Optimal compressed sensing strategies for an array of nonlinear olfactory receptor neurons with and without spontaneous activity

Shanshan Qin[a,1], Qianyi Li[b,1], Chao Tang[a,c,d,2], and Yuhai Tu[e,2]

[a]Center for Quantitative Biology, Peking University, Beijing 100871, China; [b]Integrated Science Program, Yuanpei College, Peking University, Beijing 100871, China; [c]School of Physics, Peking University, Beijing 100871, China; [d]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China; and [e]Physical Sciences Department, IBM T. J. Watson Research Center, Yorktown Heights, NY 10598

Edited by William Bialek, Princeton University, Princeton, NJ, and approved August 22, 2019 (received for review April 16, 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. For ORNs without spontaneous (basal) activity, 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 analytically that sparsity in the optimal ORIN originates from a trade-off between the broad tuning of ORNs and possible interference. Furthermore, 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 inhibitory odor–receptor interactions increases the coding capacity and the fraction of inhibitory interactions increases with the ORN basal activity. We argue that basal activities in sensory receptors in different organisms are due to the trade-off between the increase in coding capacity and the cost of maintaining the spontaneous basal activity. 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.

olfaction | information theory | coding | olfactory receptor neurons

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 ~ 300 ORs (6, 7), and the often-cited number of odors that can be distinguished is $\sim 10,000$ (8); the real number may be even larger (9) (see also refs. 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 ~ 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 (14–17). In both mammals and insects, odorants bind to receptors in the dendrites or cilia of olfactory receptor neurons (ORNs), each of which expresses only 1 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, 18–20). ORNs also exhibit a certain level of spontaneous activity in the absence of odor stimuli (21–23), and such basal activity can be suppressed by some odorants (24, 25). While largely regarded as unavoidable noise, the functional rule of spontaneous activity and odor-evoked inhibition in odor coding has only been revealed in a recent study (26). The key question that we want to address in this paper is how ORNs with these characteristics best represent external olfactory information that can be interpreted by the brain to guide an animal's behavior (4, 13, 27).

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 (28, 29) or the related InfoMax principle (30, 31). 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 (29). The receptive fields of neurons in the early visual pathway are thought to exploit statistics of natural scenes (32–36). A similar result has also been observed in the auditory system (37). In all these cases, to achieve maximum information transmission an "ideal" neuron should transform the input distribution into a uniform output

Significance

Natural odors are typically mixtures of a few odorants from a large number of possible odorants, each with a broad distribution of concentrations. Here, we study how a relatively small number of neurons with a limited response range can optimize transmission of such high-dimensional sparse odor mixture information. For neurons without basal activity, we find that the optimal coding matrix is "sparse"—a fraction of ligand-receptor sensitivities are zero and the nonzero sensitivities follow a broad distribution matching the odor mixture statistics. We show that this maximum entropy code enhances performances of the downstream reconstruction and classification tasks. For neurons with a finite spontaneous activity, we show that introducing odor-evoked inhibition further enhances coding capacity.

Author contributions: S.Q., C.T., and Y.T. designed research; S.Q., Q.L., and Y.T. performed research; S.Q., Q.L., C.T., and Y.T. analyzed data; and S.Q. and Y.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]S.Q. and Q.L. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: tangc@pku.edu.cn or yuhai@us.ibm.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1906571116/-/DCSupplemental.

distribution (29, 38) and a population of neurons should decorrelate their responses (35, 39, 40).

There are a huge number of odorants in the environment, each with its own unique molecular structure and different physiochemical properties (41, 42), which 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, 43). The sparsity of the odor mixture immediately reminds us of the powerful compressed sensing (CS) theory developed in computer science and signal processing community (44, 45). The CS theory shows that sparse high-dimensional signals can be encoded by a small number of sensors (measurements) through random projections, and the highly compressed signal can be reconstructed (decoded) with high fidelity by using an L_1-minimization algorithm (44–46). However, conventional CS theory assumes the sensors to have a linear response function with essentially an infinite dynamic range (47). In contrast, the ORN response is highly nonlinear (24, 48), with a typical dynamic range less than 2 orders of magnitude, which is far less than the typical concentration range of odorants (48).

The use of the CS theory has recently been explored in olfactory systems. For example, Zhang and Sharpee (49) proposed a fast reconstruction algorithm in a simplified setup with binary ORNs and binary odor mixtures without concentration information. In another work, Krishnamurthy et al. (50) studied how the overall "hour-glass" (compression followed by decompression) structure of the olfactory circuit can facilitate olfactory association and learning, with the assumption that ORN responses to odor mixtures are linear. Following ideas in CS theory, Singh et al. (51) recently proposed a fast olfactory decoding algorithm that might be implemented in the downstream olfactory system.

Another related study is the recent work by Zwicker et al. (52) 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 2 conditions for the binary ORNs to maximize the information transmission. The first condition is that each ORN on average responds to half of the odors, that is, 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 with a binary input–output response function and a specific prior for the sensitivity distribution (52).

In all of these previous studies, however, strong simplifying assumptions were made about neural response functions, treating them as either binary or linear. Thus, 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, we systematically study the optimal compressed coding strategy for an array of nonlinear ORNs with and without a finite basal activity.

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) $j(=1, 2, \ldots, N)$ 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 nonzero 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 $\boldsymbol{r} = (r_1, r_2, \ldots, r_M)$, 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 the i-th sensor (ORN) to the j-th odorant for all odorant–ORN pairs with $j = 1, 2, \ldots, N$ and $i = 1, 2, \ldots, M$.

For ORNs without spontaneous activity, we used a simple competitive binding model (53, 54) (also see *SI Appendix*), in which the normalized response of ORN i $(=1, 2, \ldots, M)$ 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 and η_i represents the noise. For convenience, η_i is assumed to be Gaussian with zero mean and standard deviation (SD)σ_0. In this paper, we consider the case with a small but finite noise $\sigma_0 \ll 1$. Other forms of the nonlinear response function and noise can be used without affecting the general conclusions.

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.

Eq. **1** maps the external odor stimulus $\boldsymbol{c}$ to the internal neuronal activity $\boldsymbol{r} = (r_1, r_2, \ldots, r_M)$. 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 (55).

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

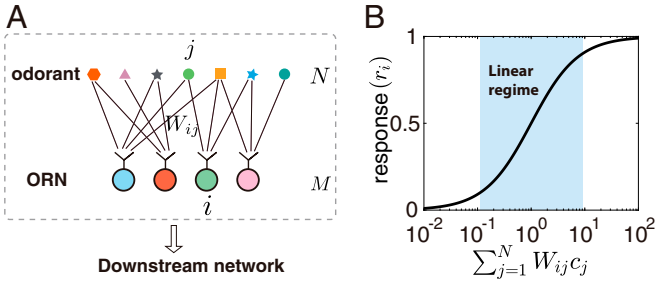

Fig. 1. Schematics of peripheral odor coding. (*A*) Illustration of interaction between odorants and ORs. There are *N* possible odorant molecules and *M* ORNs. The interaction between the *i*-th ORN and the *j*-th odorant is characterized by the sensitivity W_{ij}. The odorant–ORN interactions in the peripheral olfactory system (the dotted box) are characterized by the $(M \times N)$ sensitivity matrix $\boldsymbol{W}$. The odor information collected by the ORNs is 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.

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 is received by the peripheral ORNs and thus serves as a good "target function" to be maximized (30, 31, 52, 56–58). I is defined as

$$I(\boldsymbol{c};\boldsymbol{r}) = H(\boldsymbol{r}) - H(\boldsymbol{r}|\boldsymbol{c}) = -\int P_r(\boldsymbol{r}) \log P_r(\boldsymbol{r})\, d\boldsymbol{r} + \iint P(\boldsymbol{r},\boldsymbol{c}) \log P(\boldsymbol{r}|\boldsymbol{c})\, d\boldsymbol{r}\, d\boldsymbol{c}, \quad [2]$$

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 this paper, we consider the limit of small noise when the second term is independent of $\boldsymbol{W}$; 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 $P_{\text{env}}(\boldsymbol{c})$:

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

The optimal sensitivity matrix can be derived by maximizing the mutual information I or equivalently the differential entropy H with respect to $\boldsymbol{W}$ 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 simple cases. For more general cases, we used the covariance matrix adaptation evolution strategy (CMA-ES) algorithm to find the optimal sensitivity matrix (59, 60) (see *Materials and Methods* and *SI Appendix* for technical details).

The optimal sensitivity matrix is sparse for ORNs without basal activity. Odor concentration varies widely in the natural environment (1, 61). To capture this property, we studied the case where the odorant concentrations in an odor mixture follow 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; a more realistic consideration such as correlation among odorants will be discussed later in *Summary and Discussion*.

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 of 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, the sparsity parameter) of its elements have nonzero 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}. From the histogram of $\ln(W_{ij})$ shown in Fig. 2*B*, it is clear that elements in the optimal sensitivity matrix fall into 2 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, that is, sparsity in the odor–receptor sensitivity matrix, is supported by existing experimental measurements. As shown in Fig. 2*C*, the sparsity parameter ρ_w is estimated to be ~ 0.4 for fly larva (48) and ~ 0.1 for mouse (62). For locust, the sparsity parameter of the odor–ORN interaction matrix is estimated to be $\sim 23\%$ from the experiments reported in ref. 63, and the response sparsity of projection neurons (PNs)

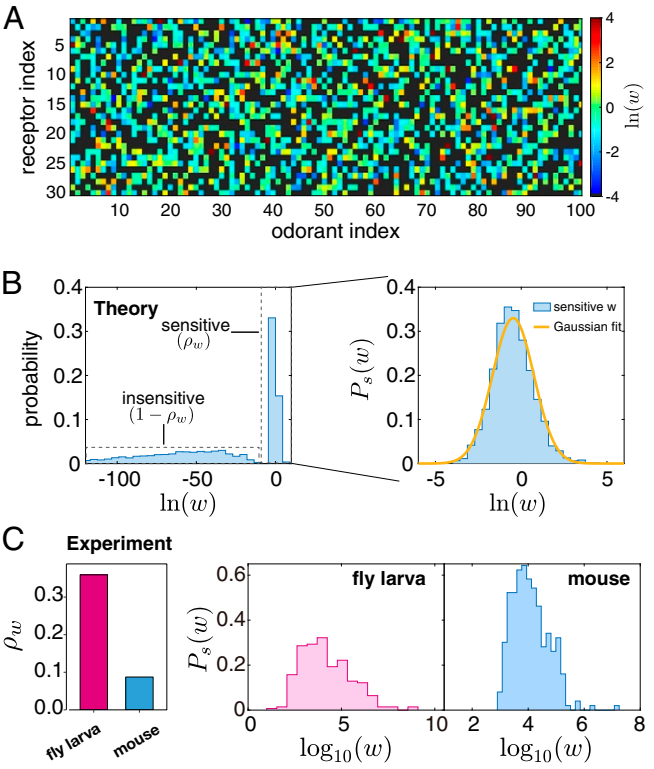

Fig. 2. Statistics of the optimal sensitivity matrix elements from theory and comparison with experiments. (*A*) Heat map of a typical optimal sensitivity matrix from our model. Color indicates the value of $\ln(W_{ij})$, and black indicates the "inactive" or negligible interactions. (*B*) Histogram of all of 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 (48) and mouse (62). (*Right*) The histogram of sensitive W_{ij}, $P_s(w)$, for fly larva and mouse. Model parameters are $N = 100$, $M = 30$, $n = 2$, $\sigma_c = 2$, and $\mu = 0$.

was found to be $\sim 50\%$ (64). Given the high concentration of odorants used in these experiments, the "zero" sensitivities cannot be simply explained by the limitation of experimental detection accuracy, although it is possible that some of the "zero" sensitivities are caused by inhibitory odorant–ORN interactions due to limitations of the measurement methods (*Summary and Discussion*). However, measurements of the excitatory odorant–ORN interactions are unambiguous and they show only a finite fraction of excitatory odorant–ORN interactions, consistent with our theory. Furthermore, the broad distribution of the excitatory sensitivities obtained in our model also agrees qualitatively with those estimated from experiments (Fig. 2 *C*, *Right*), 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. We found that both the rows and columns (*SI Appendix*, Fig. S1) 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 (*SI Appendix*, Fig. S2*A*) and a nearly uniform distribution of ORN activity for different odor mixtures (*SI Appendix*, Fig. S2 *B–D*).

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 parameter ρ_w defined as the fraction of nonzero 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 SD (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 parameter decreased (Fig. 3 *B* and *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, that is, the sensitivity matrix becomes more sparse as the number of sensors M increases (Fig. 3*D*). This somewhat counterintuitive result can be understood thus: As the system has more sensors to encode signals, each sensor can respond to a smaller number of odors to avoid interference. For all of the cases we studied, when the odor concentrations follow a log-normal distribution, then the distribution of the nonzero sensitivities in the optimal sensitivity matrix follows roughly a log-normal distribution with its mean μ_w and SD σ_w depending on the odor statistics (σ_c, n, N) and the number of ORNs M (*SI Appendix*, Fig. S3).

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 *SI Appendix*, Fig. S4 for the comparison with log-normal distribution), with different exponent β. For all values of β studied, there is always a finite sparsity parameter $\rho_w < 1$ 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 an exact log-normal distribution (*SI Appendix*, Fig. S4*B*). In fact, $P_s(w)$ is asymmetric in the $\ln(w)$ space with a skewness that depends on β as shown in Fig. 4 *B*, *Inset*.

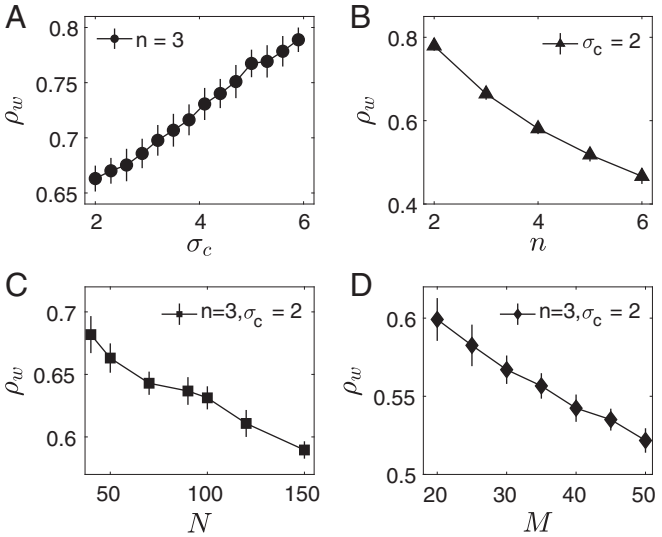

Fig. 3. Dependence of ρ_w on the width of odor log concentration σ_c (*A*), the input sparsity n (*B*), the number of total odorants N (*C*), and the number of receptors M (*D*). In *A* and *B*, $N = 50$, $M = 13$; in *C*, $M = 13$; and in *D*, $N = 50$. Error bars are SD of 40 times simulation.

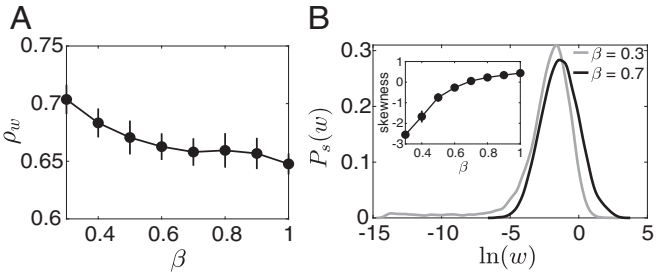

Fig. 4. The optimal sensitivity matrix for the symmetric power-law odor concentration distribution $P_{\rm env}(c) \propto \exp[-\beta|\ln c|]$. (*A*) The sparsity parameter ρ_w versus the power-law exponent β. (*B*) The distribution of the nonzero sensitivities $P_s(w)$ for $\beta = 0.3, 0.7$. (*Inset*)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

Taken together, our results suggest that sparsity in the sensitivity matrix is a robust feature for nonlinear CS problems. This theoretical existing experiments in olfactory systems (48, 62). Our study also showed that the nonzero sensitivities follow a broad distribution whose exact shape, mean, and variance depend on odor statistics and total number of ORNs.

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, that is, $M \ll N$, each sensor needs to respond to (sense) multiple types of odorant molecule so that all odorant molecules can be sensed by at least 1 sensor. However, in an odor mixture with a few types of odorant molecules, 2 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 trade-off between sensing multiple odorants and the possible interference determines the sparsity in the optimal sensitivity matrix. We demonstrate this trade-off and its effect more rigorously by developing a mean-field theory (MFT) as described below.

We begin with the simplest case in which there is only 1 odorant in the environment sensed by many receptors ($N = 1, M \gg 1$). Since the dynamic range of a nonlinear ORN is finite due to finite noise and response saturation in the ORN output (see Eq. **1**), multiple ORNs are required to represent (code) the concentration information of even a single odorant if the odorant concentration has a larger dynamic range. Obviously, there is no interference in this case. As first proposed by Laughlin (29), 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 an SD σ_c, the optimal sensitivity distribution $P_1(w)$ that maximizes $H(\boldsymbol{r})$ is also approximately a log-normal distribution:

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

where the mean $\mu_w = 0$ and the variance σ_w^2 increase 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 *SI Appendix* for details), which is verified by simulation results as shown in Fig. 5*A*:

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

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

We next consider the case where 2 odorants are sensed by multiple receptors ($N = 2, M \gg 1$). Let's denote 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 1 set will also respond to the other odorant. In the conventional linear compressive sensing scenario, the sensors respond linearly to the combination of input signals and therefore have effectively an infinite dynamic range; this linearly "compressed" representation can thus be "demultiplexed" during the decoding process, leading to accurate reconstruction of the input signals. In contrast, a realistic ORN has only a finite dynamic range and can reach its saturation level more easily when 2 or more odorants excite the ORN simultaneously. This interference effect, which is caused by saturation of the ORN response, constrains its coding capacity. On average, there are $m \times p = m^2/M$ receptors whose output is "corrupted" due to interference between 2 different odorants in a given mixture. We can write down the differential entropy as

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

where $I_1(m)$ is the maximum differential entropy for 1 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.

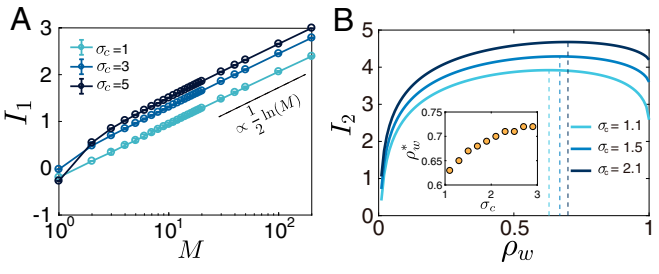

Fig. 5. The trade-off between increasing single odorant information and interference among multiple odorants. (*A*) The differential entropy with 1 odorant, I_1, versus the number of receptors M for different width (σ_c) of odor log-concentration distribution. I_1 increases monotonically with M but it only grows logarithmically with M for large M (dashed line). (*B*) Differential entropy I_2 for the case with 2 odorants in the mixture with their concentrations following the same log-normal distribution with width σ_c. I_2 depends nonmonotonically 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.

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 right-hand side 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 parameter 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 independent and identically distributed. In particular, we used the following approximation for the distribution of the sensitivity matrix $\boldsymbol{W}$:

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

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 SD 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 *SI Appendix* 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 (*SI Appendix*, Fig. S5).

The Optimal Sparse Sensitivity Matrix Enhances Downstream Decoding Performance. The response patterns of ORNs form the internal representation of external odor stimuli which is further inferred and interpreted by the brain to guide the organism's behavior. Here in this section we test whether the optimal sensitivity matrix that enables maximum odor information transmission by ORNs can enhance the downstream decoding performance by examining 2 specific learning tasks: classification and reconstruction.

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 is illustrated in Fig. 6*A*. After odor signals are sensed by ~ 50 ORNs, they are relayed by the PNs in antennal lobes to a much larger number of Kenyon cells (KCs) in the mushroom body (MB). Each KC receives sparse random connections from PNs (65). A single GABAergic neuron (APL) at each side of the brain forms negative feedback with all KCs (66). Olfactory learning is mainly mediated by the dopaminergic neurons (DANs), which control the synaptic weights between KCs and MB output neurons (MBONs) (67).

Our model "classifier" network mimics the properties of MB, as illustrated in Fig. 6*B*. It contains a high-dimensional mixed layer (KCs) and a single readout neuron for simplicity. Each KC unit pools the ORNs with a fixed random, sparse matrix. Only the synaptic weights from the KCs to the readout neuron are plastic. We assumed that odor stimuli fall into clusters whose centers represent corresponding typical odor stimuli with the radius of a cluster ΔS characterizing the variability (68) (Fig. 6*C*). Each cluster is randomly assigned with a label (attractive or aversive) (see *Materials and Methods* and *SI Appendix* for details).

The synaptic weights from the KCs to the readout neuron are trained by using a simple linear discriminant analysis method,

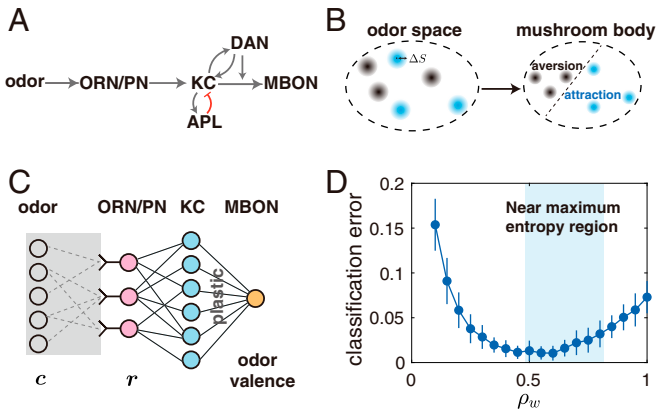

Fig. 6. Maximum entropy coding facilitates olfactory learning and classification. (*A*) Schematics of the neural circuitry for olfactory learning in the fly; see the text for detailed description. (*B*) A simplified model of the fly olfactory system shown in *A* for learning the valence of odor stimuli, where the effect of DANs 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. One hundred clusters were drawn and each cluster contains 50 variations, resulting in 5,000 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 2 odor categories. Error bars are SD from 40 simulations.

although other linear classification algorithms such as support vector machine 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 nonzero 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 gives similar results (*SI Appendix*, Fig. S6). In line with recent studies which show that sparse high-dimensional representation facilitates downstream classification (68, 69), 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.

Motivated directly by the original CS problem in computer science, we also examined how the peripheral coding regime affects the downstream reconstruction task using a generic feedforward artificial neural network as the "decoder." We found that the reconstruction error depends on the coding matrix $\boldsymbol{W}$, in particular its sparsity parameter ρ_w. The best performance is achieved around $\rho_w = 0.6$, within the region where sparse $\boldsymbol{W}$ enables nearly maximum entropy coding (*SI Appendix*, Figs. S7 and S8); this property is insensitive to the number of hidden layers in the reconstruction network (*SI Appendix*, Fig. S9).

The 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 (23, 24, 70), as shown in Fig. 7*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 (26). It is then natural to ask what the optimal design of the sensitivity matrix is to maximize coding capacity if odorants can be either excitatory or inhibitory. To answer this question, we used a 2-state model to characterize

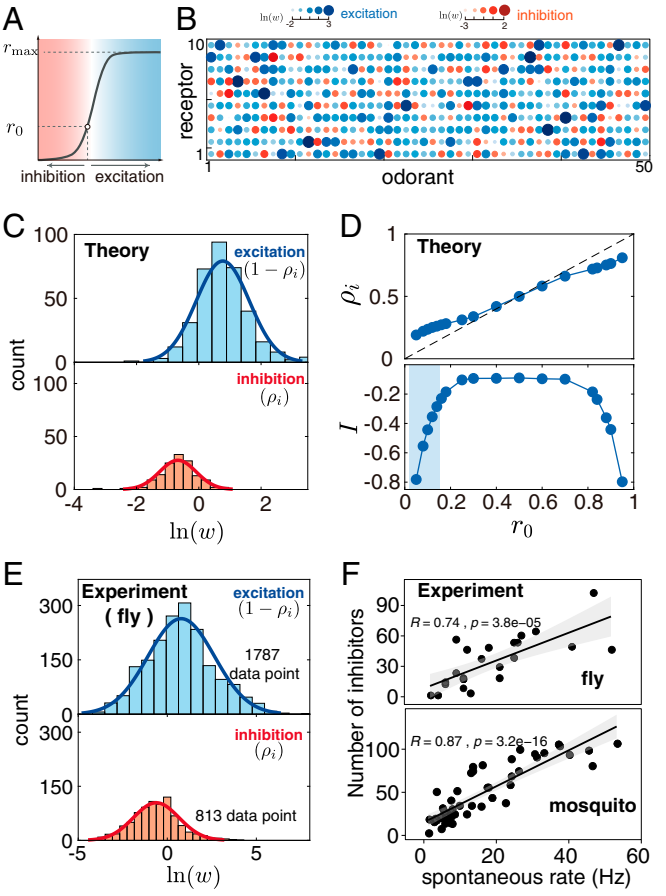

Fig. 7. The optimal sensitivity matrix for ORNs with a finite basal activity and comparison with experiments. (*A*) Schematic of ORN response to excitatory (blue region) and inhibitory odorants (red region). Note that the neuron has a finite acitivity r_0 in the absence of any stimulus. (*B*) Heat map 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*). The differential entropy I also increases with r_0 (*Lower*). The shaded region shows the range of r_0 corresponding to the fraction of inhibitory interaction estimated from experiments (24), which coincides 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 (24). 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 the fly (24) (*Upper*) and mosquito (71) (*Lower*). Each point corresponds to an ORN, the line is the linear fit, and the 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*, $r_0 = 0.18$.

both odor-evoked excitation and inhibition (26). Now, the interaction between the odorant j and ORN i has 2 possibilities: It can be either excitatory with a sensitivity W_{ij}^A or inhibitory with a sensitivity W_{ij}^I. The normalized response of i-th ORN to odor mixture $\boldsymbol{c}$ is

$$r_i = \left[1+\gamma\frac{1+\sum_{q=1}^{n_I} W_{iq}^I c_q}{1+\sum_{p=1}^{n_A} W_{ip}^A c_p}\right]^{-1} + \eta_i, \qquad [8]$$

where $\gamma(>0)$ 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. 7*B*). The strength (sensitivity) of both the excitatory and inhibitory elements follows (approximately) log-normal distributions (Fig. 7*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. 7*D*, *Upper*). 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 (*SI Appendix*, Fig. S10). As the basal activity increases, the coding capacity increases rapidly at first and quickly plateaus around $r_0 = 0.3$ (Fig. 7 *D*, *Lower*). 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 bidirectionally 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 the fly by Hallem and Carlson (24) as well as in the mosquito by Carey et al. (71). As shown in Fig. 7*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. 7*C*. The mosquito data show very similar results (*SI Appendix*, Fig. S12). Our theory also showed that the fraction of inhibitory interaction ρ_i increases with the basal activity r_0, as shown in Fig. 7 *D*, *Upper*. We have tested this theoretical result from the experimental data. As shown in Fig. 7*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. Such positive correlation is still present with a high statistical significance even with a more stringent definition of excitatory and inhibitory interactions (*SI Appendix*, Fig. S11). Finally, we note that the relative basal activity $\langle r_0 \rangle$ from the experimental data (24) is smaller than 0.16 (see *SI Appendix* for detailed analysis), where the differential entropy rises sharply with r_0 as highlighted by the shaded region in Fig. 7*D*, *Lower*. Although an even higher spontaneous activity toward $r_0 = 0.5$ can further increase the coding capacity, the gain is diminishing, while the metabolic cost increases drastically in maintaining the spontaneous activity (72). Thus, an optimal basal activity would be expected in the shaded region of Fig. 7*D* due to the trade-off between coding capacity and energy cost.

Summary and Discussion

To summarize, 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 odor information. We found that the optimal compressed coding strategy depends on whether a given ORN has a finite basal activity. For a neuron without basal activity, the best strategy is for it to only respond to a finite fraction $\rho_w(<1)$ of odorants with its sensitivities (to those odorants that it responds to) following a broad distribution and the rest $(1-\rho_w)$ fraction of the sensitivities are zero. The optimal sparsity parameter ρ_w depends on the odor mixture statistics and the number of ORNs. The sparsity in the odor–ORN sensitivity matrix is caused by the trade-off between the broad-tuning of ORNs and odor interference as a direct consequence of the finite dynamic range of the realistic nonlinear ORNs. For a neuron with a finite basal activity r_0, the optimal strategy is to have a finite fraction ρ_i of the odor–ORN interactions be inhibitory and the rest $(1-\rho_i)$ fraction of the odor–neuron interactions be excitatory. The inhibitory fraction ρ_i increases with the basal activity r_0.

The optimal strategies for the 2 types of neurons are consistent with each other. When the basal activity diminishes ($r_0 \to 0$), ρ_i approaches a finite value that is approximately the same as $1-\rho_w$. This is intuitively clear as the inhibitory interactions in the limit of $r_0 = 0$ are just the null interactions (with zero sensitivity) found in neurons without basal activity. In realistic olfactory systems, the population of ORNs is likely to be mixed: Some ORNs have no (or very small) basal activity and some have finite basal activities. In this work, we studied these 2 types of neurons separately for simplicity. However, the main results and conclusions do not change when we consider the mixed case with both types of ORNs together (*SI Appendix*, Fig. S10).

Comparison with Existing Experiments and Testable Predictions. A general conclusion from our theory is that the optimal odorant–ORN interaction matrix is only sparsely populated with excitatory interactions. In other words, there is always a finite fraction of nonexcitatory (zero or inhibitory) odorant–ORN interactions—$(1-\rho_w)$ for neurons without basal activity and ρ_i for neurons with basal activity. This general theoretical result, that is, the sparsity of the excitatory odorant–ORN interactions, is consistent with existing experimental measurements of receptor–odor sensitivity matrices in different organisms (fly larva, mouse, adult fly, and mosquito) by different measurement methods as shown in Figs. 2*C* and 7 *E* and *F*. Although the natural odor environment varies for different organisms, the broad distribution of the nonzero sensitivity obtained in our models is also consistent with the sensitivity matrices estimated from experiments in these different organisms (Figs. 2*C* and 7 *E* and *F*). The optimal coding strategy, if it 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 OR 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 (73). A recent study shows that odor-evoked inhibition can code olfactory information that drives the behavior of the fly (26). 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 the adult fly and mosquito are consistent with this prediction (Fig. 7*F*). However, high-throughput techniques such as calcium imaging, which only indirectly measure the odor–ORN interaction, seem to be incapable of detecting spontaneous activity and inhibitory response (48). Therefore, more large-scale direct measurements using electrophysiological methods such as those done for *Drosophila* (24, 25) and mosquito (71) should be carried out to test our predictions in different organisms.

By considering how the coding capacity of ORNs changes with basal activity (Fig. 7*D*) and the associated extra energy cost (72), 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 *Escherichia coli* has 5 chemoreceptors (74) which work bidirectionally with a high basal activity $r_0 \approx 1/3 - 1/2$ (75), and r_0 in the mouse is smaller than that in the fly (70). Of course, more experiments across different organisms with different numbers of sensory neurons are needed to test this hypothesis.

Possible 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 (76), it happens in a relatively slower time scale than the time required for animals to detect and respond to odor stimuli (77, 78). 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 (76, 79). Therefore, even though ORN level adaptation can further increase coding capacity at a slightly longer time scale as shown recently by Kadakia and Emonet (80), 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 2 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, that is, odorant mixture is perceived as a unit odor (17). 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 (81). 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 (24, 82). 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 CS theory with linear sensors, a random measurement matrix enables accurate reconstruction of sparse high-dimensional input signals (44, 47). By using prior information about the input, a better sensory matrix can be designed (83, 84). In many cases, the optimal matrix maximizes the entropy of compressed representation (85). 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.

In olfactory systems, ORNs of the same type converge to the same glomerulus in the second olfactory center (antennal lobe in insects and olfactory bulb in mammals). Such convergence in the glomerulus serves the purpose of controlling (averaging out) the response noise of individual ORN. Since the output considered in our study is essentially the averaged response of all of the same type of ORNs, we take the constant small noise approximation in this study and focus on the deterministic part of the response function. It would be interesting to consider explicitly the intrinsic noise in a single ORN and the noise introduced in the "averaging" process in the glomerulus when there are more information and data available about these noise sources.

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. That is, the rowwise or columnwise correlation coefficients are both centered around 0 but have a narrower distribution than those calculated from the randomly scrambled sensitivity matrix (*SI Appendix*, Fig. S1). In naturally occurring odor mixtures, co-occurrence of odorants in different odor sources is common. For example, odorants that are products in the same biochemical reaction pathway, for example fermentation, are likely to appear together (2, 86). The topography of odor space induced by the co-occurrence of odorants was recently studied by Zhou et al. (87). Although odorant-evoked ORN response patterns are not simply determined by the molecular structure, some very similar odorants do trigger similar ORN response patterns (24). On the other hand, ORNs and their responses to different odorants can be correlated due to structural similarities in their receptor proteins. Teşileanu et al. (88) recently studied a complementary problem. Given a fixed total number of ORNs, they studied the optimal fraction of the each ORN type to maximize olfactory information by considering the correlations in receptor responses. 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 future studies.

Materials and Methods

Numerical Optimization of Sensitivity Matrix. We use the CMA-ES algorithm to search the optimal sensitivity matrix (59, 60). At each iteration, a population of candidate sensitivity matrices are sampled and their performance are estimated. The subpopulation that have better performance then determines the next generation of candidates. The iteration keeps going until the solution converges. For each parameter set, we perform many simulations from random starting points. To estimate the differential entropy $I(\boldsymbol{r})$ of response pattern of ORNs $\boldsymbol{r}$, we first use the Gaussian Copula mutual information estimator to estimate the joint mutual information of the random variable MI($\boldsymbol{r}$) (89, 90); we then use the KDP algorithm (91) to estimate the entropy of each marginal distribution $I(r_i)$, then $I(\boldsymbol{r}) = \sum_i^M I(r_i) - \mathrm{MI}(\boldsymbol{r})$ (see *SI Appendix* for details).

OR–Odorant Sensitivities from Experiment. For fly larva and mouse, the sensitivities of ORs to different odorants are directly reported in refs. 48 and 62, which is used in Fig. 2*C*. For the adult fly and mosquito, both odor-evoked responses of ORNs to different odorants and spontaneous firing rate were reported in refs. 24 and 71. We assume all ORNs have the same maximum firing rate and estimate their sensitivities to excitatory and inhibitory odorants using Eq. **8** (see *SI Appendix* for details).

Downstream Decoding. In the classification task, centroids of odor clusters are sampled from $P_{\mathrm{env}}(\boldsymbol{c})$. Each cluster is randomly assigned with a label. All of the odor stimuli are randomly divided into a training set (80%) and testing set (20%). The output of the coding process $\boldsymbol{r}(\boldsymbol{c}, \boldsymbol{W})$ serves as the input of the "classifier" network. On average, each hidden unit (KC) receives

7 random connection from ORNs with the weights sampled from a truncated Gaussian distribution $\mathcal{N}(0.5, 0.25)$. Input of each KC is normalized and the response of KCs is a rectified linear function of its input with a threshold chosen such that on average 10% KCs respond to each odor (see *SI Appendix* for details).

ACKNOWLEDGMENTS. We thank Xiaojing Yang, Guangwei Si, Jingxiang Shen, Louis Tao, and Roger Traub for helpful discussions and comments. The work was supported by Chinese Ministry of Science and Technology Grant 2015CB910300 and National Natural Science Foundation of China Grant 91430217. The work by Y.T. is partially supported by NIH Grant R01-GM081747.

1. R. A. Raguso, Wake up and smell the roses: The ecology and evolution of floral scent. *Annu. Rev. Ecol. Evol. Syst.* **39**, 549–569 (2008).
2. A. Dunkel *et al.*, Nature's chemical signatures in human olfaction: A foodborne perspective for future biotechnology. *Angew. Chem. Int. Ed.* **53**, 7124–7143 (2014).
3. I. Beyaert, M. Hilker, Plant odour plumes as mediators of plant–insect interactions. *Biol. Rev.* **89**, 68–81 (2014).
4. C. Y. Su, K. Menuz, J. R. Carlson, Olfactory perception: Receptors, cells, and circuits. *Cell* **139**, 45–59 (2009).
5. K. Touhara, L. B. Vosshall, Sensing odorants and pheromones with chemosensory receptors. *Annu. Rev. Physiol.* **71**, 307–332 (2009).
6. G. Glusman, I. Yanai, I. Rubin, D. Lancet, The complete human olfactory subgenome. *Genome Res.* **11**, 685–702 (2001).
7. C. Verbeurgt *et al.*, Profiling of olfactory receptor gene expression in whole human olfactory mucosa. *PLoS One* **9**, e96333 (2014).
8. C. I. Bargmann, Comparative chemosensation from receptors to ecology. *Nature* **444**, 295–301 (2006).
9. J. P. McGann, Poor human olfaction is a 19th-century myth. *Science* **356**, eaam7263 (2017).
10. C. Bushdid, M. O. Magnasco, L. B. Vosshall, A. Keller, Humans can discriminate more than 1 trillion olfactory stimuli. *Science* **343**, 1370–1372 (2014).
11. R. C. Gerkin, J. B. Castro, The number of olfactory stimuli that humans can discriminate is still unknown. *Elife* **4**, e08127 (2015).
12. T. Weiss *et al.*, Perceptual convergence of multi-component mixtures in olfaction implies an olfactory white. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19959–19964 (2012).
13. N. Y. Masse, G. C. Turner, G. S. Jefferis, Olfactory information processing in drosophila. *Curr. Biol.* **19**, R700–R713 (2009).
14. G. Sicard *et al.*, Receptor cell responses to odorants: Similarities and differences among odorants. *Brain Res.* **292**, 283–296 (1984).
15. B. Malnic, J. Hirono, T. Sato, L. B. Buck, Combinatorial receptor codes for odors. *Cell* **96**, 713–723 (1999).
16. J. Hopfield, Odor space and olfactory processing: Collective algorithms and neural implementation. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12506–12511 (1999).
17. G. Laurent, A systems perspective on early olfactory coding. *Science* **286**, 723–728 (1999).
18. V. N. Murthy, Olfactory maps in the brain. *Annu. Rev. Neurosci.* **34**, 233–258 (2011).
19. V. Grabe *et al.*, Elucidating the neuronal architecture of olfactory glomeruli in the drosophila antennal lobe. *Cell Rep.* **16**, 3401–3413 (2016).
20. O. Gschwend *et al.*, Neuronal pattern separation in the olfactory bulb improves odor discrimination learning. *Nat. Neurosci.* **18**, 1474–1482 (2015).
21. R. Gesteland, J. Lettvin, W. Pitts, Chemical transmission in the nose of the frog. *J. Physiol.* **181**, 525–559 (1965).
22. T. S. McClintock, B. W. Ache, Hyperpolarizing receptor potentials in lobster olfactory receptor cells: Implications for transduction and mixture suppression. *Chem. Senses* **14**, 637–647 (1989).
23. J. Joseph, F. A. Dunn, M. Stopfer, Spontaneous olfactory receptor neuron activity determines follower cell response properties. *J. Neurosci.* **32**, 2900–2910 (2012).
24. E. Hallem, J. R. Carlson, Coding of odors by a receptor repertoire. *Cell* **125**, 143–160 (2006).
25. S. A. Kreher, D. Mathew, J. Kim, J. R. Carlson, Translation of sensory input into behavioral output via an olfactory system. *Neuron* **59**, 110–124 (2008).
26. L. H. Cao *et al.*, Odor-evoked inhibition of olfactory sensory neurons drives olfactory perception in drosophila. *Nat. Commun.* **8**, 1357 (2017).
27. R. L. Davis, Olfactory memory formation in drosophila: From molecular to systems neuroscience. *Annu. Rev. Neurosci.* **28**, 275–302 (2005).
28. H. B. Barlow, "Possible principles underlying the transformations of sensory messages" in *Sensory Communication*, W. A. Rosenblith, Ed. (MIT Press, 2012), pp. 217–234.
29. S. Laughlin, A simple coding procedure enhances a neuron's information capacity. *Z. Naturforschung C* **36**, 910–912 (1981).
30. R. Linsker, Self-organization in a perceptual network. *Computer* **21**, 105–117 (1988).
31. J. J. Atick, Could information theory provide an ecological theory of sensory processing?*Netw. Comput. Neural Syst.* **3**, 213–251 (1992).
32. D. J. Field, Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* **4**, 2379–2394 (1987).
33. J. J. Atick, A. N. Redlich, Towards a theory of early visual processing. *Neural Comput.* **2**, 308–320 (1990).
34. Y. Dan, J. J. Atick, R. C. Reid, Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computational theory. *J. Neurosci.* **16**, 3351–3362 (1996).
35. A. J. Bell, T. J. Sejnowski, The "independent components" of natural scenes are edge filters. *Vis. Res.* **37**, 3327–3338 (1997).
36. D. L. Ruderman, W. Bialek, "Statistics of natural images: Scaling in the woods" in *Advances in Neural Information Processing Systems 6,* J. D. Cowan, G. Tesauro, J. Alspector, Eds. (MIT Press, 1994), pp. 551–558.
37. M. S. Lewicki, Efficient coding of natural sounds. *Nat. Neurosci.* **5**, 356–363 (2002).
38. N. Brenner, W. Bialek, R. d. R. Van Steveninck, Adaptive rescaling maximizes information transmission. *Neuron* **26**, 695–702 (2000).
39. O. Schwartz, E. P. Simoncelli, Natural signal statistics and sensory gain control. *Nat. Neurosci.* **4**, 819–825 (2001).
40. X. Pitkow, M. Meister, Decorrelation and efficient coding by retinal ganglion cells. *Nat. Neurosci.* **15**, 628–635 (2012).
41. R. Haddad *et al.*, A metric for odorant comparison. *Nat. Methods* **5**, 425–429 (2008).
42. A. Keller *et al.*, Predicting human olfactory perception from chemical features of odor molecules. *Science* **355**, 820–826 (2017).
43. H. K. Dweck *et al.*, The olfactory logic behind fruit odor preferences in larval and adult drosophila. *Cell Rep.* **23**, 2524–2531 (2018).
44. E. J. Candes, T. Tao, Decoding by linear programming. *IEEE Trans. Inf. Theory* **51**, 4203–4215 (2005).
45. E. J. Candes, J. K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* **59**, 1207–1223 (2006).
46. S. Ganguli, H. Sompolinsky, Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annu. Rev. Neurosci.* **35**, 485–508 (2012).
47. E. J. Candès, M. B. Wakin, An introduction to compressive sampling. *IEEE Signal Process. Mag.* **25**, 21–30 (2008).
48. G. Si *et al.*, Structured odorant response patterns across a complete olfactory receptor neuron population. *Neuron* **101**, 950–962.e7 (2019).
49. Y. Zhang, T. O. Sharpee, A robust feedforward model of the olfactory system. *PLoS Comput. Biol.* **12**, e1004850 (2016).
50. K. Krishnamurthy, A. M. Hermundstad, T. Mora, A. M. Walczak, V. Balasubramanian, Disorder and the neural representation of complex odors: Smelling in the real world. arXiv:1707.01962 (6 July 2017).
51. V. Singh, M. Tchernookov, V. Balasubramanian, What the odor is not: Estimation by elimination. arXiv:1903.02580 (6 March 2019).
52. D. Zwicker, A. Murugan, M. P. Brenner, Receptor arrays optimized for natural odor statistics. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5570–5575 (2016).
53. V. Singh, N. R. Murphy, V. Balasubramanian, J. D. Mainland, A competitive binding model predicts nonlinear responses of olfactory receptors to complex mixtures. arXiv:1805.00563 (1 May 2018).
54. G. Reddy, J. D. Zak, M. Vergassola, V. N. Murthy, Antagonism in olfactory receptor neurons and its implications for the perception of odor mixtures. *eLife* **7**, e34958 (2018).
55. T. M. Cover, J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, 2012).
56. X. X. Wei, A. A. Stocker, Mutual information, Fisher information, and efficient coding. *Neural Comput.* **28**, 305–326 (2016).
57. C. R. Sims, Efficient coding explains the universal law of generalization in human perception. *Science* **360**, 652–656 (2018).
58. G. Tkačik, W. Bialek, Information processing in living systems. *Annu. Rev. Condens. Matter Phys.* **7**, 89–117 (2016).
59. N. Hansen, A. Ostermeier, Completely derandomized self-adaptation in evolution strategies. *Evol. Comput.* **9**, 159–195 (2001).
60. N. Hansen, "The cma evolution strategy: A comparing review" in *Towards a New Evolutionary Computation: Advances on Estimation of Distribution Algorithms*, J. A. Lozano, P. Larrañaga, E. Bengoetxea, Eds. (Springer, 2006), pp. 75–102.
61. A. Celani, E. Villermaux, M. Vergassola, Odor landscapes in turbulent environments. *Phys. Rev. X* **4**, 041015 (2014).
62. H. Saito, Q. Chi, H. Zhuang, H. Matsunami, J. D. Mainland, Odor coding by a mammalian receptor repertoire. *Sci. Signal.* **2**, ra9–ra9 (2009).
63. X. Cui, C. Wu, L. Zhang, Electrophysiological response patterns of 16 olfactory neurons from the trichoid sensilla to odorant from fecal volatiles in the locust, locusta migratoria manilensis. *Arch. Insect Biochem. Physiol.* **77**, 45–57 (2011).
64. J. Perez-Orive *et al.*, Oscillations and sparsening of odor representations in the mushroom body. *Science* **297**, 359–365 (2002).
65. S. J. Caron, V. Ruta, L. Abbott, R. Axel, Random convergence of olfactory inputs in the drosophila mushroom body. *Nature* **497**, 113–117 (2013).
66. A. C. Lin, A. M. Bygrave, A. De Calignon, T. Lee, G. Miesenböck, Sparse, decorrelated odor coding in the mushroom body enhances learned odor discrimination. *Nat. Neurosci.* **17**, 559–568 (2014).
67. P. Cognigni, J. Felsenberg, S. Waddell, Do the right thing: Neural network mechanisms of memory formation, expression and update in Drosophila. *Curr. Opin. Neurobiol.* **49**, 51–58 (2018).
68. B. Babadi, H. Sompolinsky, Sparseness and expansion in sensory representations. *Neuron* **83**, 1213–1226 (2014).
69. A. Litwin-Kumar, K. D. Harris, R. Axel, H. Sompolinsky, L. Abbott, Optimal degrees of synaptic connectivity. *Neuron* **93**, 1153–1164 (2017).
70. T. Connelly, A. Savigner, M. Ma, Spontaneous and sensory-evoked activity in mouse olfactory sensory neurons with defined odorant receptors. *J. Neurophysiol.* **110**, 55–62 (2013).
71. A. F. Carey, G. Wang, C. Y. Su, L. J. Zwiebel, J. R. Carlson, Odorant reception in the malaria mosquito anopheles gambiae. *Nature* **464**, 66–71 (2010).

72. J. E. Niven, S. B. Laughlin, Energy limitation as a selective pressure on the evolution of sensory systems. *J. Exp. Biol.* **211**, 1792–1804 (2008).
73. A. Nakashima *et al.*, Agonist-independent gpcr activity regulates anterior-posterior targeting of olfactory sensory neurons. *Cell* **154**, 1314–1325 (2013).
74. T. Grebe, J. Stock, Bacterial chemotaxis: The five sensors of a bacterium. *Curr. Biol.* **8**, R154–R157 (1998).
75. T. Shimizu, Y. Tu, H. Berg, A modular gradient-sensing network for chemotaxis in escherichia coli revealed by responses to time-varying stimuli. *Mol. Syst. Biol.* **6**, 382 (2010).
76. B. Wark, B. N. Lundstrom, A. Fairhall, Sensory adaptation. *Curr. Opin. Neurobiol.* **17**, 423–429 (2007).
77. A. Egea-Weiss, A. Renner, C. J. Kleineidam, P. Szyszka, High precision of spike timing across olfactory receptor neurons allows rapid odor coding in Drosophila. *iScience* **4**, 76–83 (2018).
78. Q. Gaudry, E. J. Hong, J. Kain, B. L. de Bivort, R. I. Wilson, Asymmetric neurotransmitter release enables rapid odour lateralization in drosophila. *Nature* **493**, 424–428 (2013).
79. C. Martelli, J. R. Carlson, T. Emonet, Intensity invariant dynamics and odor-specific latencies in olfactory receptor neuron response. *J. Neurosci.* **33**, 6285–6297 (2013).
80. N. Kadakia, T. Emonet, Front-end weber-fechner gain control enhances the fidelity of combinatorial odor coding. bioRxiv p. 475103 (21 November 2018).
81. D. Rokni, V. Hemmelder, V. Kapoor, V. N. Murthy, An olfactory cocktail party: Figure-ground segregation of odorants in rodents. *Nat. Neurosci.* **17**, 1225–1232 (2014).
82. R. A. Campbell *et al.*, Imaging a population code for odor identity in the drosophila mushroom body. *J. Neurosci.* **33**, 10568–10581 (2013).
83. Y. Weiss, H. S. Chang, W. T. Freeman, Learning compressed sensing. *CiteSeerX* (2007). http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.447.6824&rep=rep1&type=pdf. Accessed 10 September 2019.
84. D. L. Donoho, A. Javanmard, A. Montanari, Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing. *IEEE Trans. Inf. Theory* **59**, 7434–7464 (2013).
85. A. Ashok, L. C. Huang, M. A. Neifeld, Information optimal compressive sensing: Static measurement design. *JOSA A* **30**, 831–853 (2013).
86. S. Mansourian, M. C. Stensmyr, The chemical ecology of the fly. *Curr. Opin. Neurobiol.* **34**, 95–102 (2015).
87. Y. Zhou, B. H. Smith, T. O. Sharpee, Hyperbolic geometry of the olfactory space. *Sci. Adv.* **4**, eaaq1458 (2018).
88. T. Teşileanu, S. Cocco, R. Monasson, V. Balasubramanian, Adaptation of olfactory receptor abundances for efficient coding. *eLife* **8**, e39279 (2019).
89. R. A. Ince *et al.*, A statistical framework for neuroimaging data analysis based on mutual information estimated via a Gaussian copula. *Hum. Brain Mapp.* **38**, 1541–1573 (2017).
90. S. Singh, B. Pøczos, Nonparanormal information estimation. arXiv:1702.07803 (24 February 2017).
91. Z. Szabó, Information theoretical estimators toolbox. *J. Mach. Learn. Res.* **15**, 283–287 (2014).