LETTER **Communicated by Misha Tsodyks**

Learning Classification in the Olfactory System of Insects

Ramón Huerta
rhuerta@ucsd.edu
Institute for Nonlinear Science, University of California San Diego, La Jolla CA 92093-0402, U.S.A., and GNB. Escuela Técnica Superior de Informática, Universidad Autónoma de Madrid, 28049 Madrid, Spain

Thomas Nowotny
tnowotny@ucsd.edu
Institute for Nonlinear Science, University of California San Diego, La Jolla CA 92093-0402, U.S.A.

Marta García-Sanchez
m.garcia@uam.es
GNB. Escuela Técnica Superior de Informática, Universidad Autónoma de Madrid, 28049 Madrid, Spain, and Institute for Nonlinear Science, University of California San Diego, La Jolla CA 92093-0402, U.S.A.

H. D. I. Abarbanel
hdia@jacobi.ucsd.edu
Institute for Nonlinear Science, University of California San Diego, La Jolla CA 92093-0402, U.S.A., and Department of Physics and Marine Physical Laboratory, Scripps Institute of Oceanography, University of California San Diego, La Jolla, CA 92093-0402, U.S.A.

M. I. Rabinovich
mrabinovich@ucsd.edu
Institute for Nonlinear Science, University of California San Diego, La Jolla CA 92093-0402, U.S.A.

We propose a theoretical framework for odor classification in the olfactory system of insects. The classification task is accomplished in two steps. The first is a transformation from the antennal lobe to the intrinsic Kenyon cells in the mushroom body. This transformation into a higher-dimensional space is an injective function and can be implemented without any type of learning at the synaptic connections. In the second step, the encoded odors in the intrinsic Kenyon cells are linearly classified in the mushroom body lobes. The neurons that perform this linear classification are equivalent to hyperplanes whose connections are tuned by local Hebbian learning and by competition due to mutual inhibition. We

calculate the range of values of activity and size of the network required to achieve efficient classification within this scheme in insect olfaction. We are able to demonstrate that biologically plausible control mechanisms can accomplish efficient classification of odors.

1 Introduction

Odor classification by nervous systems involves several quite different computational tasks: (1) similar odor receptions need to be classified as the same odor, (2) distinct odors have to be discriminated from each other, and (3) some quite different odors might carry the same meaning and therefore must be associated with each other. From studies of insect olfaction, we have developed a framework that may provide insights into how the olfactory system of insects accomplishes these different tasks with the available natural technology.

Insects have three known processing stages of odor information before classification: the antenna, the antennal lobe (AL), and the mushroom body (MB) (see Figure 1 for a description). To put our work into the right context, we will describe the known facts about the system. Each olfactory receptor cell in the antenna expresses one type of receptor, and all olfactory receptor cells expressing the same receptor type connect to the same glomerulus in the AL (Gao, Yuan, & Chess, 2000; Vosshall, Wong, & Axel, 2000; Scott et al., 2001). Thus, a chemosensory map of receptor activity in the antenna is formed in the AL: the genetically encoded architecture induces a stimulus-dependent spatial code in the glomeruli (Rodrigues, 1988; Distler, Bausenwein, & Boeckh, 1998; Joerges, Kuettner, Galizia, & Menzel, 1997; Galizia, Joerges, Kuettner, Faber, & Menzel, 1997; Galizia, Nagler, Holldobler, & Menzel, 1998). Moreover, the spatial code is conserved across individuals of the same species (Galizia, Sachse, Rappert, & Menzel, 1999; Wang, Wong, Flores, Vosshal, & Axel, 2003) as would be expected given the genetic origin of the code. There are many fewer neurons in the first relay station (the AL) than the number of receptors types by a ratio of 1:10. The reasons for such a strong convergence are not well understood and remain an interesting theoretical question for future study.

It is known that increasing odor concentrations recruit increasing numbers of glomeruli (Ng et al., 2002; Wang et al., 2003). A simple transduction of the glomerular activity would increase the number of active neurons in the MB with increasing odor concentration as well. Calcium recordings in the MB of *Drosophila* show, however, that the activity in the MB indicated by calcium concentrations is independent of the odor concentration for the odors ethyl acetate and benzaldehyde (Wang et al., 2001). Moreover, recent recordings from the AL in the locust indicate that the activity in the projections of the excitatory neurons in the AL into the MB is nearly constant in time (Stopfer, Jayaraman, & Laurent, 2003). Therefore, a gain control mechanism maintaining a nearly constant average neuronal activity in the AL

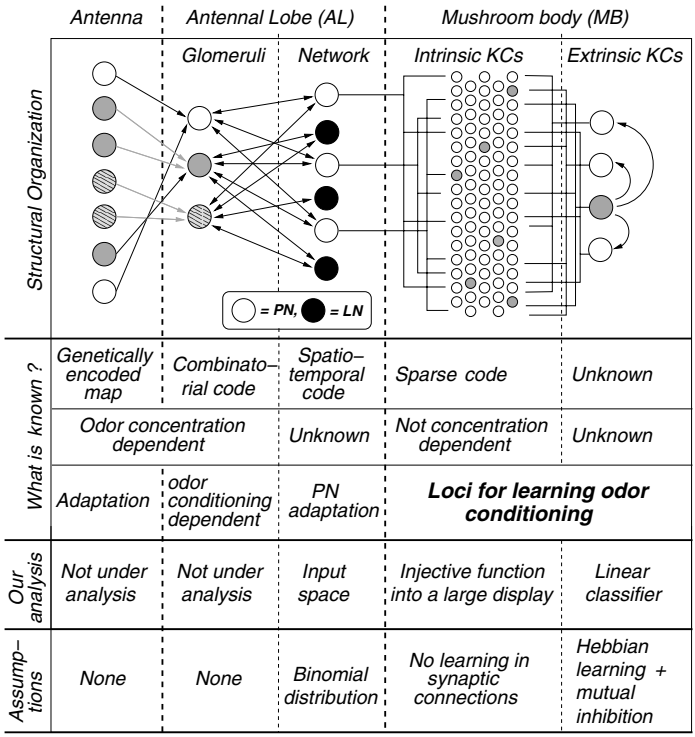

Figure 1: Description of the structural organization of the first few processing layers of the olfactory system of insects, including a list of what is known in terms of coding, odor concentration dependence, and learning of odor conditioning.

must exist. It seems that the AL performs some preprocessing of the data to feed an adequate representation of it into the area of the insect brain that is responsible for learning odor conditioning, the MB.

Although the MBs are not critical for normal behavior of insects, they are crucial for odor conditioning (Heisenberg, Borst, Wagner, & Byers, 1985). Genetic manipulation has proved to be a powerful tool to identify the MB as the locus for learning (de Belle & Heisenberg, 1994; Zars, Fischer, Schulz, & Heisenberg, 2000; Pascual & Preat, 2001; Dubnau, Grady, Kitamoto, & Tully 2001; McGuire, Le, & Davis, 2001; Connolly et al., 1996; Menzel, 2001; Heisenberg, 2003; Dubnau, Chiang, & Tully, 2003). Dubnau and colleagues (2001) propose that "the Hebbian processes underlying olfactory associative learning reside in mushroom body dendrites or upstream of the mushroom body and that the resulting alterations in synaptic strength modulate mushroom body output during memory retrieval." This is an important idea that supports our hypothesis that the important synaptic changes to achieve

classification of odors have to occur in the connections from the intrinsic KCs to the MB lobes, not earlier or later.

We propose a theory of odor classification that relies on a large coding display in the MB, random connectivity among neurons, Hebbian learning, and mutual inhibition. We show that these elements of an odor identification network are sufficient to accomplish the classification tasks described above (see Figure 1).

Our main hypothesis as stated in Figure 1 is that the classification decision occurs in the MB lobes that receive inputs from the intrinsic KCs. This hypothesis is suggested by the concept of support vector machines (SVM) (Cortes & Vapnik, 1995) and Cover's work (Cover, 1965), which were developed in the context of the classification problem based on linear threshold devices when the input classes are not linearly separable. The strategy consists of casting the input into a high-dimensional space by means of a nonlinear transformation and then using linear threshold devices for classification in this higher-dimensional representation. The nonlinear transformation is designed to separate the previously inseparable input classes in the new high-dimensional space. In this article, we consider the AL as the input space, the intrinsic KCs form the high-dimensional space, and the neurons in the MB lobes are the linear threshold devices that can separate the classes. The experimental support for having such an SVM is based on the known large divergence of connections from the AL to the MB and the observed role of the MB as the focus of learning odor conditioning.

Let us elaborate on the necessary processing layers of the classification circuit and our hypotheses about these different stages of information processing. We consider the AL as the input layer to the classification layers in the MB. Figure 2 shows the proposed classification scheme. The first processing stage is the nonlinear transformation of the information contained in the AL activity into the large screen of intrinsic KCs. Our hypothesis for this stage is that the mapping of activity patterns resulting from the transformation from the AL to the MB should be statistically injective (Garcia-Sanchez & Huerta, 2003). Injectivity means that if two different states of the AL, arising from distinct odors, are projected into the screen of intrinsic KCs, the resulting states in the intrinsic KC layer should be distinct as well.

The next hypothesis is that once the odor representations in the AL have been widely separated in the intrinsic KC layer, they can be classified at the next processing stage using simple linear threshold devices (see Figure 2); we interpret the neurons of the MB lobes as such linear threshold devices. At this stage, we then employ biologically feasible learning mechanisms to allow memory formation and self-organization in this part of the network.

In addition to showing that the developed framework indeed allows successful classification, we also determine the range of possible values for the sparseness of activity in the intrinsic KC layer. The results are consistent with experimental recordings of intrinsic KCs in the MB of locust (Perez-Orive et al., 2002) and thus give a theoretical explanation for the observed

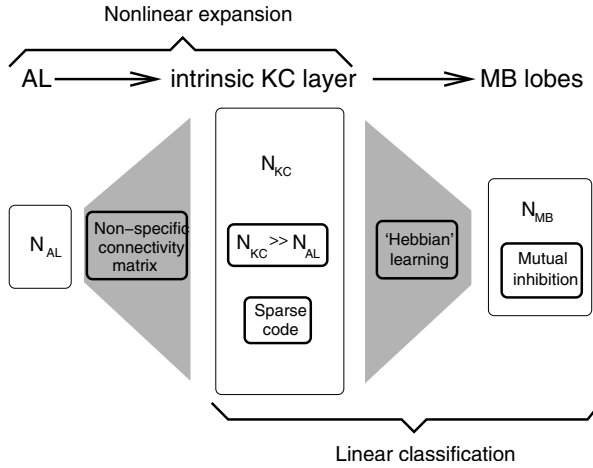

Figure 2: Main elements required to account for efficient classification in the insect olfactory system. The first stage is an injective function from the AL to the intrinsic KC layer using nonspecific, connectivity matrices (left). By nonspecific, we mean that the connectivity matrix does not need to be learned or be genetically encoded in detail. For classification purposes, a large number of neurons in the intrinsic KC layer and a sparse code are needed, as we will show. The second phase (linear classification) is the decision phase, where linear classification takes place. It is characterized by converging connections, Hebbian learning, and mutual inhibition between MB lobe neurons.

sparseness of the code. Finally, we demonstrate that the large size of the display or screen in the intrinsic KC layer is critical for efficient classification and quantify the interdependence of storage capacity and KC layer size.

Previous theoretical studies have revealed the possibility of using the temporal dynamics of the first relay station of olfactory processing for odor recognition (Hendin, Horn, & Hopfield, 1994; Hendin, Horn, & Tsodyks, 1998; White, Dickinson, Walt, & Kauer, 1998; Li & Hertz, 2000). White et al., (1998) propose delay lines to improve discrimination between odors. Hendin et al. (1994, 1998) show how the olfactory bulb (equivalent to the AL) can, in principle, implement an associative memory and segmentation of odors. Li and Hertz (2000) emphasize the importance of the feedback from the cortex (equivalent to the MB, the location of odor recognition) to the olfactory bulb (equivalent to the AL) for odor segmentation. In this work, we do not consider time or odor segmentation. Time is important to enhance the separation of similar odors (Friedrich & Laurent, 2001; Linster & Cleland, 2001; Laurent, 2002; Laurent et al., 2001). The problem of segmentation addresses the question of recognizing odor A or odor B when

encountering a mixture of both. Both the role of time and the problem of odor segmentation require detailed knowledge of the AL representation of odors and mixtures of odors. To date, there is no theoretical work that addresses these issues, especially of the code for mixtures of odors in the AL, in a satisfactory way. We therefore restrict our analysis to the spatial aspects of the representation of pure odors in an appropriate sense.

The model that we present in this article is to our knowledge the first to address quantitatively the levels of neural activity, convergent and divergent connectivities, and the function of th MB in odor classification. The role of time is left out for further investigation. We need to understand the limitations of spatial code processing before exploring the advantages of using time in the neural code.

The organization of the letter is as follows. We start with a description of the elements of the system. Then we present two analyses: an analytical description of a single decision neuron and a systematic numerical analysis demonstrating that the AL and MB structure together with Hebbian learning and mutual inhibition account for efficient classification.

2 Description of the Processing Stages

As mentioned above and shown in Figure 2, there are two essential stages in our model of odor classification: a nonlinear transformation followed by a linear classification. We will see that the similarities of this core concept of SVMs to the observed biological system are striking even though the specific implementation is different because of the actual composition, the available "wetware," of the insect olfactory network.

2.1 Nonlinear Transformation from the AL to the MB. Recently, strong evidence has been collected in the locust (Perez-Orive et al., 2002) that the spatiotemporal activity in the AL is read by the MB as pictures or snapshots. The two observations supporting this hypothesis are the strong feedforward inhibition from the lateral horn interneurons onto the Kenyon cells and the short integration time of the intrinsic KCs. The periodic strong inhibition resets the activity in the KC layer every 50 ms, whereas the short integration time of the KCs makes them coincidence detectors that cannot process input across more than one inhibition cycle.

Because we are not addressing the temporal aspects of the system in this work, the input for our classification system is a single snapshot. The hypothesis for the nonlinear transformation from the AL to the MB is then that every such snapshot or code word in the AL has a unique corresponding code word in the MB: the nonlinear transformation needs to be an injective function at least in a statistical sense. In previous work (Garcia-Sanchez & Huerta, 2003), we proposed a method to select the parameter values that allow one to construct such an injective function from the AL to the intrinsic KC layer with very high probability. The appropriate parameters are used

throughout this article for designing the nonlinear transformation from the AL to the MB.

It is known that the activity among intrinsic KCs is very sparse (Perez-Orive et al., 2002). Most of the intrinsic KCs fire just one or two spikes during one odor presentation of a few seconds. Given this low endogenous activity of the intrinsic KC neurons, we chose simple McCulloch-Pitts "neurons" (McCulloch & Pitts, 1943) to represent them. The state of this model neuron is described by a binary number (0 = no spike and 1 = spike). In particular, the McCulloch-Pitts neuron is described by

$$y_j = \Theta\left(\sum_{i=1}^{N_{\mathrm{AL}}} c_{ji}x_i - \theta_{\mathrm{KC}}\right) \quad j = 1, 2, \ldots, N_{\mathrm{KC}}. \tag{2.1}$$

$\mathbf{x}$ is the state vector of the AL neurons. It has dimension N_{AL}, where N_{AL} is the number of AL neurons. The components of the vector $\mathbf{x} = [x_1, x_2, \ldots, x_{N_{\mathrm{AL}}}]$ are 0s and 1s. $\mathbf{y}$ is the state vector for the intrinsic KC layer; it is N_{KC} dimensional. The c_{ij} are the components of the connectivity matrix, which is $N_{\mathrm{AL}} \times N_{\mathrm{KC}}$ in size; its elements are also 0s and 1s. θ_{KC} is an integer number that gives the firing threshold in an intrinsic KC. The Heaviside function $\Theta(\cdot)$ is unity when its argument is positive and zero when its argument is negative.

To determine the statistical degree of injectivity of the connectivity between the AL and intrinsic KC, we first calculate the probability of having identical outputs given different inputs for a given connectivity matrix: $P(\mathbf{y} = \mathbf{y}' \mid \mathbf{x} \neq \mathbf{x}', C)$, where C is one of the possible connectivity matrices (see Garcia-Sanchez & Huerta, 2003, for details) and the notation $\mathbf{x} \neq \mathbf{x}'$ is $\{(\mathbf{x}, \mathbf{x}') : \mathbf{x} \neq \mathbf{x}'\}$. We want this probability, which we call the probability of confusion, to be as small as possible, on average, over all inputs and over all connectivity matrices.

We write this average as $P(\text{confusion}) = \left\langle \langle P(\mathbf{y} = \mathbf{y}' \mid \mathbf{x} \neq \mathbf{x}', C) \rangle_{\mathbf{x} \neq \mathbf{x}'} \right\rangle_C$, where $\langle \cdot \rangle_{\mathbf{x} \neq \mathbf{x}'}$ is the average over all nonidentical input pairs $(\mathbf{x}, \mathbf{x}')$, and $\langle \cdot \rangle_C$ is the average over all connectivity matrices C. This gives us a measure of clarity, the opposite of confusion, as

$$I = 1 - P(\text{confusion}). \tag{2.2}$$

The closer I is to 1, the better is our statistically injective transformation from the states $\mathbf{x}$ of the AL to the states $\mathbf{y}$ of the intrinsic KCs.

There are two parameters of the model that can be adjusted using the measure of clarity. One is the probability p_C of having a connection between a given neuron in the AL and a given intrinsic KC. The second is the firing threshold θ_{KC} of the intrinsic KCs. Fixed parameters in the model are the probability p_{AL} of having an active neuron in the AL layer, the number N_{AL} of input neurons, and the number N_{KC} of intrinsic KCs. p_C and θ_{KC} can be

estimated using the following inequality,

$$I \leq 1 - \{{p_{\mathrm{KC}}}^2 + (1 - p_{\mathrm{KC}})^2 + 2\sigma^2\}^{N_{\mathrm{KC}}}, \tag{2.3}$$

where p_{KC} is the firing probability of a single neuron in the intrinsic KC layer. It can be calculated for inputs and connection matrices generated by a Bernoulli process with probabilities p_{AL} and p_C as

$$p_{KC} = \sum_{i=\theta_{KC}}^{N_{AL}} \binom{N_{AL}}{i} (p_{AL}p_C)^i (1 - p_{AL}p_C)^{N_{AL}-i}. \tag{2.4}$$

This probability has variance (σ^2 above) when we average over all possible inputs and connectivity matrices.

The formula for the probability of confusion can be intuitively understood if we assume that the activity of every intrinsic KC is statistically independent of the activity of the others. If so, the probability of confusion in one output neuron is the sum of the probability of having a one for two inputs plus the probability of having a zero for both: $p_{\mathrm{KC}}^2 + (1-p_{\mathrm{KC}})^2$. Thus, the probability of confusion in all N_{KC} output neurons is $(p_{\mathrm{KC}}^2 + (1 - p_{\mathrm{KC}})^2)^{N_{\mathrm{KC}}}$ in the approximation of independent inputs. This bound on I should be close to unity for any set of parameter values we choose. The inequality for the measure of clarity becomes an equality for sparse connectivity matrices. This, fortunately, is the case in the locust, where every intrinsic KC receives an average of only 10 to 20 connections from the AL where nearly 900 are possible.

2.2 Linear Classification in the MB Lobes. As shown in Figure 2, we hypothesize that the classification decision takes place at the MB lobes. The neurons in the MB lobes are again modeled by McCulloch-Pitts neurons, which are simple linear threshold devices,

$$z_l = \Theta\left(\sum_{j=1}^{N_{\mathrm{KC}}} w_{lj} \cdot y_j - \theta_{\mathrm{LB}}\right), \quad l = 1, 2, \ldots, N_{\mathrm{LB}}. \tag{2.5}$$

Here, the index *LB* denotes the MB *Lob*es. The vector $\mathbf{z}$ is the state of the MB lobes; it has dimension N_{LB}. θ_{LB} is the threshold for the decision neurons in the MB lobes. The $N_{\mathrm{KC}} \times N_{\mathrm{LB}}$ connectivity matrix w_{lj} has entries 0 or 1. We introduce a plasticity rule for the entries w_{lj} below.

Every column vector $\mathbf{w}_l$ (w_{lj} for fixed j) of a connectivity matrix defines a hyperplane in the intrinsic KC layer coordinates $\mathbf{y}$ (see Figure 3 for an illustration of this). This is the plane that is normal to $\mathbf{w}_l$. This hyperplane discriminates between the points that are "above it" and the points that are "below it." There is a different hyperplane for each MB lobe neuron, and the combinatorial placement of the planes determines the classification space.

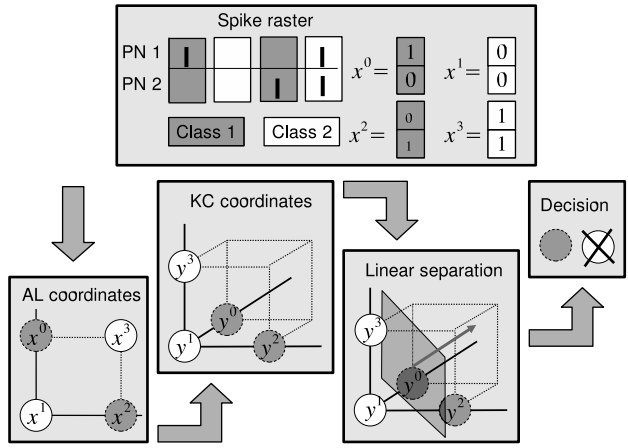

Figure 3: Hypothesized functional diagram of the olfactory system of insects. The data of the AL are encoded in snapshots. Every snapshot is a binary code of spikes (1) or no spikes (0). The projection from the AL to the intrinsic KC layer separates the encoded odors to be classified. The connectivity matrix from the intrinsic KC layer to the MB lobes defines a set of hyperplanes $\mathbf{w}$ that are able to discriminate between different sets of encoded odors. Note that without the projection into a higher-dimensional space, it would not be possible to linearly discriminate the dark points from the light ones.

2.2.1 Local Synaptic Changes: Hebbian Learning. Another key ingredient in addressing the classification problem is a learning mechanism. Hebbian learning is the classical choice for local synaptic changes (Hebb, 1949). The resulting local synaptic modifications are made at the linear classification stage (see Figure 2) and efficiently solve the classification problem. Hebbian plasticity is used to adjust the connectivity matrix w_{lj} from the intrinsic KC layer to the MB lobes. No other areas need to be involved in learning for our model of classification.

The plasticity rule is applied by first choosing a connectivity matrix with some randomly chosen initial entries. Then inputs are presented to the system. The entries of the connectivity matrix at the time of the nth input are denoted by $w_{lj}(n)$. The values after the next input, $w_{lj}(n+1)$, are given by the rule

$$w_{lj}(n+1) = H\left(z_l, y_j, w_{lj}(n)\right), \tag{2.6}$$

where

$$H(1,1,w_{lj}(n)) = \begin{cases} 1 & \text{with probability } p_+, \\ w_{lj}(n) & \text{with probability } 1-p_+, \end{cases}$$

$$H(1,0,w_{lj}(n)) = \begin{cases} 0 & \text{with probability } p_-, \\ w_{lj}(n) & \text{with probability } 1-p_-, \end{cases} \tag{2.7}$$
$$H(0,1,w_{lj}(n)) = w_{lj}(n),$$
$$H(0,0,w_{lj}(n)) = w_{lj}(n).$$

A synaptic connection is activated (it becomes 1 when it was 0) with probability p_+ if the input activity is accompanied by an output activity. The connection is removed (it becomes 0 if it was 1) with probability p_- if an output occurs in the absence of input activity. In the remaining cases, no input and no output and input but no output, the synapse remains unaltered.

For example, let us apply the plasticity rule to a given input $\mathbf{y}$ and one output $z = 1$. This is the case that we primarily analyze in the next section. It can be easily calculated that the average number of iterations it takes for the connection w_j to become 1 if $y_j = 1$ and $w_j = 0$ is $1/p_+$. If w_j was already 1, then the number of iterations is 0. On the other hand, the average number of iterations it takes for the connection w_j to become 0 if $y_j = 0$ and $w_j = 1$ is $1/p_-$. In other words, the inverses of the probabilities p_+ and p_- are just the timescales of synaptic changes. Therefore, if we apply this rule sufficiently long for one given input activity at the intrinsic KC layer and an active response $z = 1$ in an output neuron, the set of active connections will eventually be the set of active inputs itself.

2.2.2 Mutual Inhibition: n_W-Winner-Take-All. We hypothesize that mutual inhibition exists and, in joint action with Hebbian learning, is able to organize a nonoverlapping response of the decision neurons in the MB lobes. This is often considered a neural mechanism for self-organization in the brain. The combination of Hebbian learning and mutual inhibition has already been proposed as a biologically feasible mechanism to account for learning in neural networks (O'Reilly, 2001).

Mutual inhibition is implemented artificially in the finite automata model with McCulloch-Pitts neurons. We allow only a subset of decision neurons that receive the highest synaptic input to fire. The size of this subset is fixed to allow n_W winners in a winner-take-all configuration for every odor presentation. Let us define the vector $u_\mu = \sum_{j=1}^{N_{\mathrm{KC}}} w_{\mu j}\, y_j - \theta_{\mathrm{LB}}$. Then equation 2.5 is rewritten as

$$z_l = \Theta\left(\sum_{j=1}^{N_{\mathrm{KC}}} w_{lj}\, y_j - \theta_{\mathrm{LB}} - \{\mathbf{u}\}_{n_W+1}\right), \tag{2.8}$$

where $\{\mathbf{u}\}_{n_W+1}$ denotes the $(n_W + 1)$th largest component of the vector $\mathbf{u}$. Since we force the system to have exactly n_W active extrinsic KCs, θ_{LB} does not play any role because it is canceled in equation 2.8 and we can simply

Table 1: Summary of the Parameters and Variables of All the Processing Layers.

	AL	AL→KC	KC	KC→ lobe	M lobe
Variables	$\mathbf{x}$	c_{ij}	$\mathbf{y}$	w_{ij}	$\mathbf{z}$
Parameters	$N_{\mathrm{AL}}, p_{\mathrm{AL}}$	p_C	$N_{\mathrm{KC}}, \theta_{\mathrm{KC}}$ $p_{\mathrm{KC}}(N_{\mathrm{AL}}, p_{\mathrm{AL}}, p_C, \theta_{\mathrm{KC}})$	p_+, p_-	N_{LB}, n_W

write

$$z_l = \Theta\left(\sum_{j=1}^{N_{\mathrm{KC}}} w_{lj}\, y_j - \left\{\sum_{j=1}^{N_{\mathrm{KC}}} w_{\mu j}\, y_j\right\}_{n_W+1}\right). \tag{2.9}$$

The parameters and variables of the model are summarized in Table 1. The AL parameters are the number of neurons N_{AL} and the probability of firing p_{AL}. The connectivity probability from the AL to the MB is p_C. The parameters of the intrinsic KC layer are the number of neurons N_{KC} and the threshold for activation θ_{KC}. The firing probability p_{KC} is a function of N_{AL}, p_{AL}, p_C, and θ_{KC}. The dynamics of the connections from the intrinsic KC layer to the MB lobes is governed by the probabilities (timescales) for increasing and decreasing connection strength, p_+ and p_-. Finally, the response of the decision layer depends on the number of allowed active neurons in the MB lobes, n_W. For clarity, we also include a summary of the equations of the model in Table 2.

3 Results

In the following sections, we divide our analysis into two parts. First, we analyze the parameter values that allow classification with one single output neuron. For this analysis, we use basic probability theory in

Table 2: Summary of the Equations.

Intrinsic KC layer	$y_j = \Theta\left(\sum_{i=1}^{N_{\mathrm{AL}}} c_{ji}\, x_i - \theta_{\mathrm{KC}}\right)$
Mushroom body lobes	$z_l = \Theta\left(\sum_{j=1}^{N_{\mathrm{KC}}} w_{lj}\, y_j - \left\{\sum_{j=1}^{N_{\mathrm{KC}}} w_{\mu j}\, y_j\right\}_{n_W+1}\right)$,
Plasticity rule	$w_{lj}(n+1) = \begin{cases} 1 & \text{with } p_+ \text{ if } y_j = 1 \text{ and } z_l = 1 \\ 0 & \text{with } p_- \text{ if } y_j = 0 \text{ and } z_l = 1 \\ w_{lj}(n) & \text{in all other cases} \end{cases}$

Note: The index n represents the iteration number for presentations of a snapshot of AL activity to the intrinsic KCs.

a similar framework as Caticha, Palo Tejada, Lancet, & Domany, (2002). In principle, Shannon information can be employed as well, as in Nadal (1991) and Frolov & Murav'ev (1993) for associative memories. For our purposes, basic probability theory yields straightforward results that are easy to interpret.

For simplicity, we will use a simple convention to separate the classification problem into two subproblems we call discrimination and association:

- The discrimination task is the ability of a single output neuron to distinguish one input from the rest, assuming uncorrelated input patterns in the antennal lobe. The probability space of AL activity will be unchanged throughout our analysis
- The association problem is defined as the capability of a single output neuron to fire for d given uncorrelated inputs while not responding to the rest. The discrimination problem can be seen as a special case of the association problem with $d = 1$.

In the second part of the results, we present simulations of a complete system with many output neurons that organizes via the plasticity rule into winner-take-all configurations. This analysis is different from the previous one in that the system self-organizes without supervision.

3.1 Discrimination for a Single Output Neuron. We consider a single output neuron in the MB lobes, usually referred to as an extrinsic Kenyon cell (eKC). We will show how the ability of this neuron to discriminate one odor from the rest depends on the total number of neurons in the intrinsic KC layer and the level of KC activity. The obtained discrimination ability can be enhanced by redundancy, for example, by using several decision neurons. The single neuron calculations presented here are in this sense a lower bound on the ability of the system to perform discrimination. The application of error correcting codes (Hamming, 1950) based on redundancy will be investigated in future work.

The goal of the investigation presented in this section is to obtain exact quantitative statements rather than general trends to be able to compare our results to experimental findings. Let us proceed with the analysis as illustrated in Figure 4. As a first step, we need to calculate the probability distribution of the number of active intrinsic KC neurons (iKCs). The expectation value for the number of active iKCs is given by $\mathbb{E}\, n_{KC} = N_{KC}\, p_{KC}$, where N_{KC} is the number of iKC and p_{KC} is given by equation 2.4, (see section A.1 in the appendix). The probability distribution for the number of active iKC is, however, not binomial despite the random connectivity matrix between the AL and the MB. It can be calculated using the same procedure as in Nowotny and Huerta (2003). The details are explained in the appendix. The probability distribution $P(n_{KC} = k)$ cannot be simplified into a more compact form. However, it can be calculated for given p_C

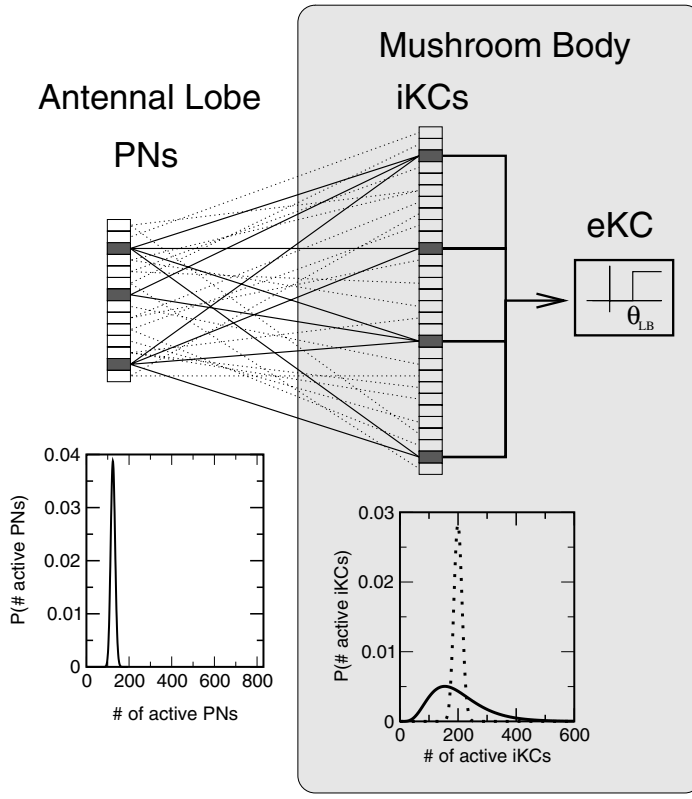

Figure 4: Characteristics of the first processing stage, the nonlinear expansion. The input probability space in the AL is fixed as a Bernoulli process with $p_{AL} = 0.15$ in 50 millisecond intervals, as observed experimentally. The probability distribution, which is a binomial distribution, is shown in the lower left graph. The intrinsic KCs receive direct input from the PNs according to equation 2.1. The main parameter, here, is the probability of connection from the AL to the MB, p_C. By changing this parameter, we can regulate the expected activity in the intrinsic KC layer. The resulting probability distribution for the number of active iKCs is not a binomial distribution (dotted line in the lower right graph) but a much wider distribution (solid line in the lower right graph) with an expected value given by equation 2.4. Given this distribution regulated by p_C, we study the ability to discriminate uncorrelated inputs in the AL by using a single extrinsic KC. This investigation gives us a lower limit on the performance of this system that can be quantitatively compared to experimental data.

and p_{AL} and stored in the computer for further calculations. As shown in Figure 4, the distribution $P(n_{KC} = k)$ is three to four times wider than the binomial distribution obtained if assuming independence of firing events in the iKCs. The distributions in the figure were obtained for the typical parameter values of the locust.

To quantitatively determine the ability of the system to discriminate, we then generate a set of random inputs $\mathbf{x}^0, \mathbf{x}^1, \ldots, \mathbf{x}^N$ in the AL by independent identical Bernoulli processes with probability p_{AL}. In section 2.1, we give the conditions such that the map from the AL activity to the MB activity is an injective function. If these conditions are fulfilled, the set of random inputs in the AL corresponds (with probability close to one) uniquely to a set $\mathbf{y}^0, \mathbf{y}^1, \ldots, \mathbf{y}^N$ of activity vectors in the MB. The activities $n_{KC}(\mathbf{y}^j)$ of the activity vectors obey the probability distribution given in equation A.4.

The first pattern $\mathbf{y}^0$ is then trained in one single classification output neuron: one neuron is forced to spike in response to this pattern, while for all other input patterns, the system acts autonomously. The Hebbian plasticity rule given in equation 2.8 generates a learned connectivity vector $\mathbf{w} = \mathbf{y}^0$ if the rule is applied in the order of $\max\{1/p_-, 1/p_+\}$ times. Given this learned vector $\mathbf{w}$, we determine the structural parameters such that $z(\mathbf{y}^1) = z(\mathbf{y}^2) = \cdots = z(\mathbf{y}^N) = 0$ with probability close to one. The probability of discrimination $P(z(\mathbf{y}^1) = 0, z(\mathbf{y}^2) = 0, \ldots, z(\mathbf{y}^N) = 0)$ is calculated in the appendix in section A.2. For successful discrimination, it needs to be as close to one as possible. To obtain the probability of discrimination, we basically estimate the degree of overlap of each of the uncorrelated inputs to the learned one. If the learned vector has $l(\mathbf{w})$ ones, the probability of having i overlapping 1s between $\mathbf{w}$ and any given $\mathbf{y}^j$ is

$$p(i \text{ overlapping ones}) = \frac{\binom{l(\mathbf{w})}{i}\binom{N_{KC}-l(\mathbf{w})}{l(\mathbf{y}^j)-i}}{\binom{N_{KC}}{l(\mathbf{y}^j)}}. \tag{3.1}$$

The overlapping activity vectors are read by an eKC that has a threshold of activation, θ_{LB}. In order for the eKC not to spike in response to the wrong activity vector $\mathbf{y}^j$, the probability for overlaps with $i \geq \theta_{LB}$ has to be close to 0. It can quickly be seen that for sparse activity, the probability of overlaps decreases. The advantage of using sparse code for associative memory was pointed out by Marr (1969) and others (Willshaw, Buneman, & Longuet-Higgins, 1969) 30 years ago using nonoverlapping patterns of activity. In this work, we rigorously calculate the probability of discrimination in order to determine the minimum size of the intrinsic KC layer, the maximal and minimal activity in the intrinsic KC layer, the capacity of the system, and the ability to discriminate similar odors. This will allow us to compare our results to the real system on a quantitative level.

3.1.1 Dependence on the Intrinsic KC Layer Size . It is interesting to investigate the size of the intrinsic KC layer because of the contradiction that nature seems to have chosen in insect olfaction. On the one hand, the number of 50, 000 intrinsic KCs in locust is very large. On the other hand, the locust seems to use only a few of these cells. That raises the question: Is there an optimal size of the intrinsic KC layer that allows the system to work most efficiently?

There are two main cases: keep p_{KC} constant, leading to an increasing number of active neurons $n_{KC} = p_{KC} \cdot N_{KC}$, with increasing size N_{KC}, or keep the number of active neurons n_{KC} constant by adjusting p_{KC} appropriately. Using constant p_{KC} is problematic because the learned vector will have too many active connections due to the Hebbian plasticity rule when N_{KC} is large. Since the threshold θ_{LB} of the decision neuron is not size dependent, this causes the neuron to fire for any input, and the network's discrimination ability drops drastically for large sizes. We therefore concentrate on the more interesting second case ensuring a constant average number of active neurons by using equation 2.4 to modify p_C appropriately.

In Figure 5A, the probability of discrimination for a total of $N = 100$ inputs, for various values of the (fixed) average number n_{KC} of active intrinsic KCs as a function of the intrinsic KC layer size N_{KC}, is shown. For better discrimination, one obviously needs to increase the number of neurons in the intrinsic KC layer. Note that although the scale shown is logarithmic, one can see a relatively small area in which the system makes a transition from failure to function. This points to the possibility that there is a critical number of intrinsic KCs that can be quantitatively compared to experimental data.

The most important result is that the minimum iKC layer size for successful discrimination has scaling property as a function of the average number of active iKCs. The results are shown in Figure 5B for three different values of N. As one can see in the log-log plot, the exponent does not depend on N. When we fit these values, $N_{KC} \propto n_{KC}^{2.12}$. We can theoretically estimate that $N_{KC} \propto n_{KC}^2$ (this will be published elsewhere). However, we do not know at this point where the discrepancy between the exponents stems from because we took severe limiting cases, like $p_{AL} p_C \to 0$ and $N_{AL} \gg \theta_{KC}$, which has an effect on the scaling exponent. The implication of this result is that for discrimination purposes, if the internal representation in the intrinsic KC layer is large, the size of this layer needs to be even larger. The effects of noise can be strong, because it would require increasing the number of active neurons, as explained in Garcia-Sanchez & Huerta (2003). We can conjecture that the intrinsic KC neurons need to have very low endogenous noise.

3.1.2 What Is the Maximum Number of Active iKC That Allows the System to Discriminate? It has been experimentally observed that the representation

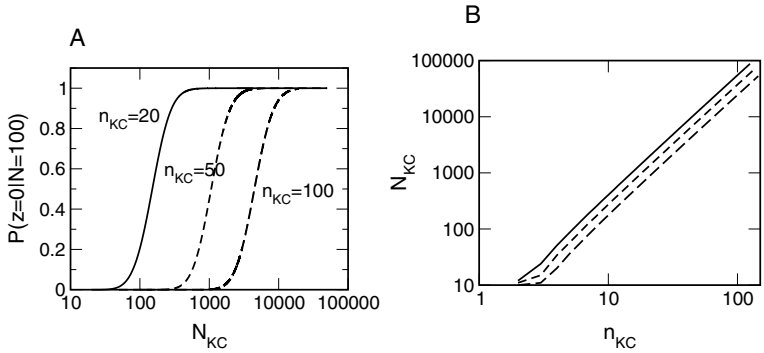

Figure 5: (A) Probability of discrimination for $N = 100$ inputs for several numbers of active neurons n_{KC} in the intrinsic KC layer for a threshold value of activation in the MB lobe layer of $\theta_{\mathrm{LB}} = 7$. (B) Scaling law properties of the boundary of discrimination. Solid line: $N = 1000$ inputs; middle dashed line: $N = 100$; dashed line-$N = 10$. The exponent of the power law is 2.12, $N_{KC} \propto n_{KC}^{2.12}$.

of information in the intrinsic KC layer of the locust is sparse (Perez-Orive et al., 2002). The percentage of iKC neurons that do not respond to the presentation of any given odor is 90% (the statistic was gathered with fewer than 100 neurons). This means that only $\approx 10\%$ are active during 1 second of odor response. Therefore, the probability that an iKC will be active in a window of 50 milliseconds is $p_{KC} \sim 0.005$. This means that on average, there are 250 active neurons in every 50 ms snapshot. If we compare this value to the theoretical sparseness values presented in Figure 6A, we obtain that the probability of discrimination is 97% for $N = 10$ uncorrelated inputs, 84% for $N = 100$, and 56% for $N = 1000$. As explained in section 3.1 these values can be improved by using redundant error-correcting code techniques.

It turns out that there is a mechanism that increases the boundary of successful discrimination and thus induces a better correspondence to the experimental observations. Let us assume in the following, for convenience of the analysis, that all connections from the iKCs to the extrinsic KC are initially very unlikely. Recall that our learning paradigm so far assumed that the learning time is much larger than $\max\{1/p_-, 1/p_+\}$ such that $\mathbf{w} = \mathbf{y}(\mathbf{x}^0)$. For finite time, however, there is a nonzero probability of $\mathbf{w} \neq \mathbf{y}(\mathbf{x}^0)$. The probability of having an active connection at time t during learning is $p_t = p_+(1-p_+)^{t-1}$, where t denotes the number of presentations of input $\mathbf{x}^0$. This probability is valid only for the connections whose presynaptic neuron is active. Therefore, we can write

$$p(l(\mathbf{w})) = \sum_{i=l(\mathbf{w})}^{N_{KC}} \binom{i}{l(\mathbf{w})} (1-p_t)^{l(\mathbf{w})}\, p_t^{i-l(\mathbf{w})}\, P(n_{KC} = i), \tag{3.2}$$

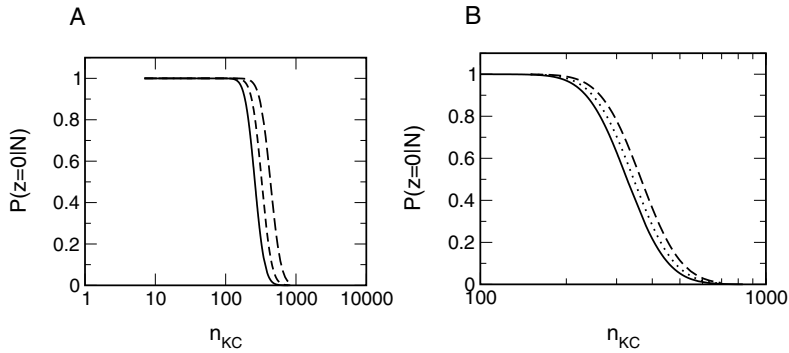

Figure 6: (A) Probability of discrimination for $N = 1000$ (solid line), $N = 100$ (middle dashed line), $N = 10$(dashed line) inputs as a function of the average number of active neurons n_{KC} in the intrinsic KC layer for a threshold value of activation in the MB lobe layer of $\theta_{\mathrm{LB}} = 7$ and $N_{KC} = 50000$ as in the locust. The experimental values for the locust are located near the boundary of the discrimination regime. (B) Effects of having less time to reinforce the positive odor. The solid line represents $p_t = 1$, the dotted line is for $p_t = 0.9$, and the dashed line for $p_t = 0.8$. A limited time for learning increases the boundaries of the discrimination regime.

where the $P(n_{KC} = i)$ is calculated in the appendix. The probability $p(l(\mathbf{w}))$ of having $l(\mathbf{w})$ connections from the iKCs to an eKC is used in equations A.17 and A.22 in the appendix. If we recalculate the probability of discrimination, we obtain the results shown in Figure 6B. If the probability of transition p_+ is decreased, the sparse region increases. So in this case, a limited time to learn the vector to be discriminated can increase the boundary of the parameter region for successful discrimination.

3.1.3 Discrimination Capacity. Given a set of structural parameters such as threshold for activation, θ_{LB}, size N_{KC} of the intrinsic KC layer, and level of activity p_{KC}, we try to determine the maximal number of inputs that can be discriminated; we determine the maximum value $N_{\max}$ of N such that $P(z(\mathbf{y}^1) = 0, \ldots, z(\mathbf{y}^N) = 0) \geq 1 - \epsilon$ where ϵ is a fixed small error tolerance. In the appendix (section A.2.4) we calculate the upper bound for the capacity

$$N_{\max} \leq \frac{\log P(z(\mathbf{y}^1) = 0, \ldots, z(\mathbf{y}^{N-1}) = 0)}{\log P(z = 0)} \approx \frac{\epsilon}{P(z = 1)}. \tag{3.3}$$

The result shows that capacity is proportional to the inverse of the probability of misclassification. It is trivial to say that if we increase the probability of discrimination, we can increase the capacity to discriminate. However, it

is not trivial to conclude that the capacity to discriminate is proportional to the inverse of the probability of misclassification.

3.1.4 Resolution of Discrimination. In the previous sections, we dealt with an uncorrelated set of inputs in the AL. This is a good starting point for achieving good discrimination abilities. The next question is what the resolution of the insect olfaction system is when the inputs are highly correlated or very similar.

Let us consider an input $\mathbf{x}$ that is learned as the vector $\mathbf{x}^0$, that is, the representation of $\mathbf{x}$ in the intrinsic KC layer, $\mathbf{y}(\mathbf{x})$, corresponds to $\mathbf{w}$. Now let us consider another vector $\mathbf{x}'$, which is at distance $l(\mathbf{x}' - \mathbf{x}) = \mu$. This means that the vector $\mathbf{x}'$ and $\mathbf{x}$ differ in μ ones. Let us define f as the event that the input $\mathbf{x}$ produces a spike in a neuron of the intrinsic KC layer and f' as the event that the input $\mathbf{x}'$ produces a spike in the same neuron of the intrinsic KC layer. We want to determine the probability that the input $\mathbf{x}'$ produces a spike given that the (similar) input $\mathbf{x}$ already produced a spike and that the distance between $\mathbf{x}'$ and $\mathbf{x}$ is μ, $l(\mathbf{x}' - \mathbf{x}) = \mu$. As always, we also need to condition for a given number of active AL neurons. We denote the desired probability as $q \equiv P(f' \mid n_{AL} = k, f, l = \mu)$. In the appendix (section A.2.5), we explain how to calculate q.

After obtaining q, we can calculate whether there are sufficient 1s in the intersection of $\mathbf{y}(\mathbf{x}')$ and $\mathbf{w} = \mathbf{y}(\mathbf{x})$ such that the eKC under consideration is above threshold for input $\mathbf{x}'$. Given that there are $n_{KC} = r$ active intrinsic KCs in $\mathbf{y}(\mathbf{x})$, the probability of having i 1s in the new vector $\mathbf{y}(\mathbf{x}')$ is

$$P(i \text{ coincident 1s} \mid n_{AL} = k, n_{KC} = r) = \binom{r}{i} q^i (1-q)^{n_{KC}-i}. \tag{3.4}$$

Therefore,

$$\begin{aligned} P(i \text{ coincident 1s}) = \sum_{k=0}^{N_{AL}} \sum_{r=i}^{N_{KC}} & P(i \text{ coincident 1s} \mid n_{AL} = k, n_{KC} = r) \\ & \times P(n_{KC} = r \mid n_{AL} = k) P(n_{AL} = k), \end{aligned} \tag{3.5}$$

where $P(n_{KC} = r \mid n_{AL} = k)$ and $P(n_{AL} = k)$ are given in the appendix (section A.1). We now have the tools to calculate the probability of $z(\mathbf{x}') = 1$ given that $z(\mathbf{x}) = 1$ and $l(\mathbf{x}' - \mathbf{x}) = \mu$:

$$P(z(\mathbf{x}') = 1 \mid z(\mathbf{x}) = 1, l(\mathbf{x}' - \mathbf{x}) = \mu) = 1 - \sum_{i=0}^{\theta_{LB}-1} P(i \text{ coincident 1s}) \tag{3.6}$$

Figure 7 shows the ability of the system to discriminate similar inputs as a function of the separation of the input states and for different average levels of activity in the intrinsic KC layer. The system is able to separate or

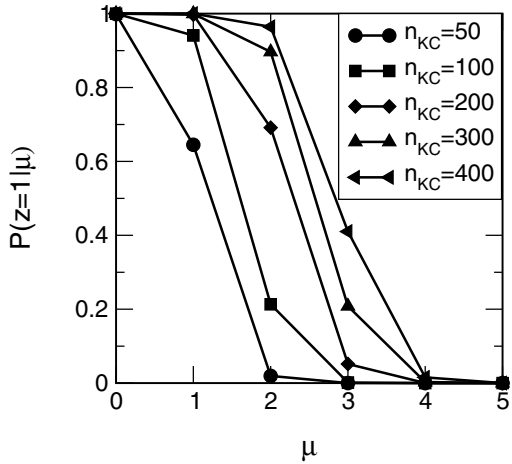

Figure 7: Ability to discriminate similar inputs as a function of distance between the inputs and different levels of average activity in the intrinsic KC layer. The probability of separating the first learned vector from a similar vector is high when the distance is larger than 2, 3, or 4 depending on the level of sparseness of the activity in the KC layer.

discriminate when the distance between the inputs is larger than 4, which is a rather small distance considering the dimension of the input space, that is, the number of AL neurons. We therefore conclude that it has a great ability to discriminate similar inputs.

3.2 Association for One Single Output. In the general problem of association, one has a set of uncorrelated inputs in the intrinsic KC layer $\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^d$. All these uncorrelated inputs need to produce a 1 in the classification neuron, $z(\mathbf{y}) = 1$, using Hebbian learning. The rule in equation 2.8 has two limiting cases depending on the choice of p_+ and p_-. The first limiting case is given for $p_+ \to 0$ and $p_- = 1$. Note that in this case, every time that there is a 0 in the input y_j, the connectivity w_j immediately drops to 0. Only the 1s occurring in all d inputs will produce a $w_j = 1$ in an average time $1/p_+$. Therefore, the connectivity vector $\mathbf{w}$ is eventually equal to the intersection of all the inputs, that is, $\mathbf{w} = \mathbf{y}^1 \cap \mathbf{y}^2 \cap \cdots \cap \mathbf{y}^d$. The second limiting case is given in the extreme $p_+ = 1$ and $p_- \to 0$. Following a similar argument as before, it can easily be seen that $\mathbf{w} = \mathbf{y}^1 \cup \mathbf{y}^2 \cup \cdots \cup \mathbf{y}^d$ in an average time $1/p_-$. Playing with p_+ and p_-, one can move $\mathbf{w}$ between these two limiting cases.

Let us study the disadvantages of the first case, $(p_+ \to 0, p_- = 1)$. To allow any activation of the output neuron, the number of 1s in $\mathbf{w}$ needs to be greater than or equal to θ_{LB}. One therefore needs to find parameters such

that the probability of having less than θ_{LB} 1s in $\mathbf{w}$ is very small, that is, $P(l(\mathbf{w}) < \theta_{\mathrm{LB}}) = \epsilon$ for some small given ϵ. Here, $l(\mathbf{w}) = \sum_j w_j$ the number of 1s in $\mathbf{w}$. The proper method to calculate this probability is explained in the appendix in section A.3. However, it is not possible to carry out due to its computational cost. Nevertheless, we can make a simple approximation that can contribute to clarifying the association problem. Let us approximate the probability distribution $p_{iKC}(n_{KC})$ by a binomial distribution with probability p_{KC}, where p_{KC} is the expected value of number of active neurons in the intrinsic KC layer. Then $P(l(\mathbf{w}) < \theta_{\mathrm{LB}})$ can easily be calculated as

$$P(l(\mathbf{w}) < \theta_{LB}) = \sum_{i=0}^{\theta_{\mathrm{LB}}-1} \binom{N_{KC}}{i} (p_{KC}^d)^i (1 - p_{KC}^d)^{N_{KC}-i}. \tag{3.7}$$

Using this probability, one can calculate the necessary level of activity p_{KC} in the intrinsic KC layer as a function of d for a fixed tolerance ϵ. Figure 8 shows that in order to be able to associate a few uncorrelated inputs, one needs to strongly increase the level of activity in the network up to unrealistic levels. And this is not all. One also needs to make sure that one can still discriminate the learned inputs from another sequence of uncorrelated inputs $\mathbf{y}^{d+1}, \mathbf{y}^{d+2}, \ldots, \mathbf{y}^{N+d+1}$ with the activity p_{KC} obtained from Figure 8. The way to measure this ability is to calculate $P(z(\mathbf{y}^{d+1}) = 0, z(\mathbf{y}^{d+2}) = 0, \ldots, z(\mathbf{y}^{N+d+1}) = 0)$. How to calculate this probability follows from equation A.22. The result of this calculation is that for $N = 1$ and $d = 2$, $P(z(\mathbf{y}^{d+1}) = 0, z(\mathbf{y}^{d+2}) = 0, \ldots, z(\mathbf{y}^{N+d+1}) = 0) = 10^{-7}$, whereas for larger

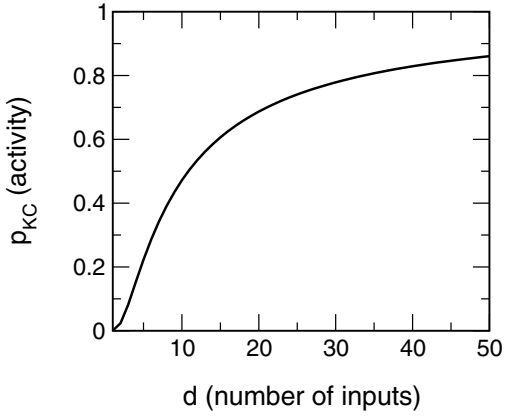

Figure 8: Necessary activity level p_{KC} for an intrinsic KC layer of 50,000 neurons to be able to associate d independent inputs in the intersection scenario. The threshold value of activation in the MB lobe layer was $\theta_{\mathrm{LB}} = 7$, and we used a tolerance level of $\epsilon = 10^{-6}$. Details are explained in the text.

values of d, the probability is practically 0. This means that in principle, one can associate d uncorrelated inputs. However, later, one can no longer distinguish them from the remainder of the N inputs. Therefore, the plasticity parameters that led to the intersection of the d inputs, ($p_+ \to 0, p_- = 1$), are not suitable for association.

Let us proceed to the second scenario in which one has the union of d inputs, ($p_+ = 1$, $p_- \to 0$). In this training process, active states of the preceding stimulus are diffused to avoid the restrictive condition of the intersecting sets. One obtains a connectivity vector that is the union of all the training input vectors, $\mathbf{w} = \mathbf{y}^1 \cup \mathbf{y}^2 \cup \cdots \cup \mathbf{y}^d$. The probability distribution $p(l(\mathbf{y}^1 \cup \mathbf{y}^2 \cup \cdots \cup \mathbf{y}^d))$ can in principle be calculated using the inclusion exclusion principle to obtain $l(\bigcup_{1 \le i \le d} \mathbf{y}^i)$, namely, $l(\bigcup_{1 \le i \le d} \mathbf{y}^i) = \sum_{1 \le i_1 \le d} l(\mathbf{y}^{i_1}) - \sum_{1 \le i_1 < i_2 \le d} l(\mathbf{y}^{i_1} \cap \mathbf{y}^{i_2}) + \sum_{1 \le i_1 < i_2 < i_3 \le d} l(\mathbf{y}^{i_1} \cap \mathbf{y}^{i_2} \cap \mathbf{y}^{i_3}) - \ldots + (-1)^{d+1} l(\bigcap_{i=1}^{d} \mathbf{y}^i)$ and the known probability distributions for the intersections discussed in the appendix. As it is already virtually impossible to evaluate the probability distributions for the intersections, the probability distribution for the length of the union is even less accessible. We therefore again assume independent firing events in the KC layer with probability p_{KC}. Then the probability of not having a connection in $\mathbf{w}$ is $(1 - p_{KC})^d$, where d is the number of associated inputs. With our independence assumption, the probability of having $l(\mathbf{w})$ 1s in $\mathbf{w}$ is

$$P(l(\mathbf{w})) = \binom{N_{KC}}{l(\mathbf{w})} (1 - (1 - p_{KC})^d)^{l(\mathbf{w})} ((1 - p_{KC})^d)^{N_{KC} - l(\mathbf{w})}. \tag{3.8}$$

This probability can be used in equation A.22 to obtain the probability for discriminating the rest of the nonassociated inputs, $P(z(\mathbf{y}^{d+1}) = 0, z(\mathbf{y}^{d+2}) = 0, \ldots, z(\mathbf{y}^{N+d+1}) = 0) = 1$. Note that we do not need to make sure that the output neuron fires for all the d inputs because $\mathbf{w}$ is the union of all those inputs. One can then calculate the maximum $d_{\max}$ of associated inputs such that $P(z(\mathbf{y}^{d+1}) = 0, z(\mathbf{y}^{d+2}) = 0, \ldots, z(\mathbf{y}^{N+d+1}) = 0) \ge 1 - \epsilon$ for a given tolerance level ϵ and a given N.

To test the validity of the simplifying independence assumption that led to the distribution 3.8 which differs significantly from the exact distribution, we take a semianalytical approach and produce the probability distribution for the length of the union of the d KC activity vectors numerically. In order to do so, we choose 1000 independent random connectivities and generate 1000 different input sets of d independent random inputs for each connectivity. We then calculate the length of the union of the KC activity vectors for the 10^6 chosen connectivity-input pairs. The accumulated histogram of relative numbers of occurrences for different lengths of the union vector is used as the approximation for the probability for this length. Once we have the numerical estimation of the distribution, we again use equation A.22.

Figure 9A shows the probability of discriminating N inputs given that d 1s have been associated. We can see that the semianalytical calculation and

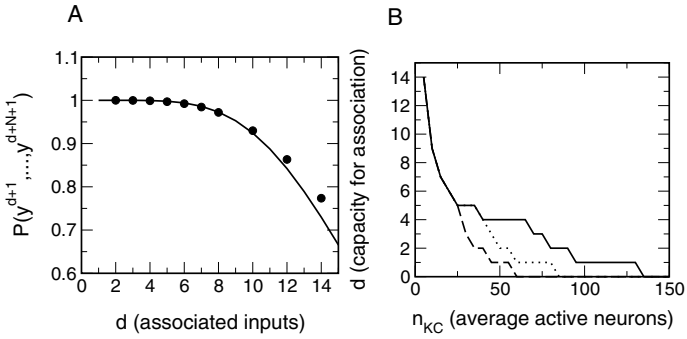

Figure 9: (A) Probability of discriminating d associated inputs from $N = 10$ uncorrelated inputs. The solid line represents the values using the approximation explained in the text. The circles show the semianalytical values, for which $p(l(\mathbf{y}^1 \cup \mathbf{y}^2 \cup \cdots \cup \mathbf{y}^d))$ was estimated from simulations. (B) Maximum number of associated inputs, $d_{\max}$ for an intrinsic KC layer of 50,000 neurons as a function of the average number of active neurons n_{KC} for a threshold value of activation in the MB lobe layer of $\theta_{\text{LB}} = 7$. The lines from left to right are for $N = 10^2, 10^4, 10^6$. We used a tolerance level of $\epsilon = 10^{-6}$.

the approximation start to separate for probability values far from 1. Since we are interested in determining $d_{\max}$ such that $P(z(\mathbf{y}^{d+1}) = 0, z(\mathbf{y}^{d+2}) = 0, \ldots, z(\mathbf{y}^{N+d+1}) = 0) \geq 1 - \epsilon$, that is, we work in a region where the probabilities are close to 1, we can reasonably use the approximation.

Figure 9B shows the maximum number of associated inputs $d_{\max}$ as a function of the average number of active neurons, $n_{KC} = p_{KC} N_{KC}$, in a KC layer of $N_{KC} = 50,000$ neurons. Sparse levels of activity achieve the higher values for the number of inputs that can be associated. However, the overall ability of the system to associate inputs is not very good. This implies a prediction on how many uncorrelated odors an insect can associate. From this analysis, we can see that there are surely no more than 10.

3.3 Self-Organization for Classification with Many Output Neurons. In the preceding sections, we were able to quantify the dependence of the classification performance of a single output neuron on the KC layer size and the level of activity in this layer. In this section, we investigate the performance of the full system with several output neurons in the decision layer. It remains to be seen whether the dependencies observed for the single output neuron are preserved for the system that self-organizes by mutual inhibition between eKCs and Hebbian learning as described in section 2.2.1.

First, let us be specific about the input classes used in the AL layer. We create a basis for a set of input classes by a Bernoulli process such that the probability of having a 1 in an AL neuron is p_{AL}; that is, we create N_c inde-

pendently chosen basis vectors $\eta^1, \eta^2, \ldots, \eta^{N_c}$. For each basis vector η, we then create a set or class of N_r inputs that are highly correlated to the basis vector η^μ. The vectors in each of the classes μ are denoted as $\mathbf{x}^j(\eta^\mu)$ where $\mu = 1, \ldots, N_c$ and $j = 1, \ldots, N_r$. The inputs $\mathbf{x}^j(\eta^\mu)$ belonging to class μ are generated by relocating each 1 in η^μ with probability p_r. For $p_r = 0$, all the inputs of the class are therefore exactly the same, $\mathbf{x}^j(\eta^\mu) = \eta^\mu$, and they become increasingly different from each other and the basis η^μ with increasing p_r. This mechanism to generate classes allows us to automatically create a large number of correlated and uncorrelated inputs with a controlled degree of correlation. The correlated inputs are those generated from the same basis vector η^μ, and the uncorrelated ones are those produced from different such vectors. We shall see that the correlated inputs will be classified by the same output neurons and the uncorrelated ones by different ones.

Many parameters need to be adjusted at this point. Let us first specify which parameters are fixed due to computational constraints. The number of input neurons, N_{AL}, is set to 100. This number is close to the number of projection neurons in the AL of the *Drosophila*. Note that the locust has 830, which makes it very difficult to gather sufficient statistics in this case due to long computation times. The intrinsic KC layer size, N_{KC}, is mainly set to 2500, again as in *Drosophila*. This parameter is occasionally varied to test the classification performance as a function of the screen layer. The activity level in the AL of locust is experimentally at 15% in a 50 millisecond snapshot, $p_{\mathrm{AL}} = 0.15$. We make the assumption that this is similar in *Drosophila*. The rest of the parameters, p_C and θ_{KC}, are chosen to keep the mapping from the AL activity onto the KC activity an injective function as explained in section 2.1 and to regulate the level of average activity, n_{KC}, in the intrinsic KC layer. The number of extrinsic KCs is not known, so we assume that $N_{LB} = 100$. Note that the qualitative results do not depend on this number.

The output response of the decision layer has n_W active neurons for every input $\mathbf{x}^j(\eta^\mu)$. This is achieved via an appropriate level of mutual inhibition between eKCs. The correlated inputs that belong to the same class should be classified by the same output response, while different inputs from different classes are to be separated, resulting in different output responses. To measure the degree of such separation or nonseparation, we define two quantities that quantify the distance within and among sets of outputs. Let us first denote the average output in response to a class μ of inputs by

$$\langle \mathbf{z}_\mu \rangle = \langle \mathbf{z}(\mathbf{x}^j(\eta^\mu)) \rangle_{\mathbf{x}^j(\eta^\mu)}, \tag{3.9}$$

where j takes the values $1, \ldots, N_r$ and labels the inputs within the class. Then the distance within a given class is defined as

$$D_{\mathrm{intra}} = \frac{1}{N_c} \sum_{\mu=1}^{N_c} \frac{1}{N_r} \sum_{j=1}^{N_r} |\mathbf{z}(\mathbf{x}^j(\eta^\mu)) - \langle \mathbf{z}_\mu \rangle|, \tag{3.10}$$

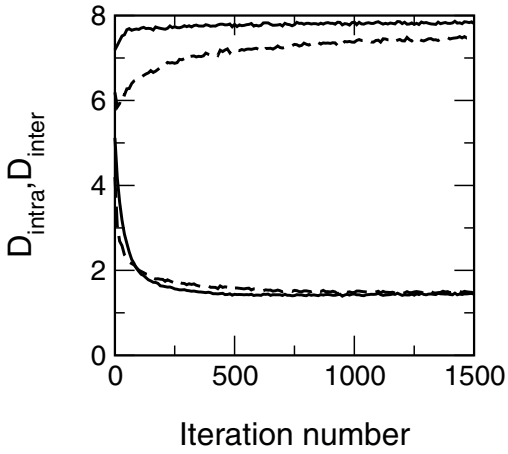

Figure 10: An example of the evolution of learning through several presentations of odors for $p_r = 0.9$, $n_W = 5$, $p_+ = 0.2$, $p_- = 0.5$, and average number of active neurons in the intrinsic KC layer set to $n_{KC} = 35$. The solid lines represents the inter- (upper) and intradistance (lower) for an initial probability of connection from the intrinsic KC to the MB lobe of 0.1. The dashed line represents an initial probability of connection of 0.05. This initial probability of connection is an important parameter for minimizing learning time.

and the interclass distance is expressed as

$$D_{\text{inter}} = \frac{1}{(N_c^2 - N_c)/2} \sum_{\mu=1}^{N_c-1} \sum_{\nu>\mu}^{N_c} |\langle \mathbf{z}_\mu \rangle - \langle \mathbf{z}_\nu \rangle|. \tag{3.11}$$

Note that we calculated the distance between classes using the average of the outputs in response to a class. We chose this definition over using the distance between all pairs of outputs of different classes to reduce computational costs. For a given n_W, the values of the distances are bounded, that is, $D_{\text{intra}}, D_{\text{inter}} \leq 2\, n_W$.

As an illustration of how the system self-organizes by means of mutual inhibition and Hebbian learning, Figure 10 shows the time evolution of the system through several odor presentations. The main parameter under study in this figure is the initial probability of connections from the intrinsic KC layer to the MB lobe. Lower initial connection probabilities lead to a high number of iterations before a steady state is achieved. In all our simulations, we chose the initial probability that leads to short learning times.

3.3.1 Level of Inhibition: Number of n_W Active MB Lobe Neurons. In contrast to the AL and the intrinsic KC layer, not much is known about the

connectivity and levels of activity within the MB lobes. Thus, the number of active neurons in the MB lobes, the threshold level, the probability of modification of synaptic connectivity, and the initial number of connections in the system (before learning starts) are free parameters. We need to elaborate on these parameters before running a parametric search.

First, theoretically, the number of classes that can be stored in the MB lobes is $\binom{N_{LB}}{n_W}$, which increases quickly with the number of active extrinsic KCs, n_W. If we knew the number of classes that the insect stores, we would be able to extract an estimate for n_W from this knowledge. It seems that the larger n_W, the better the system becomes. It is, however, unclear whether the system is able to learn a high n_W/N_{LB} ratio. n_W also determines the optimal value of the interdistance, D_{inter}.

Let us now consider two vectors that belong to different classes, and let us assume for a first analysis that $\mathbf{z}$ and $\mathbf{z}'$ are randomly chosen vectors with probability n_W/N_{LB} to have a 1 at any position. The probability that $\mathbf{z}$ and $\mathbf{z}'$ share a given active neuron is $(n_W/N_{LB})^2$. Then the probability of having j common active neurons is $P(j) = \binom{N_{LB}}{j}((n_W/N_{LB})^2)^j\,(1-(n_W/N_{LB})^2)^{n_W-j}$. So the average number of common neurons is n_W^2/N_{LB}. The expectation value for the optimal interdistance is then $D_{inter} \approx 2\,n_W - 2n_W^2/N_{LB}$. Therefore, although the capacity grows with n_W, the overlaps among responses increase proportional to n_W^2. These theoretical arguments are rather academic, as the activity vectors are clearly not independent Bernoulli processes in the real system.

In the simulations of the real system, we randomly choose 2000 inputs from the set of $N = 400$ generated inputs and present them to the system. Usually the system reaches a stationary state after 250 presentations. To produce reliable statistics, we perform a sufficient number of independent trials. Five hundred repetitions appeared to be the appropriate compromise between maximizing precision and minimizing computer time. Figure 11 shows one general example of the efficiency of classification as a function of the number of active neurons n_W in the MB lobes. The solid line represents the difference between the inter- and intradistance normalized to the maximum possible value, which is $2\,n_W$. As we can see, the optimal value to discriminate between classes while grouping the inputs belonging to the same class is $n_W = 1$. This result is consistent with associative networks, where sparseness is helpful to achieve minimum overlaps in the output (Frolov & Murav'ev, 1993; Nadal & Tolouse, 1990).

Shall we then choose $n_W = 1$ as a default value? There are additional criteria that we have not considered here: noise, the ability of the downstream reader to see one single neuron, and the implementation of the system with real conductance-based models. In all the calculations that we have shown here, we did not consider noise. Real neurons are, however, always noisy, and some redundancy in the system has to be expected. Second, when considering who reads the output of the MB lobes, we need to ask whether it

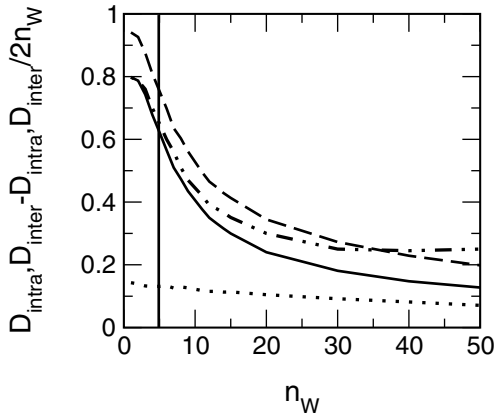

Figure 11: Estimates for the representation quality in the MB lobes as a function of the number of active neurons n_W in the lobes. The dotted line is D_{intra} normalized to the maximum value $2\,n_W$, the dashed line is D_{inter}, and the thick solid line is the difference between them normalized to the maximum. The dash-dotted line represents $D_{inter} - D_{intra}$ normalized to the expected maximal difference, which is $2\,n_W - 2n_W^2/N_{LB}$. The values used to obtain this curve are $N_c = 40$, $N_r = 10$, $p_r = 0.9$, $p_- = 0.5$, $p_+ = 0.2$, and $p_C = 0.21$. The different parameter sets used in the next sections produce the same qualitative results. In particular, increasing the number of classes N does not change the slope of the decreasing functions.

is possible to read the response of a single neuron. This is another reason for some redundancy in the representation. The last issue concerns the implementation of an n_W-representation with more realistic neuron models. It is difficult to adjust the level of mutual inhibition, then represented by realistically modeled inhibitory synapses with given maximal conductance, such that any input produces a single neuron to respond. Therefore, to avoid being unrealistic, we will consider $n_W = 5$ as a moderate choice for the level of activity in the MB lobes.

3.3.2 Intrinsic KC Layer Size. We have shown the dependence of the discrimination performance as a function of the intrinsic KC layer size, N_{KC}. To compare in the same terms as in previous sections, the activity of the intrinsic KC layer is again regulated to have a constant average number of active neurons n_{KC}. Here, we chose the same parameter values as in the previous section: $N_c = 40$, $N_r = 10$, $p_r = 0.1$, and $p_- = 0.5$. The probability of connections from the AL to the MB is set from equation 2.4 such that the average number of active neurons is $n_{KC} = 35$.

Figure 12 shows from left to right the observed inter- and intradistances after learning for different intrinsic KC layer sizes. The discrimination ability

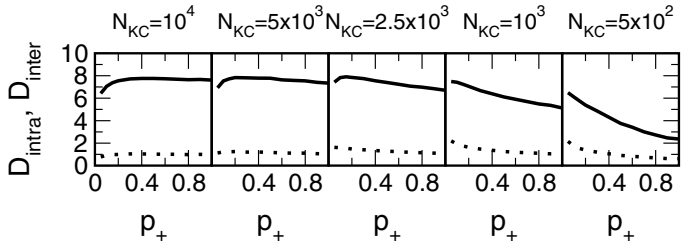

Figure 12: Classification ability among classes versus several sizes of the intrinsic KC layer. From left to right, the size runs from 10,000 neurons to 500. The solid line is the interdistance between clusters, and the dotted line is the average intradistance within the clusters. There is a clear improvement with increasing size of the intrinsic KC layer.

clearly improves with increasing N_{KC}. It is also clear that the positive plasticity timescale p_+ is critical for smaller KC layer sizes. There is an optimal value of p_+ where the discrimination ability among classes is best. However, we can also observe that the intradistance within a class gets worse with lower values of p_+. It seems reasonable that there is a balance between the ability of the system to separate classes and the ability to put together inputs that belong to the same class. Therefore, to be able to have an understanding of the overall performance between these two antagonistic goals, we plot $D_{\text{inter}} - D_{\text{intra}}$ versus p_+ in Figure 13. One of the main messages one can infer from this plot is that for larger sizes of the intrinsic KC layer, the learning rate is not so important. This means that one of the additional advantages of having a larger system size is more robust learning. For sizes larger than 5000, the system reaches a saturation level regardless of the learning rate. This size is smaller than the one in Figure 5 due to different numbers of inputs used.

3.3.3 Level of Activity in the Intrinsic KC Layer. We have demonstrated that sparse code in the intrinsic KC layer for one output neuron is optimal, as can be seen in Figure 6. We now analyze the optimal level of activity in the KC layer with a complete decision layer, including mutual inhibition between eKCs. To test the ability of the system to classify properly, we increase the difficulty of the problem by increasing the number of classes to 200. We then perform the same type of simulations as in the previous section, but now for different activity levels of the intrinsic KC layer. To regulate this activity level, we adjust p_C appropriately. To increase the speed of the simulations, we consider only a single MB of *Drosophila*, which has $N_{KC} = 2500$. Figure 14 shows the results of the simulations. They are consistent with the conclusions drawn in the previous section. High activity levels on the intrinsic KC layer do not perform very well. It appears that

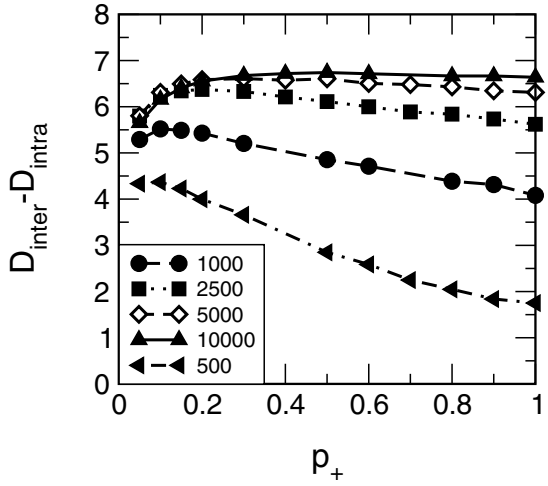

Figure 13: Difference between the inter- and intradistances observed versus the inverse learning rate p_+ extracted from Figure 12.

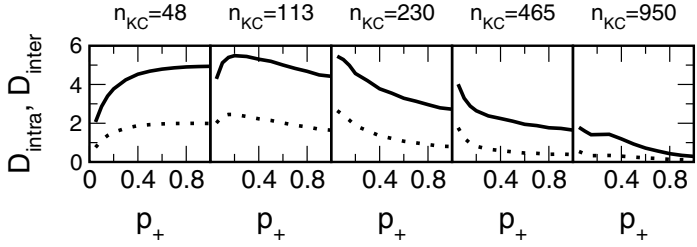

Figure 14: Classification ability among classes versus several activity levels on the intrinsic KC layer, n_{KC}, for a system with $N_{KC} = 2500$. The solid line represents the interdistance between clusters, and the dotted line is the intradistance within a cluster. It seems that $n_{KC} = 113$ classifies better than the rest, and large activity in the intrinsic KC layer produces very poor performance.

the best performance is achieved at activity levels near 100. Note that the interdistance is not very high because there are 200 different classes with 10 inputs for each class. This means that there are 2000 inputs that each trigger 5 output neurons in these simulations. Therefore, there have to be many output neurons that share classes.

To quantify the performance better, we plot the maximum over all tested p_+ of the difference between the inter- and intradistance as a function of the KC activity levels used, that is, $\max_{p_+}\{D_{\text{inter}} - D_{\text{intra}}\}$ as a function of n_{KC}.

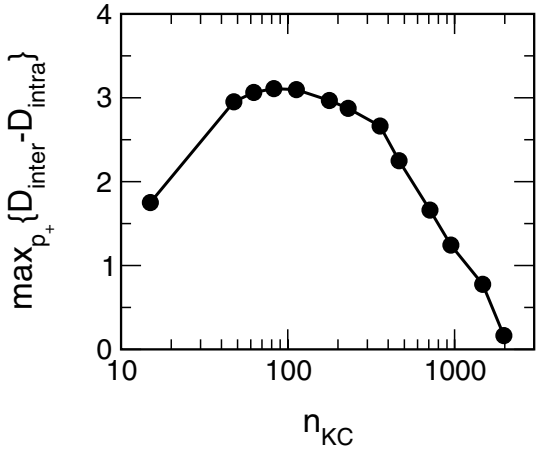

Figure 15: $\max_{p_+}\{D_{\text{inter}}-D_{\text{intra}}\}$ versus increasing values of activity n_{KC} for $N_{KC} = 2500$. There is an optimal activity level in the sparse regime close to $n_{KC} = 100$.

The result is shown in Figure 15. The best performance is obtained for n_{KC} roughly around 100. Note the logarithmic scale on the n_{KC} axis.

4 Conclusions

Our results clearly show that the classification task can be implemented in the first few relay stations of the olfactory system of insects. We have shown that there is no need for complicated, nonrealistic global learning rules or highly specific connections. The fan-out and fan-in network structure, nonspecific or random connections, simple Hebbian learning, and mutual inhibition can implement discrimination between classes and grouping of similar inputs into the same class.

Within this classification scheme are two main parameters that heavily affect the performance of the system: intrinsic KC layer size and level of activity in the MB. In simple terms, the larger the number of neurons, the better the insect will perform in the classification task. For the level of activity in the MB, there is an optimal value at a very sparse level. A straightforward prediction arises. The honeybee with 170,000 intrinsic KC neurons will outperform the *Drosophila* in any classification task. However, if we were somehow able to increase the level of activity in the honeybee MB by genetic mutation or chemical intrusion that increases the excitability of the neurons, our theory predicts that the performance will quickly decay.

We have also shown that the structural organization of the AL and the MB provides the basic means to obtain a high level of discrimination for very similar odors, as we have demonstrated. This indicates that the main

function of the intrinsic KC layer is to separate all the odor inputs as much as possible to be associated in the next layer, the MB lobe.

The analysis of the association of unrelated inputs into a single output response showed that only a limited number of inputs can be associated (see Figure 9). This can be a specific prediction of the analysis of this article. The number of uncorrelated odors that an insect can associate is limited to a few. An experiment to confirm this prediction would need to use a pool of uncorrelated odors. The detection of uncorrelated odors is in itself not a simple matter, but we take it for granted for the moment. Then the insect would need to be forced to associate d odors, and we would test whether the insect is able to discriminate these odors from N new, uncorrelated ones. A value of $d = 10$ associated odors is the maximum number that the insect might be able to achieve according to our analysis.

In summary, the correct level of sparseness and an appropriate system size are crucial for association and discrimination in this model that replicates the structural organization of the olfactory system of insects. The importance of sparseness for the capacity of associative memories is an old theoretical idea that was first proposed by Marr (1969) in a model of the cerebellar cortex, and Willshaw and coworkers (1969) in an associative memory model. Since then, a lot of progress has been made concerning the role of sparse representations to increase the capacity of artificial associative memories (Palm, 1980; Tsodyks & Feigel'man, 1988; Amari, 1989; Buhman, 1989; Perez-Vicent & Amit, 1989). Given that a sparse code has been found experimentally in the intrinsic KC neurons of the MB of locust (Perez-Orive et al., 2002), it is feasible to locate the classification task in the MB lobes. This is also consistent with experimental observations about the location of learning in odor conditioning.

5 Discussion

We have explored sufficient conditions to achieve classification in the olfactory system of insects. An issue that we have disregarded in this work is the temporal aspects of the neuronal code. Is there a need for a temporal code? From the results obtained, it is clear that increasing the activity in the MB degrades the performance of the system. So what happens if the presented odor concentrations are increased? With increasing odor concentration, the number of active sensory receptors increases, and increased activity at the glomerular level is recorded (Ng et al., 2002; Wang et al., 2003). On the other hand, recordings in the MB of *Drosophila* show that in normal specimens the levels of activity do not seem to change for different odor concentrations (Wang et al., 2001). Recent investigations have shown that the average activity of the excitatory neurons in the locust AL (Stopfer et al., 2003) remains nearly constant across odor concentrations as well. In between the glomeruli and the MB, there obviously is a gain control mechanism that regulates the activity being fed into the MB. The network of excitatory and

inhibitory neurons in the AL is a good candidate to perform this function. In this work, we did not consider this function and how to integrate the temporal snapshots of activity in the AL. We explored just the boundaries and limitations of a neural network using biologically feasible properties while disregarding time. Summarizing, the olfactory system certainly needs a gain control mechanism in order to maintain the activity of the MB within its working regime.

Another suggested function of the AL that we did not address here is the role of spatiotemporal code to separate similar odors. Hosler, Buxton, and Smith (2000) Stopfer, Bhagavan, Smith, and Laurent (1997) showed that temporal activity plays a role in behavior. Although different odors can be discriminated without temporal activity, similar odors require spatiotemporal activity to be separated (Hosler et al., 2000; Stopfer et al., 1997). It has been shown experimentally (Friedrich & Laurent, 2001) and theoretically (Friedrich & Laurent, 2001) that time improves discrimination. In this article, we show that the transformation from the AL to the MB using a random connectivity matrix can produce a very good level of resolution for similar odors (see Figure 7) without using temporal mechanisms.

We like to think that the connections from the AL to the MB are random not only because of the mathematical simplicity but also because it works well. We think that random connections from the AL to the MB allow the system to be ready for any type of input, such that it does not matter whether the insect knows the odor. A similar way of thinking can be found in O'Reilly and McClelland (1994), where random transformations from the entorhinal cortex to the dentate gyrus in the hippocampus are found to be better in achieving pattern separation. In principle, we do not think that there is need for learning at this stage. One possible functional role of synaptic plasticity has been shown in (Finelli et al., 2004), where synaptic learning improves the sparse representation in the intrinsic KC layer.

Appendix

A.1 $P(n_{KC} = r)$**.** The probability that a given intrinsic neuron fires is given by

$$p_{KC} = \sum_{i=\theta_{KC}}^{N_{AL}} \binom{N_{AL}}{i} (p_{AL} p_C)^i (1 - p_{AL} p_C)^{N_{AL}-i}. \tag{A.1}$$

However, in order to estimate the probability of discrimination, we need to calculate the probability distribution $P(n_{KC} = r)$ to have $n_{KC} = r$ active intrinsic KCs. The conditional probability that an intrinsic KC fires given that $n_{AL} = k$ PNs are active is

$$p(\text{fire} \mid n_{AL} = k) = \sum_{i=\theta_{KC}}^{k} \binom{k}{i} p_C^i (1 - p_C)^{k-i}. \tag{A.2}$$

Then the conditional probability to have $n_{KC} = r$ active intrinsic KCs given $n_{AL} = k$ active PNs is

$$P(n_{KC} = r \mid n_{AL} = k) = \binom{N_{KC}}{r} p(\text{fire} \mid n_{AL} = k)^r \times (1 - p(\text{fire} \mid n_{AL} = k))^{N_{KC}-r}. \quad \text{(A.3)}$$

because of the independently chosen connections. Finally, the probability distribution for the number of active n_{KC} is

$$P(n_{KC} = r) = \sum_{k=0}^{N_{AL}} P(n_{KC} = r \mid n_{AL} = k) \binom{N_{AL}}{k} p_{AL}^k (1 - p_{AL})^{N_{AL}-k}. \quad \text{(A.4)}$$

Unfortunately, this probability distribution cannot be expressed in simpler terms and needs to be evaluated numerically.

The expectation value of this distribution is by definition

$$\mathbb{E}\, n_{KC} = \sum_{r=0}^{N_{KC}} r P(n_{KC} = r) \quad \text{(A.5)}$$

$$= \sum_{k=0}^{N_{AL}} \binom{N_{AL}}{k} p_{AL}^k (1 - p_{AL})^{N_{AL}-k} \sum_{r=0}^{N_{KC}} r \binom{N_{KC}}{r} \times p(\text{fire} \mid n_{AL} = k)^r (1 - p(\text{fire} \mid n_{AL} = k))^{N_{KC}-r}. \quad \text{(A.6)}$$

The second sum is the expectation value of a binomial distribution with parameters N_{KC} and $p(\text{fire} \mid n_{AL} = k)$ and thus is equal to $N_{KC}\, p(\text{fire} \mid n_{AL} = k)$ such that

$$\mathbb{E} n_{KC} = N_{KC} \sum_{k=0}^{N_{AL}} \binom{N_{AL}}{k} p_{AL}^k \left(1 - p_{AL}\right)^{N_{AL}-k} \sum_{i=\theta_{KC}}^{N_{AL}} \binom{k}{i} p_C^i (1 - p_C)^{k-i} \quad \text{(A.7)}$$

$$= N_{KC} \sum_{i=\theta_{KC}}^{N_{AL}} \binom{N_{AL}}{i} (p_{AL} p_C)^i (1 - p_{AL} p_C)^{N_{AL}-i}. \quad \text{(A.8)}$$

Note that this is the same expectation value as one obtains assuming independence of firing events of intrinsic KCs. The standard deviation of the probability distribution, equation A.4, is, however, quite a bit larger than the standard deviation of the binomial distribution obtained under this simplifying assumption.

A.2 $P(z(\mathbf{y}^1) = 0, z(\mathbf{y}^2) = 0, \ldots, z(\mathbf{y}^N) = 0)$**.**

A.2.1 Preliminary Analysis. All the activity vectors in the intrinsic KC layer are obtained from independently, randomly generated activity vectors in the AL. Assuming that a randomly chosen activity vector $\mathbf{y}'(\mathbf{x}')$ has been trained sufficiently long, the resulting connectivity vector will be $\mathbf{w} = \mathbf{y}'$, as explained in section 3.1. If another input vector $\mathbf{y}(\mathbf{x})$ is chosen randomly, the probability of a proper classification is

$$P_{\text{proper}} = P(z = 0 \mid \mathbf{y} \neq \mathbf{y}'). \tag{A.9}$$

It is problematic to calculate this conditional probability directly, so let us try to simplify it:

$$P(z = 0) = P_{\text{proper}} P(\mathbf{y} \neq \mathbf{y}') + P(z = 0 \mid \mathbf{y} = \mathbf{y}') P(\mathbf{y} = \mathbf{y}') \tag{A.10}$$

$$= P_{\text{proper}}(1 - P(\mathbf{y} = \mathbf{y}')). \tag{A.11}$$

The probability $P(\mathbf{y} = \mathbf{y}')$ of having obtained the same activity vector twice is identical to $P(\mathbf{x} = \mathbf{x}')$ if we assume that the mapping from the AL to the MB is an injective function. In section 2.1, we give the conditions on the parameters that ensure an injective function. Appropriate parameters are used throughout this article. We conclude,

$$P(\mathbf{y} = \mathbf{y}') = P(\mathbf{x} = \mathbf{x}') = \sum_{\mathbf{x}} P(\mathbf{x}) P(\mathbf{x})$$

$$= \sum_{l=0}^{N_{\text{AL}}} \binom{N_{AL}}{l} p_{AL}^{2l} (1 - p_{AL})^{2(N_{AL} - l)} \tag{A.12}$$

$$= \left(p_{AL}^2 + (1 - p_{AL})^2 \right)^{N_{AL}} \tag{A.13}$$

For example, the locust has $N_{AL} = 830$ neurons and $p_{\text{AL}} = 0.15$; this probability is approximately 10^{-106}. Another example is the *Drosophila* with $N_{\text{AL}} = 140$. In this example, the probability of collision is 10^{-17}. We therefore can safely neglect $P(\mathbf{y} = \mathbf{y}')$ in equation A.11 and call $P(z = 0)$ the probability of proper classification.

A.2.2 Probability to Discriminate Another Randomly Chosen Input. The probability of having i coincident ones in a randomly chosen activity vector $\mathbf{y}$ with $l(\mathbf{y})$ 1s, that is, $l(\mathbf{y}) = \sum_i y_i$, and a given connectivity vector $\mathbf{w}$ is

$$P(i \mid \mathbf{w}, l(\mathbf{y})) = \frac{\binom{l(\mathbf{w})}{i} \binom{N_{KC} - l(\mathbf{w})}{l(\mathbf{y}) - i}}{\binom{N_{KC}}{l(\mathbf{y})}}. \tag{A.14}$$

Then the probability of proper classification given $\mathbf{w}$ and $l(\mathbf{y})$ is

$$P(z = 0 \mid \mathbf{w}, l(\mathbf{y})) = \sum_{i=0}^{\theta_{LB} - 1} P(i \mid \mathbf{w}, l(\mathbf{y})). \tag{A.15}$$

In the next section, we will use

$$P(z=0\,|\,\mathbf{w}) = \sum_{l(\mathbf{y})=0}^{N_{KC}} P(z=0\,|\,\mathbf{w}, l(\mathbf{y}))P(n_{KC}=l(\mathbf{y})). \tag{A.16}$$

Finally, the total probability of proper classification for two random inputs is

$$P(z=0) = \sum_{l(\mathbf{w})=0}^{N_{KC}} P(z=0\,|\,\mathbf{w})P(n_{KC}=l(\mathbf{w})). \tag{A.17}$$

A.2.3 Probability of Discrimination Failure for a Set of N Inputs. We now generalize the result obtained in the previous section to a set of N inputs. The conditional probability for a given connectivity $\mathbf{w}$ and a single input is generalized to the joint conditional probability to produce $z=0$ with each of a set of N inputs. First,

$$\begin{aligned} &P(z(\mathbf{y}^1)=0, z(\mathbf{y}^2)=0, \ldots, z(\mathbf{y}^N)=0\,|\,\mathbf{w}) \\ &\quad = \prod_{j=1}^{N} P(z(\mathbf{y}^j)=0\,|\,\mathbf{w}), \end{aligned} \tag{A.18}$$

because all $\mathbf{y}^j$ are statistically independent (as they are functions of the independently chosen $\mathbf{x}^j$). As all $\mathbf{x}^j$ are generated by identical Bernoulli processes,

$$P(z(\mathbf{y}^j)=0\,|\,\mathbf{w}) = P(z(\mathbf{y}^1)=0\,|\,\mathbf{w}) \quad \forall j, \tag{A.19}$$

such that we obtain

$$P(z(\mathbf{y}^1)=0, z(\mathbf{y}^2)=0, \ldots, z(\mathbf{y}^N)=0) \tag{A.20}$$

$$= \sum_{l(\mathbf{w})=0}^{N_{KC}} P(z(\mathbf{y}^1)=0, z(\mathbf{y}^2)=0, \ldots, z(\mathbf{y}^N)=0\,|\,\mathbf{w}) \times P(n_{KC}=l(\mathbf{w})) \tag{A.21}$$

$$= \sum_{l(\mathbf{w})=0}^{N_{KC}} (P(z=0\,|\,\mathbf{w}))^N P(n_{KC}=l(\mathbf{w})), \tag{A.22}$$

where we used $P(z(\mathbf{y}^1)=0\,|\,\mathbf{w}) = P(z=0\,|\,\mathbf{w})$.

A.2.4 Discrimination Capacity. To estimate the capacity, we need to find the maximum value of N such that $1 - P(z(\mathbf{y}^1) = 0, \ldots, z(\mathbf{y}^N) = 0)$ is below a given small tolerance level ϵ. Thus, we take equation A.22 and apply the Jensen inequality to the convex function $f(x) = x^N$, leading to

$$\begin{aligned} &P(z(\mathbf{y}^1) = 0, \ldots, z(\mathbf{y}^N) = 0) \\ &\quad = \sum_{l(\mathbf{w})=0}^{N_{KC}} (P(z = 0 \mid \mathbf{w}))^N P(n_{KC} = l(\mathbf{w})) \qquad \text{(A.23)} \\ &\quad \geq \left(\sum_{l(\mathbf{w})=0}^{N_{KC}} P(z = 0 \mid \mathbf{w}) P(n_{KC} = l(\mathbf{w})) \right)^N \qquad \text{(A.24)} \\ &\quad = (P(z = 0))^N. \qquad \text{(A.25)} \end{aligned}$$

Note that $P(z = 0)$ has been calculated in section A.2. So N is bounded by

$$N \leq \frac{\log P(z(\mathbf{y}^1) = 0, \ldots, z(\mathbf{y}^N) = 0)}{\log P(z = 0)}. \qquad \text{(A.26)}$$

Using $P(z(\mathbf{y}^1) = 0, \ldots, z(\mathbf{y}^N) = 0) = 1 - \epsilon$ and taking into account that with $P(z(\mathbf{y}^1) = 0, \ldots, z(\mathbf{y}^N) = 0)$ close to one automatically also $P(z = 0)$ is very close to one, we can write in very good approximation

$$N \leq \frac{\log P(z(\mathbf{y}^1) = 0, \ldots, z(\mathbf{y}^N) = 0)}{\log P(z = 0)} \approx \frac{\epsilon}{P(z = 1)}. \qquad \text{(A.27)}$$

A.2.5 $P(f' \mid n_{AL} = k, f, l = \mu)$. The probability q can be expressed as

$$\begin{aligned} q &= P(f' \mid n_{AL} = k, f, l = \mu) \qquad \text{(A.28)} \\ &= \sum_{\eta=0}^{N_{AL}} P(f' \mid \textstyle\sum_j c_j x_j = \eta, n_{AL} = k, f, l = \mu) \\ &\quad \times P(\textstyle\sum_j c_j x_j = \eta \mid n_{AL} = k, f, l = \mu) \qquad \text{(A.29)} \\ &= \sum_{\eta=\max\{\theta_{KC},\mu\}}^{N_{AL}} P(f' \mid \textstyle\sum_j c_j x_j = \eta, n_{AL} = k, l = \mu) \\ &\quad \times P(\textstyle\sum_j c_j x_j = \eta \mid n_{AL} = k, f). \qquad \text{(A.30)} \end{aligned}$$

The last factor can be rewritten as

$$\begin{aligned} P(\textstyle\sum_j c_j x_j = \eta \mid n_{AL} = k, f) &= \frac{P(\sum_j c_j x_j = \eta \cap f \cap n_{AL} = k)}{P(f \mid n_{AL} = k) P(n_{AL} = k)} \\ &= \frac{P(\sum_j c_j x_j = \eta \mid n_{AL} = k)}{P(f \mid n_{AL} = k)}. \qquad \text{(A.31)} \end{aligned}$$

All factors are now known, namely,

$$P(f' \mid \textstyle\sum_j c_j x_j = \eta, n_{AL} = k, l = \mu) \tag{A.32}$$

$$= \begin{cases} \displaystyle\sum_{i=\max\{\mu-(\eta-\theta_{KC}),0\}}^{\mu} \binom{\mu}{i} (p_C)^i (1-p_C)^{\mu-i} & \text{for } k \geq \eta,\ \ \eta \geq \theta_{KC}, \mu \\ 0 & \text{otherwise} \end{cases},$$

$$P(\textstyle\sum_j c_j x_j = \eta \mid n_{AL} = k) = \binom{k}{\eta} p_C^{\eta} (1-p_C)^{k-\eta}, \tag{A.33}$$

and $P(f \mid n_{AL} = k)$ is given by equation A.2.

A.3 $P(l(\mathbf{y}^1 \cap \mathbf{y}^2 \cap \cdots \cap \mathbf{y}^d) < \theta_{LB})$**.** The probability of having i_1 intersecting 1s given two vectors $\mathbf{y}^1$ and $\mathbf{y}^2$, whose lengths are l_1 and l_2, is

$$P(i_1 \mid l_1, l_2) = \frac{\binom{l_1}{i_1}\binom{N_{KC}-l_1}{l_2-i_1}}{\binom{N_{KC}}{l_2}}. \tag{A.34}$$

The probability of having i_2 intersecting 1s of the vector $\mathbf{y}^3$ given i_1 coincidences of the first two vectors is

$$P(i_2 \mid l_1, l_2, l_3, i_1) = \frac{\binom{i_1}{i_2}\binom{N_{KC}-i_1}{l_3-i_2}}{\binom{N_{KC}}{l_3}}. \tag{A.35}$$

Given that we know $P(i_1 \mid l_1, l_2)$, we can now write

$$P(i_2 \mid l_1, l_2, l_3) = \sum_{i_1=0}^{\min\{l_1,l_2,l_3\}} \frac{\binom{i_1}{i_2}\binom{N_{KC}-i_1}{l_3-i_2}}{\binom{N_{KC}}{l_3}} \frac{\binom{l_1}{i_1}\binom{N_{KC}-l_1}{l_2-i_1}}{\binom{N_{KC}}{l_2}}. \tag{A.36}$$

We can continue this procedure to calculate the probability of having i_{d-1} 1s in the vector $(\mathbf{y}^1 \cap \mathbf{y}^2 \cap \cdots \cap \mathbf{y}^d)$, which is

$$P(i_{d-1} \mid l_1, l_2, \ldots, l_d) = \frac{1}{\prod_{j=2}^{d}\binom{N_{KC}}{l_j}} \sum_{\substack{i_{d-2},i_{d-3},\ldots,i_1=0 \\ i_{d-2}\leq i_{d-3}\leq\ldots\leq i_1}}^{l_{\min}} \binom{l_1}{i_1}\binom{N_{KC}-l_1}{l_2-i_1} \times \prod_{j=1}^{d-2}\binom{i_j}{i_{j+1}}\binom{N_{KC}-i_j}{l_{j+2}-i_{j+1}}, \tag{A.37}$$

where $l_{\min} = \min\{l_1, l_2, \ldots, l_d\}$.

The not-firing probability of the extrinsic KC is $P(l(\mathbf{y}^1 \cap \mathbf{y}^2 \cap \cdots \cap \mathbf{y}^d) < \theta_{LB})$. Therefore, the conditional not-firing probability given $(l_1, l_2, \ldots, l_d)$ is

$$P(\overline{\text{fire}} \mid l_1, l_2, \ldots, l_d) = \sum_{i_{d-1}=0}^{\theta_{LB}-1} P(i_{d-1} \mid l_1, l_2, \ldots, l_d). \tag{A.38}$$

Finally,

$$P(l(\mathbf{y}^1 \cap \mathbf{y}^2 \cap \cdots \cap \mathbf{y}^d) < \theta_{LB}) = \sum_{l_1, l_2, \ldots, l_d=0}^{N_{KC}} P(\overline{\text{fire}} \mid l_1, l_2, \ldots, l_d) \prod_{j=1}^{d} P(n_{KC} = l_j), \tag{A.39}$$

where $P(n_{KC} = l_j)$ is given in Section A.1. The main problem with this calculation is the computer time required to evaluate it. The minimum number of calls to the hypergeometric function based on the multiprecision library (Torbjorn, 2001) is $(N_{\max}(d-1)\theta_{LB})^d$. For the parameter values corresponding to the locust we have $\theta_{LB} = 7$ and $N_{\max} = 2000$, where $N_{\max}$ is defined by the condition

$$\sum_{i=0}^{N_{\max}} P(n_{KC} = i) = 1 - \epsilon, \tag{A.40}$$

with $\epsilon = 10^{-6}$. The number of associated inputs that we can analyze is therefore with $d \leq 4$ rather small. An approximation is required to draw conclusions for larger d.

Acknowledgments

We are very grateful to Gilles Laurent for conversations about this work. We also thank Ildiko Aradi and Gabriel Mindlin for constructive criticisms. This work was partially supported by the National Science Foundation under grants NSF/EIA-0130708 and NSF PHY0097134, and the U. S. Department of Energy, Office of Basic Energy Sciences, Division of Engineering and Geosciences, under grants DE-FG03-90ER14138 and DE-FG03- 96ER14592. This work was also partially supported by M. Ciencia y Tecnología BFI2000-0157 (M.G.).

References

Amari, S. (1989). Characteristics of sparsely encoded associative memory. *Neural Networks, 2*, 451–457.

Buhmann, J. (1989). Oscillations and low firing rates in associative memory neural networks. *Phys. Rev. A, 40*(7), 4145–4148.

Caticha, C., Palo Tejada, J. E., Lancet, D., & Domany, E. (2002). Computational capacity of an odorant discriminator: The linear separability of curves. *Neural Comput., 14*, 2201–2220.

Connolly, J. B., Roberts, I. J., Armstrong, J. D., Kaiser, K., Forte, M., Tully, T., & O'Kane, C. J. (1996). Associative learning disrupted by impaired Gs signaling in Drosophila mushroom bodies. *Science, 274*(5295), 2104–2107.

Cortes, C,. & Vapnik, V. (1995). Support vector networks. *Machine Learning, 20*, 273–297.

Cover, T. (1965). Geometric and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE T Elect. Comput., 14*, 326.

de Belle, J. S., & Heisenberg, M. (1994). Associative odor learning in *Drosophila* abolished by chemical ablation of mushroom bodies. *Science, 263*, 692–695.

Distler, P. G., Bausenwein, B., & Boeckh, J. (1998). Localization of odor-induced neuronal activity in the antennal lobes of the blowfly *Calliphora vicina*, a [3H] 2-deoxyglucose labeling study. *Brain Res.*, 805, 263–266.

Dubnau, J., Chiang, A. S., & Tully, T. (2003). Neural substrates of memory: From synapse to system. *J. Neurobiol., 54*(1), 238–253.

Dubnau, J., Grady, L., Kitamoto, T., & Tully, T. (2001). Disruption of neurotransmission in *Drosophila mushroom* body blocks retrieval but not acquisition of memory. *Nature, 411*(N6836), 476–480.

Finelli, L. A., Haney, S., Bazhenov, M., Stopfer, M., Sejnowski, T. J., & Laurent, G. (2004). *Effects of a synaptic learning rule on the sparseness of odor representations in a model of the locust olfactory system.* Unpublished manuscript, Salk Institute, San Diego, CA.

Friedrich, R,. & Laurent, G. (2001). Dynamical optimization of odor representations in the olfactory bulb by slow temporal patterning of mitral cell activity. *Science, 291*, 889–894.

Frolov, A. A. & Murav'ev, I. P. (1993). Informational characteristics of neural networks capable of associative learning based on Hebbian plasticity. *Network, 4*, 495–536.

Galizia, C. G., Joerges, J., Kuettner, A., Faber, T., & Menzel, R. (1997). A semi-in-vivo preparation for optical recording of the insect brain. *J. Neurosci. Meth., 76*, 61–69.

Galizia, C. G., Nagler, K., Holldobler, B., & Menzel, R. (1998). Odor coding is bilaterally symmetrical in the antennal lobes of honeybees (*Apis mellifera*). *Eur. J. Neurosci., 10*, 2964–2974.

Galizia, C. G., Sachse, S., Rappert, A., & Menzel, R. (1999). The glomerular code for odor representation is species specific in the honeybee *Apis mellifera R. Nature Neurosci., 2*, 473–478.

Gao, Q., Yuan, B., & Chess, A. (2000). Convergent projections of *Drosophila* olfactory neurons to specific glomeruli in the antennal lobe. *Nat. Neurosci., 3*(8),780–785.

Garcia-Sanchez, M., & Huerta, R. (2003). Design parameters of the fan-out phase of sensory systems. *J. Comput. Neurosci., 15*, 5–17.

Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell System Technical Journal, 26*(2), 147–160.

Hebb, D. (1949). *The organization of behavior*. New York: Wiley.

Heisenberg, M. (2003). Mushroom body memoir, from maps to models. *Nature Rev. Neurosci., 4*, 266–275.

Heisenberg, M., Borst, A., Wagner, S., and Byers, D. (1985). *Drosophila* mushroom body mutants are deficient in olfactory learning. *J. Neurogenet., 2*, 1–30.

Hendin, O., Horn, D., & Hopfield, J. J. (1994). Decomposition of a mixture of signals in a model of the olfactory bulb. *P. Natl. Acad. Sci. USA, 91*(13), 5942–5946.

Hendin, O., Horn, D., & Tsodyks, M. V. (1998). Associative memory and segmentation in an oscillatory neural model of the olfactory bulb. *J. Comput. Neurosci., 5*(2), 157–169.

Hosler, J. S., Buxton, K. L., & Smith, B. H.(2000). Impairment of olfactory discrimination by blockade of GABA and nitric oxide activity in the honey bee antennal lobes. *Behav. Neurosci., 114*(3), 514-525.

Joerges, J., Kuettner, A., Galizia, C. G., & Menzel, R. (1997). Representations of odors and odor mixtures visualized in the honeybee brain. *Nature, 387*, 285–288.

Laurent, G. (2002). Olfactory network dynamics and the coding of multidimensional signals. *Nature Rev. Neurosci., 3*, 884–895.

Laurent, G., Stopfer, M., Friedrich, R. W., Rabinovich, M. I., Volkovskii, A., & Abarbanel, H. D. I. (2001). Odor encoding as an active, dynamical process, Experiments, computation, and theory. *Annual Rev. Neurosci., 24*, 263–297.

Li, Z., & Hertz, J. (2000). Odour recognition and segmentation by a model olfactory bulb and cortex. *Network Comput. Neural Syst., 11*, 83–102.

Linster, C., & Cleland, T. A. (2001). How spike synchronization among olfactory neurons can contribute to sensory discrimination. *J. Comput. Neurosci., 10*(2), 187–193.

Marr, D. (1969). A theory of cerebellar cortex. *J. Physiol., 202*, 437–470.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5*, 115–133.

McGuire, S. E., Le, P. T., & Davis, R. L. (2001). The role of *Drosophila mushroom* body signaling in olfactory memory. *Science, 293*(5533), 1330–1333.

Menzel, R. (2001). Searching for the memory trace in a mini-brain, the honeybee. *Learn. Mem., 8*(2), 53–62.

Nadal, J. P. (1991). Associative memory, on the (puzzling) sparse coding limit. *J. Phys. A, 24*, 1093–1101.

Nadal, J. P., & Tolouse, G. (1990). Information storage in sparsely coded memory nets. *Network, 1*, 61–74.

Ng, M., Roorda, R. D., Lima, S. Q., Zemelman, B. V., Morcillo, P., & Miesenbock, G. (2002). Transmission of olfactory information between three populations of neurons in the antennal lobe of the fly. *Neuron, 36*(3), 463–474.

Nowotny, T., & Huerta, R. (2003). Explaining synchrony in feed-forward networks: Are McCulloch-Pitts neurons good enough? *Biol. Cyber., 89*(4), 237-241.

O'Reilly, R. C. (2001). Generalization in interactive networks: The benefits of inhibitory competition and Hebbian learning. *Neural Comput., 13*,1199–1241

O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall, avoiding a trade-off. *Hippocampus, 4*(6), 661-682.

Palm, G. (1980). On associative memory. *Biol. Cyber., 36*, 59–71.

Pascual, A., & Preat, T. (2001). Localization of long-term memory within the Drosophila mushroom body. *Science, 294*(5544), 1115–1117.

Perez-Orive, J., Mazor, O., Turner, G. C., Cassenaer, S., Wilson, R. I., & Laurent, G. (2002). Oscillations and the sparsening of odor representations in the mushroom body. *Science, 297*, 359–365

Perez Vicente, C. J., & Amit, D. J. (1989). Optimised network for sparsely coded patterns. *J. Phys. A, 22*(5), 559–569.

Rodrigues, V. (1988). Spatial coding of olfactory information in the antennal lobe of *Drosophila melanogaster. Brain Res., 453*, 299–307.

Scott, K., Brady, R. Jr., Cravchik, A., Morozov, P., Rzhetsky, A., Zuker, C., & Axel, R. (2001). A chemosensory gene family encoding candidate gustatory and olfactory receptors in *Drosophila. Cell, 104*(5), 661–673.

Stopfer, M., Bhagavan, S., Smith, B. H., & Laurent, G. (1997). Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Nature, 390*(6655), 70-74.

Stopfer, M., Jayaraman, V., & Laurent, G. (2003). Intensity versus identity coding in an olfactory system. *Neuron, 39*(6), 991–1004.

Torbjorn, G. (2001). *GNU MP: The GNU multiple precision arithmetic library, version. 4.0.1.* Available on-line at: http://www.swox.com/gmp/.

Tsodyks, M. V., & Feigel'man, M. V. (1988). The enhanced storage capacity in neural networks with low activity level. *Europhys. Lett., 6*, 101–105.

Vosshall, L. B., Wong, A. M., & Axel, R. (2000). An olfactory sensory map in the fly brain. *Cell, 102*(2), 147–159.

Wang, Y., Wright, J. D., Guo, H.-F., Zuoping, X., Svoboda, K., Malinow, R., Smith, D. P., & Zhong, Y. (2001). Genetic manipulation of the odor-evoked distributed neural activity in the *Drosophila* mushroom body. *Neuron, 29*, 267–276.

Wang, J. W., Wong, A. M., Flores, J., Vosshall, L. B., & Axel, R. (2003). Two-photon calcium imaging reveals an odor-evoked map of activity in the fly brain. *Cell, 112*, 271–282.

Willshaw, D., Buneman, O. P., & Longuet-Higgins, H. C. (1969). Nonholographic associative memory. *Nature, 222*, 960.

White, J., Dickinson, T. A., Walt, D. R., & Kauer, J. S. (1998). An olfactory neuronal network for vapor recognition in an artificial nose. *Biol. Cyber., 78*, 245-251.

Zars, T., Fischer, M., Schulz, R., & Heisenberg, M. (2000). Localization of a short-term memory in *Drosophila. Science, 288*, 672–675.

Received June 9, 2003; accepted January 30, 2004.