (\varphi_1(c_2) - \varphi_2(c_2,c_1))[\rho\varphi_2(c_1,c_2) + (1-\rho)\varphi_1(c_1)] + 2\rho\varphi_{12}^2(c_1,c_2)\}dc_1dc_2 = \rho^{-1}. \tag{S37}$$

To obtain the optimal functional form of $p_s(v)$ directly through variational method is difficult, therefore we further parameterize the distribution $p_s(v)$ with a log-normal distribution, i.e., $p_s(\ln(v)) \sim \mathcal{N}(\mu_v, \sigma_v)$, which is supported by our numerical simulation. Therefore we can take derivative of Eq. (S33) with respect to μ_v and σ_v

$$D_\zeta \equiv \frac{\partial \Delta H}{\partial \zeta} = \frac{1}{2}\int \frac{P(c_1,c_2)}{\Sigma(c_1,c_2)}\frac{\partial \Sigma(c_1,c_2)}{\partial \zeta}dc_1dc_2 = 0, \quad \zeta = \mu_v, \sigma_v, \tag{S38}$$

where

$$\Sigma_\zeta(c_1,c_2) = [\rho G_2^\zeta(c_1,c_2) + (1-\rho)G_1^\zeta(c_1)] \times [\rho\varphi_2(c_1,c_2) + (1-\rho)\varphi_1(c_2)] + [\rho\varphi_2(c_1,c_2) + (1-\rho)\varphi_1(c_1)] \times [\rho G_2^\zeta(c_2,c_1) + (1-\rho)G_1^\zeta(c_2)] - 2\rho^2 G_{12}^\zeta(c_2,c_1)\varphi_{12}(c_2,c_1), \ \zeta = \mu_v, \sigma_v,$$

and

$$G_1^\zeta \equiv \frac{\partial \varphi_1(c)}{\partial \zeta} = \int \frac{v^2}{(1+vc)^4}\frac{\partial p_s(v)}{\partial \zeta}dv, \quad G_2^\zeta \equiv \frac{\partial \varphi_2(c_1,c_2)}{\partial \zeta} = \int \frac{v_1^2}{(1+v_1c_1+v_2c_2)^4}\frac{\partial [p_s(v_1)p_s(v_2)]}{\partial \zeta}dv,$$
$$G_{12}^\zeta \equiv \frac{\partial \varphi_{12}(c_1,c_2)}{\partial \zeta} = \int \frac{v_1v_2}{(1+v_1c_1+v_2c_2)^4}\frac{\partial [p_s(v_1)p_s(v_2)]}{\partial \zeta}dv_1dv_2, \ \zeta = \mu_v, \sigma_v.$$

We then solve the above three coupled equations by turning the problem into an optimization problem

$$\{\mu_v, \sigma_v, \rho\} = \arg\min_{\{\mu_v,\sigma_v,\rho\}}[(D_\rho - \rho^{-1})^2 + D_\mu^2 + D_\sigma^2]. \tag{S39}$$

The solution is shown in Fig. S5, which agrees with the direct simulation qualitatively [see Fig. 3(a)]. The quantitative difference is due to the omission of the interference between odor samples.

D. The interference between mixtures of odors

Eq. (S14) can also be approximated through a phenomenological equation by considering the interference between the two odorants. As in the main text, an intuitive mean-field equation can be written down

$$I_2(\rho M) = 2I_1(\rho M) - \alpha\rho^2 \frac{\partial I_1(\rho)}{\partial(\rho M)}, \tag{S40}$$

where $\rho \equiv \frac{m}{M}$ is what we defined as the "sparsity" of the sensitivity matrix with m the average number of receptors activated by one odorant, α is a phenomenological factor that quantifies how much the information conveyed by one receptor is "corrupted" when the receptor is active to two odorants in the same mixture. Comparing with Eq. (S35), we can see that α is a function of ρ and σ_c, i.e., $\alpha = \alpha(\sigma_c, \rho)$. The first term of Eq. (S40) increases logarithmically with m , while the second term increases supralinearly with m, and there will be a crossover point which leads to a peak of $I_2(\rho M)$.

We next extend our mean-field theory to more general cases where there are n odorants (can be larger than 2) in odor mixtures. While unable to derive analytically how ρ and $p_s(v)$ depend on input parameters, we will show that our previous result set an upper limit of ρ_w, thus introducing the interference would only cause the optimal ρ_w to be an even smaller value. Starting with $N = 3$ and $n = 2$, the mutual information is

$$I_3^2(\rho M) = 2I_1(\rho M) - \alpha\rho^2 \frac{\partial I_1(\rho M)}{\partial \rho} - (\vartheta - \alpha)\rho^3 \frac{\partial I_1(\rho M)}{\partial \rho}. \tag{S41}$$

The contribution of the higher order term depend on the difference between ϑ and α, and ϑ is a phenomenological factor that quantifies how much the information conveyed by one receptor is "corrupted" when the receptor is active to all 3 odorants. It is obvious that $\vartheta - \alpha > 0$. This can be extended to more general case, i.e., $n > 2, N > 2$ where there are n odorants in each mixture and N possible odorants, we have

$$I_N^n(\rho M) = nI_1(\rho M) - \sum_{n_0=n}^{N} \alpha_{n_0} \rho^{n_0} (1-\rho)^{N-n_0} \frac{\partial I_1(\rho M)}{\partial \rho} \tag{S42}$$

with

$$\alpha_{n_0} = \binom{N}{n} \sum_{m=max\{2,n-N+n_0\}}^{min\{n_0,n\}} \alpha_m \frac{\binom{n}{m}\binom{N-n}{n_0-m}}{\binom{N}{n_0}}$$

and

$$\alpha_m < \alpha_{m+1} < \cdots < \alpha_{min\{n_0,n\}},$$

α_m should increase with m sublinearly, because α_m describes the loss of information when a receptor is active to m odorants, and this information has a lower bound, as n increases, the loss increases but saturates gradually, and therefore α_{n_0} also increases with n_0. It is easy to see from Eq. (S41) that the interference between odor mixtures adds more negative terms, thus would lead to a smaller optimal sparsity level.

S3. NUMERICAL OPTIMIZATION PROCEDURE

In this section, we describe the procedure of numerical simulation to search for optimal sensitivity matrix in detail. We first explain how we estimate the target function, i.e., differential entropy, at each step of the iteration, and how the genetic algorithm works to find a local minimum of the target function in each trial. Then, we explain the difference between optimization procedures for the excitation-only case and for receptors with both excitatory and inhibitory response.

A. Estimation of differential entropy

The response of ORNs $\boldsymbol{r}$ is generated through Eq. (S3) using 50000 odor stimuli sampled from $P_{\text{env}}(\boldsymbol{c})$. The differential entropy of $P_r(\boldsymbol{r})$ is estimated using a copula-based method, the Gaussian Copula Mutual Information(GCMI) estimator[8, 9], which uses the concept of statistical copula to give a Gaussian parametric estimation of mutual information for variables with any marginal distributions. The differential entropy $I(\boldsymbol{r})$ is the sum of the entropy of each marginal distribution $I(r_i)$ subtracting the joint mutual information of the random variable $\text{MI}(\boldsymbol{r})$, thus

$$I(\boldsymbol{r}) = \sum_i^M I(r_i) - \text{MI}(\boldsymbol{r}), \tag{S43}$$

where $I(r_i)$ is estimated by KDP algorithm provided in the *Information theoretical estimators toolbox* [10].

B. Genetic algorithm for optimization

We use the covariance matrix adaptation-evolution strategy (CMA-ES) algorithm to do the numerical optimization[11, 12]. First, a population of initial candidates in the solution space are randomly sampled from a probability model, their performance is estimated according to Eq. (S43), and part of the best "performers" are selected and used to update the parameters in the probability model. Then, a new population of solutions are generated according to the updated probability model. The iteration keeps going until the solution converges. Since this method does not guarantee global optimal solution,

for each combination of different parameters N,M,n and σ_c, we perform many simulations from random starting points, and the results are robust with different initial conditions (very similar values of final target function). In all our simulations, the algorithm runs for 5000 iterations to ensure convergence.

C. CMA-ES algorithm

In the case where both odor-evoked excitation and inhibition exist, the sensitivity matrix $\boldsymbol{W}$ can be divided into two parts, W^A and W^I. Since the target function $H(\boldsymbol{r})$ does not continuously change as the values of W_{ij} go from positive to negative values, instead of directly searching in the $\boldsymbol{W}$-space, we search solutions in the "normalized response space"(response to unit concentration). Thus, the elements of W_{ij} is mapped to a quantity f_{ij} which we refer as a "standard response", based on whether it is excitatory or inhibitory,

$$f_{ij}^{ex} = r_{\max}[1+\gamma\frac{1}{1+W_{ij}^A}]^{-1},$$
$$f_{ij}^{in} = r_{\max}[1+\gamma(1+W_{ij}^I)]^{-1},$$

where γ is the constant that determines the basal activity of ORN. In this way, at each iteration, we update f_{ij} continuously and W_{ij} can be derived from f_{ij} at the final stage.

S4. DOWNSTREAM DECODING TASKS

The potential advantage of having maximum entropy coding at ORNs are demonstrated using two tasks: reconstruction and classification. For a set of parameters (N, M, n, σ_c), we have found the optimal sensitivity matrix $\boldsymbol{W}^*$through numerical search. The coding scheme is changed by only changing the sparsity ρ_w of $\boldsymbol{W}$ without changing the distribution of sensitive elements, i.e., they are from the same distribution as in $\boldsymbol{W}^*$. By doing so, we can control the coding capacity simple by varying ρ_w and examine its consequences on downstream decoding tasks.

A. Reconstruction task

We use a multilayer artificial neural network (ANN) as the decoder to reconstruct odor vectors. During the training phase, 10^6 sparse odor stimuli are generated from $P_{\text{env}}(\boldsymbol{c})$, and the activity patterns of ORNs they elicit are calculated by Eq. 1 in the main text, which are the inputs of the first layer of the decoding network. The decoder infers both the composition and intensity of odorants, so the cost function consists of two parts, the "identity error" and the "intensity error" [Eq. (7) in main text]. After training, the performance of ANN is tested using a new set of odors that are drawn from the same distribution as used in the training phase.

B. Classification task

Odors are presented as clusters whose centroids are sampled from $P_{\text{env}}(\boldsymbol{c})$, members in a cluster represent variations of the centroid. Denote the stimulus of a centroid as $\bar{\boldsymbol{c}}$, variations are generated according to $\mathcal{N}(\bar{\boldsymbol{c}}, \Delta S^2\bar{\boldsymbol{c}}\bar{\boldsymbol{c}}^T)$, where ΔS is termed "radius" of a odor cluster. In each simulation, each of the 100 centroids (clusters) is randomly assigned with a category (label), i.e., attraction or aversion. Each cluster contains 50 variations with the same label as the corresponding centroid, resulting in an ensemble of 5000 stimuli, which are randomly divided into training set (80%) and testing set (20%). The representations at ORNs are used as the input of the single hidden layer decoder, 500 neurons in hidden layer (KC) and a single output neuron (MBON) are used. To model the sparse KC response observed in experiments[13, 14], numbers of connection from ORNs/PN to KC are randomly drawn from a binomial distribution, resulting on average 7 random inputs from ORNs/PNs to a KC (if $M < 7$, the maximum number of connection is M). The strength of connections are in the range $[0, 1]$, which are randomly drawn from the truncated Gaussian distribution $\mathcal{N}(0.5, 0.25)$. To further encapsulate the global inhibition from APL to KCs[15], following[16, 17], the input of ith KC evoked by an ORNs/PNs response vector $\vec{r}$ (see Eq. 1 in main text) is $h_i = \langle \vec{J}_i, (\vec{r} - \langle \vec{\mu}, \vec{r}\rangle\vec{\mu})\rangle$, where $\langle \cdot, \cdot \rangle$ is the inner-product, $\vec{J}_i$ is the weight vector from ORNs/PN to KC, and $\vec{\mu}$ is the average ORNs/PNs response vector over all odors. The activity of each KCs is a rectified linear function of its input: $\Theta(h_i - \theta)$, where $\Theta(x) = x$ if $x > 0$, and $\Theta(x) = 0$ otherwise. The threshold is chosen such that on average 10% KCs respond to each odor, matching the *in vivo* experimental data[13, 18].

[1] V. Singh, N. R. Murphy, V. Balasubramanian, and J. D. Mainland, arXiv preprint arXiv:1805.00563 (2018).

[2] G. Reddy, J. D. Zak, M. Vergassola, and V. N. Murthy, eLife **7**, e34958 (2018).

[3] L.-H. Cao, D. Yang, W. Wu, X. Zeng, B.-Y. Jing, M.-T. Li, S. Qin, C. Tang, Y. Tu, and D.-G. Luo, Nature communications **8**, 1357 (2017).

[4] G. Si, J. K. Kanwal, Y. Hu, C. J. Tabone, J. Baron, M. Berck, G. Vignoud, and A. D. Samuel, Neuron (2019).

[5] E. a. Hallem and J. R. Carlson, Cell **125**, 143 (2006).

[6] H. Saito, Q. Chi, H. Zhuang, H. Matsunami, and J. D. Mainland, Sci. Signal. **2**, ra9 (2009).

[7] A. F. Carey, G. Wang, C.-Y. Su, L. J. Zwiebel, and J. R. Carlson, Nature **464**, 66 (2010).

[8] R. A. Ince, B. L. Giordano, C. Kayser, G. A. Rousselet, J. Gross, and P. G. Schyns, Human brain mapping **38**, 1541 (2017).

[9] S. Singh and B. Pøczos, arXiv preprint arXiv:1702.07803 (2017).

[10] Z. Szabó, The Journal of Machine Learning Research **15**, 283 (2014).

[11] N. Hansen and A. Ostermeier, Evolutionary computation **9**, 159 (2001).

[12] N. Hansen, in *Towards a new evolutionary computation* (Springer, 2006) pp. 75–102.

[13] G. C. Turner, M. Bazhenov, and G. Laurent, Journal of neurophysiology **99**, 734 (2008).

[14] S. J. Caron, V. Ruta, L. Abbott, and R. Axel, Nature **497**, 113 (2013).

[15] A. C. Lin, A. M. Bygrave, A. De Calignon, T. Lee, and G. Miesenböck, Nature neuroscience **17**, 559 (2014).

[16] S. X. Luo, R. Axel, and L. Abbott, Proceedings of the National Academy of Sciences **107**, 10713 (2010).

[17] K. Krishnamurthy, A. M. Hermundstad, T. Mora, A. M. Walczak, and V. Balasubramanian, arXiv preprint arXiv:1707.01962 (2017).

[18] R. A. Campbell, K. S. Honegger, H. Qin, W. Li, E. Demir, and G. C. Turner, Journal of Neuroscience **33**, 10568 (2013).

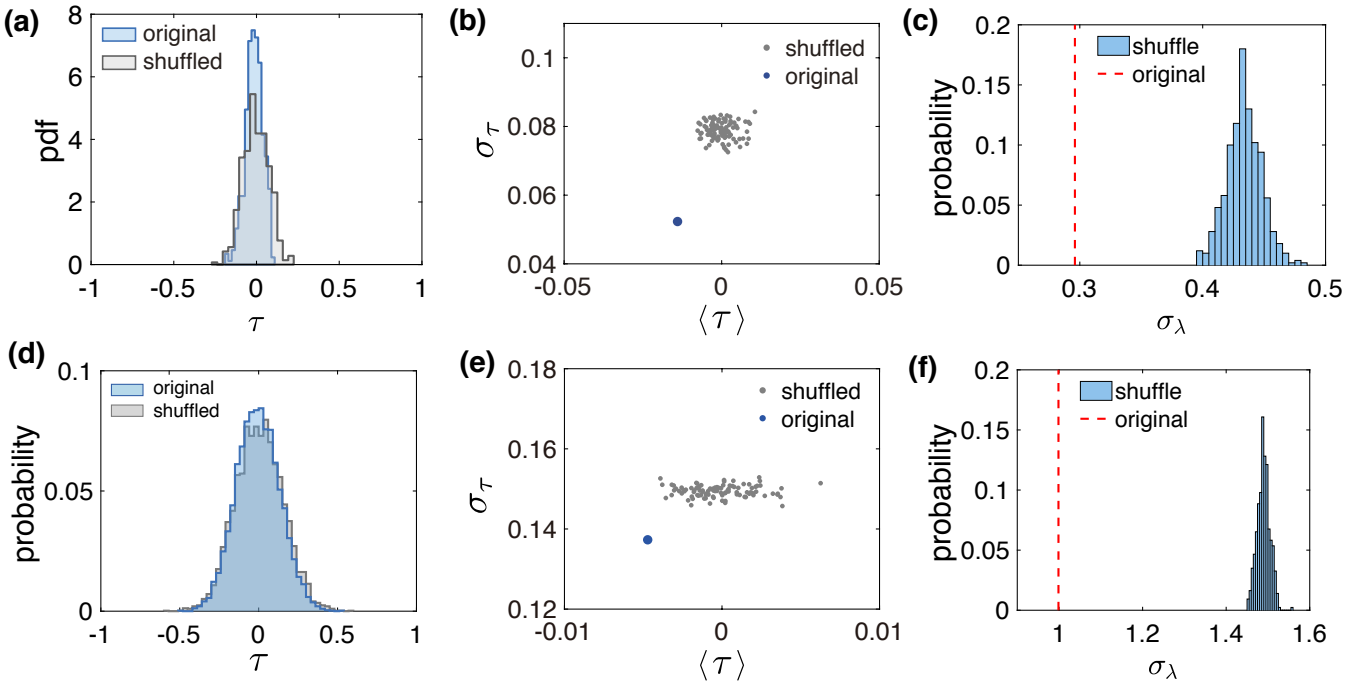

FIG. S1. (a) Histogram of row-wise (odor-wise) Kendall's tau rank correlation coefficient (τ) of the optimal matrix (blue, the matrix in Fig. 2 in main text) and its randomly shuffled matrix (gray). Note that the "insensitive" elements in the matrix has been set to be 0 before calculating the rank correlation. (b) Comparison of the average and standard deviation of τ for the original optimal sensitivity matrix (blue dot) and 100 randomly shuffled matrices (gray dots). (c) Optimal sensitivity matrix is more orthogonal than random matrix. Comparison of the eigenvalues of row-wise Kendall's tau rank correlation matrix. Histogram is the standard deviation of eigenvalues from 500 randomly shuffled matrices, red dashed line corresponds to the optimal sensitivity matrix. Notice that, as a property of correlation matrix, the average of eigenvalues is 1 for both shuffled and original correlation matrix. (d) (f) correspond to (a) (c) respectively, but for column-wise rank correlations. Parameters: $N = 100, M = 30, n = 2, \sigma_c = 2$.

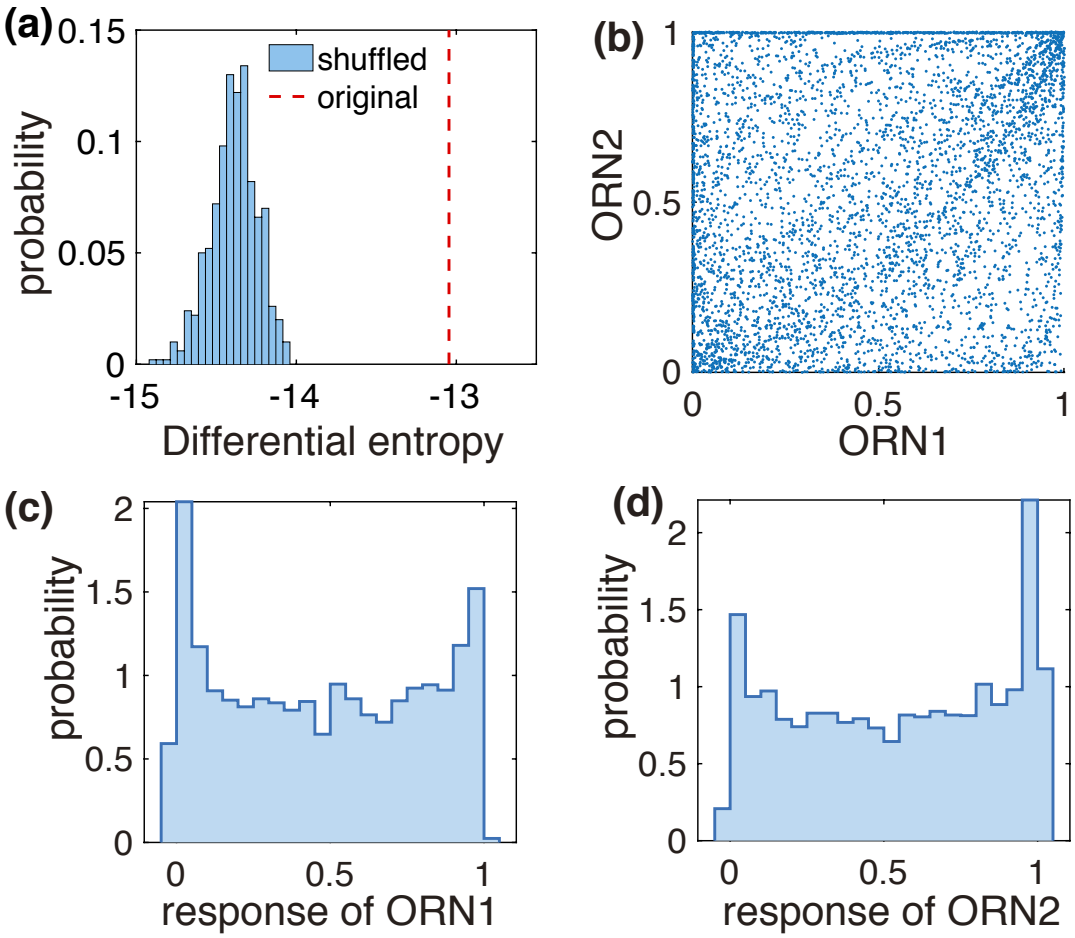

FIG. S2. (a) Comparison of the differential entropy of ORN response I(r) with optimal sensitivity matrix (dashed line) and with randomly shuffled matrix (histogram). Parameters: $N = 100, M = 30, n = 2, \sigma_c = 2$. (b)-(d) Optimal receptor-odor sensitivity matrix ensures near uniformly distributed and decorrelated ORN response patterns. (b) Scatter plot of responses of two ORNs to a sample of 10^4 odor stimuli drawn from Penv(c). (c) and (d), histogram of corresponding ORN response showing roughly uniform distribution in the whole response range. Simulation parameters used in (b)-(d) $N = 50, M = 13, n = 3, \sigma_c = 2$.

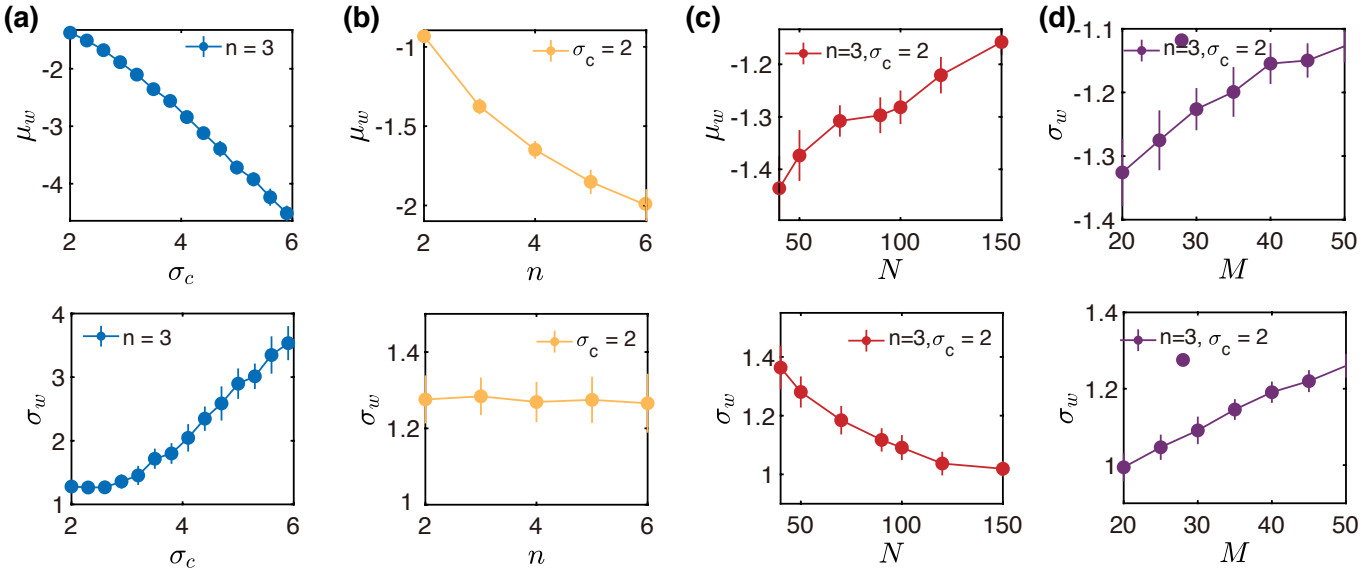

FIG. S3. Dependence of the average μ_w (upper panel) and standard deviation σ_w (lower panel) of the sensitive elements in the optimal W on the width of odor log concentration σ_c (a), the input sparsity n (b), the number of total odorants N (c), and the number of receptors M (d). In (a-b), N = 50, M = 13, in (c), M = 13, and in (d), N = 50. Error bars are standard deviation of 40 times simulation.

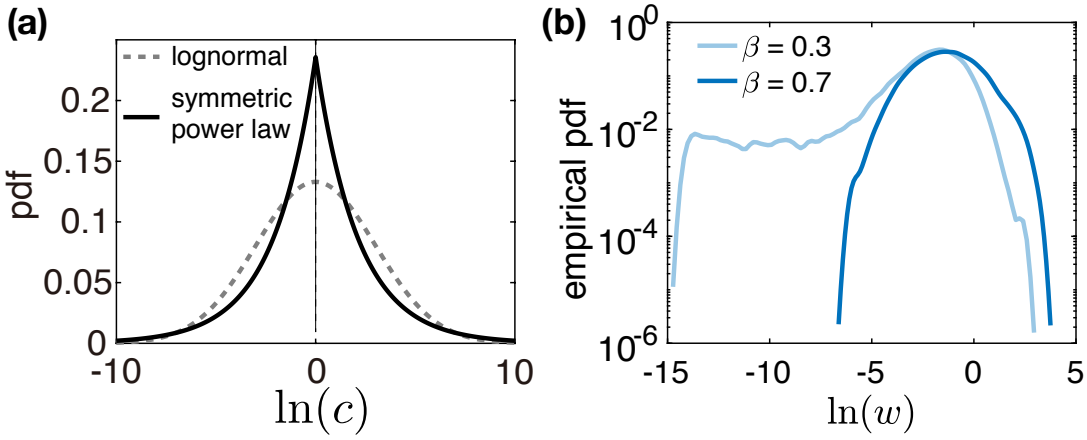

FIG. S4. (a) Odorant concentration distribution for the symmetric power law situation (solid line), compared with log-normal distribution (dashed line) with the same variance. (b) The shape of distribution of the sensitive elements depends on the exponent of the symmetric power law. Same as figure 4 (a) in the main text, but using a log scale of the y axis.

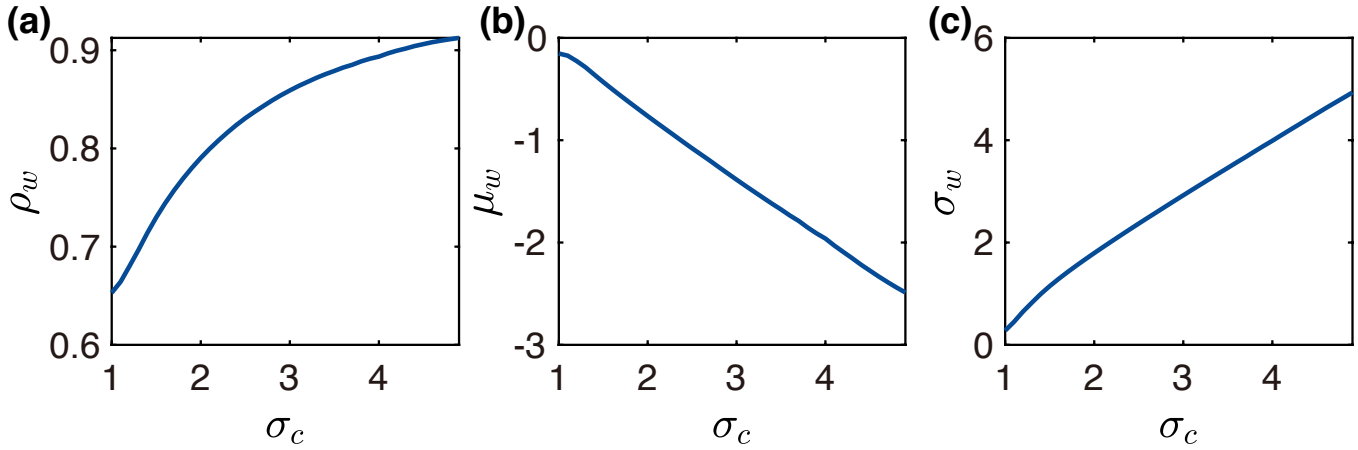

FIG. S5. Statistics of the optimal sensitivity matrix from the mean-field theory with two odorants in mixtures. The sparsity (a), average (b) and standard deviation (c) of optimal W (log scale) versus the standard deviation of log-concentration of odorant. Note the qualitatively similar trends as in Fig. 3a.

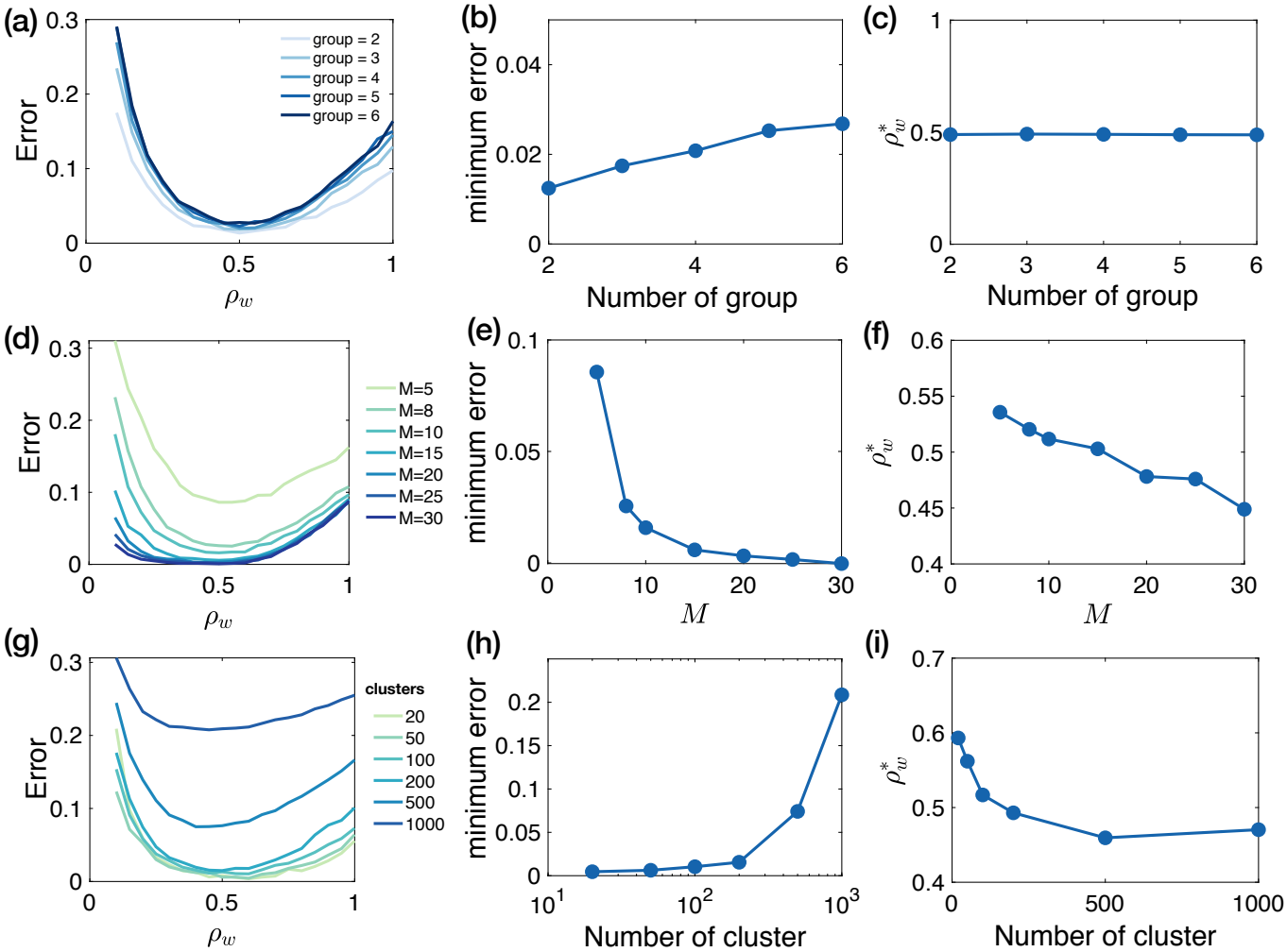

FIG. S6. Classification performance versus the sparsity of the receptor-odor sensitivity matrix with various parameters of the decoder. (a)-(c) Classification performance for different number of groups (labels) of odor stimuli. (a) Classification error (fraction of mis-classification) v.s sparsity of W. Colors indicate different number of groups that odor stimuli have been assigned. (b) Minimum classification error with respect to number odor groups. (c) For different number of odor groups, optimal performance is achieved around $\rho_w = 0.5$. Parameters used: $N = 100, M = 12, n = 2, \sigma_c = 2, \Delta S = 0.1, \sigma_0 = 0.1$, and 100 clusters. (d)-(f) Classification performance for different number of ORNs, M. (d) Classification error v.s number of ORNs. Colors indicate different M. (e) Minimum classification error with respect to number odor ORNs. (f) For different numbers of ORN, optimal performance is achieve around $\rho_w^* = 0.5$ and slightly decreases as M increases. Parameters used: $N = 100, M = 12, n = 3, \sigma_c = 2, \Delta S = 0.1, \sigma_0 = 0.1$ and two groups of labels. (g)-(i) Classification performance for different number of clusters of odor stimuli. (g) Classification error v.s sparsity of W. Colors indicate different number of clusters. (h) Minimum classification error with respect to number of clusters. (i) Optimal sparsity of W v.s the number of clusters. Parameters used: $N = 100, M = 20, n = 3, \sigma_c = 2, \Delta S = 0.1, \sigma_0 = 0.1$, 100 clusters, and two groups of labels. In all these figures, line are the average classification error, error bars have been omitted to avoid clutter.

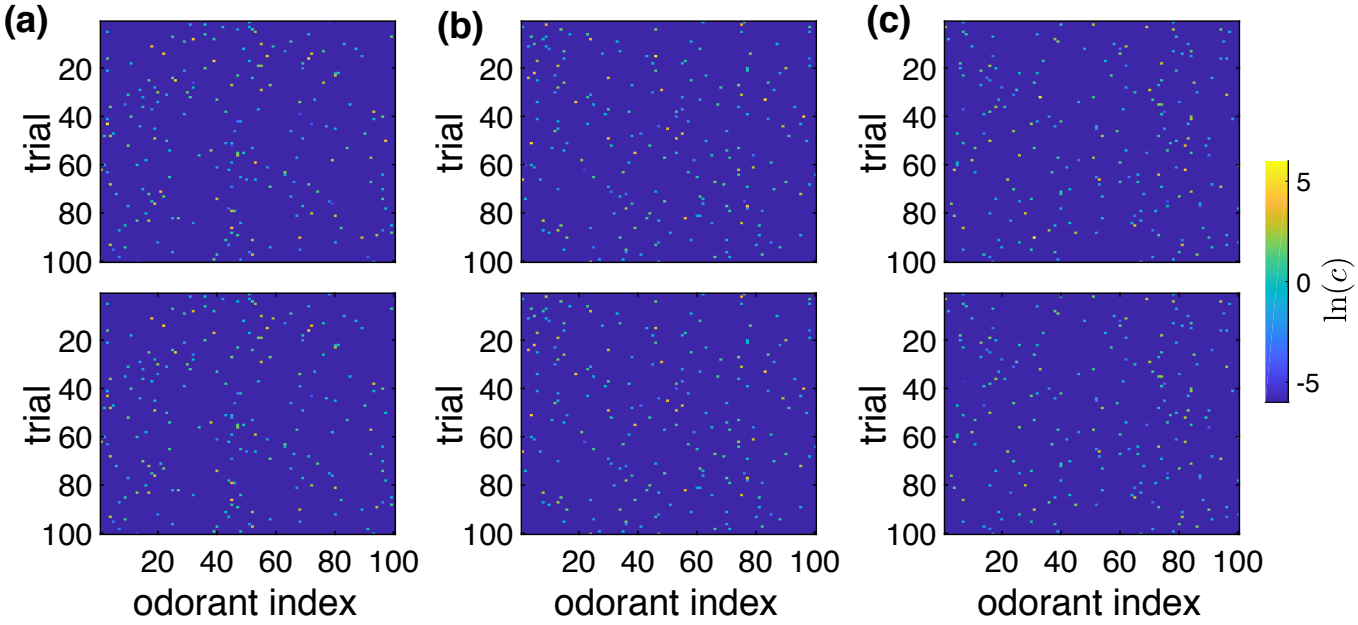

FIG. S7. Related to Fig. 5 in main text. Comparison of the reconstructed and the input odor vector when the mapping from odor space to ORN space only differ in the sparsity of sensitivity matrix W (with fixed average and standard deviation of nonzero entries, $\mu_w = -1$, $\sigma_w = 2$). Odor stimuli were randomly generated from $P_{\text{env}}(\boldsymbol{c})$. Up: 100 example sparse vectors $\boldsymbol{c}$ whose nonzero entries are coded with color (log scale). Bottom: the reconstructed sparse vector $\hat{c}$. From (a) to (c), $\rho_w = 0.1, 0.6, 0.95$ respectively. Parameters: $N = 100, M = 20, n = 2, \sigma_c = 2, \sigma_0 = 0.05$.

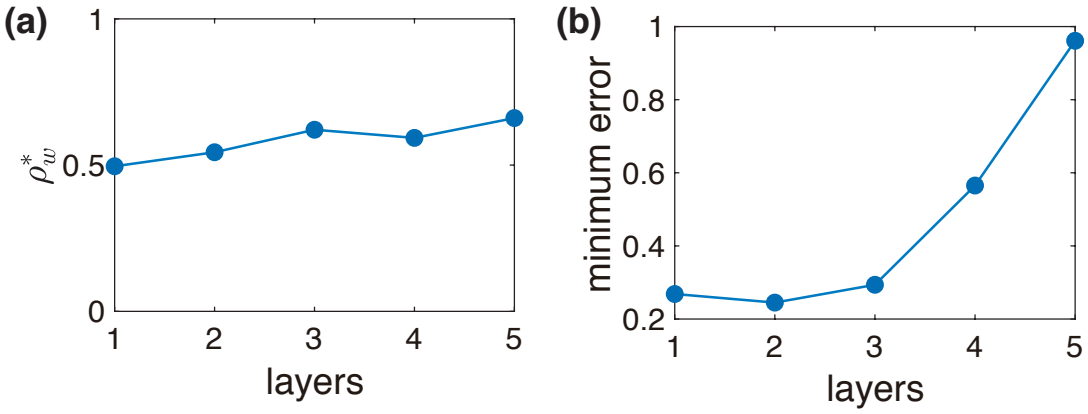

FIG. S8. Performance of reconstruction task with different number of hidden layers. (a) Sparsity of sensitivity matrix that achieves best performance v.s the number of hidden layers in the reconstruction network. (b) Minimum reconstruction error with respect to the number of hidden layers. Parameters: $N = 100$,$M = 20$,$n = 2$,$\sigma_c = 2$, $\sigma_0 = 0.05$ and each hidden layer contains 100 units.

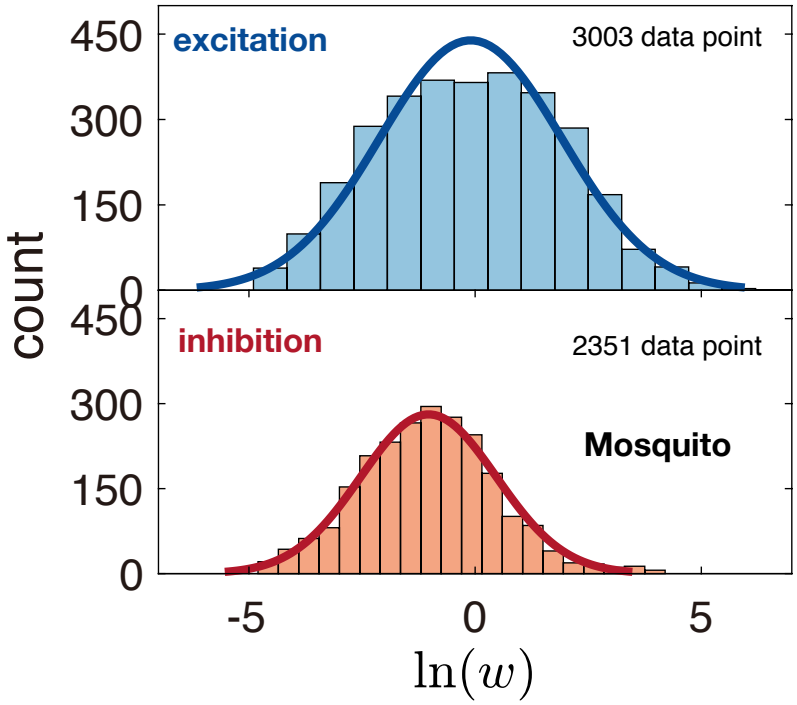

FIG. S9. The distribution of estimated relative excitatory (upper) and inhibitory (lower) receptor-odor sensitivities from experimental data for mosquito. Solid lines are log-normal distribution fit. Data are from [7].