

## Computational Capacity of an Odorant Discriminator: The Linear Separability of Curves

**N. Caticha**

*nestor@if.usp.br*

**J. E. Palo Tejada**

*debianaqp@terra.com*

*Instituto de Física, Universidade de São Paulo, CEP 05315-970, São Paulo, SP, Brazil*

**D. Lancet**

*Doron.Lancet@weizmann.ac.il*

*Department of Membrane Research and Biophysics, Weizmann Institute of Science, Rehovot 76100, Israel*

**E. Domany**

*eytan.domany@weizmann.ac.il*

*Department of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel*

We introduce and study an artificial neural network inspired by the probabilistic receptor affinity distribution model of olfaction. Our system consists of  $N$  sensory neurons whose outputs converge on a single processing linear threshold element. The system's aim is to model discrimination of a single target odorant from a large number  $p$  of background odorants within a range of odorant concentrations. We show that this is possible provided  $p$  does not exceed a critical value  $p_c$  and calculate the critical capacity  $\alpha_c = p_c/N$ . The critical capacity depends on the range of concentrations in which the discrimination is to be accomplished. If the olfactory bulb may be thought of as a collection of such processing elements, each responsible for the discrimination of a single odorant, our study provides a quantitative analysis of the potential computational properties of the olfactory bulb. The mathematical formulation of the problem we consider is one of determining the capacity for linear separability of continuous curves, embedded in a large-dimensional space. This is accomplished here by a numerical study, using a method that signals whether the discrimination task is realizable, together with a finite-size scaling analysis.

### 1 Introduction ---

The basic machinery for olfaction, our ability to smell, is an array of a few hundred different types of sensory neurons. Each of these expresses molecular receptors that belong to a single type. When this small neuronal as-

sembly is exposed to external stimuli, its cooperative response is capable of detecting and recognizing a wide variety of odorants and measuring their concentrations. We use *odorant* to describe any chemically homogeneous substance (ligand) that elicits a response from the olfactory system.

The response of the array of neurons to any particular odorant is determined by the responses of the individual constituent neurons. This response is, however, governed by the extent to which the receptors expressed by the particular neuron bind the odorant, that is, by the affinity  $K$  of the neuron's receptors to the odorant. According to a model proposed by Lancet, Sadvsky, and Seidman (1993), these affinities can be viewed as independent random variables, drawn from a single receptor affinity distribution (RAD), denoted by  $\psi(K)$ . Once a set of affinities (for all odorants and all sensory neurons) has been generated, the response of the entire sensory assembly to any odorant is determined.

This information is transferred from the sensory neurons to the olfactory bulb, onto which the axons of the sensory neurons project. They form synapses within the olfactory glomeruli with secondary neurons (mitral and tufted cells). Each glomerulus appears to receive input from cells that bear the same type of receptor protein. The olfactory bulb also contains a complex network of local lateral connections, which provide for initial cross-talk between the receptor-specific units, mediated by periglomerular and granule interneurons. This integration of the sensory input that takes place in the olfactory bulb forms the first step of the information processing that takes place in the olfactory pathway. Additional processing and integration is believed to take place in the prepyriform cortex, entorhinal cortex, and other higher brain areas.

In this article, we evaluate, on the basis of a very simple model, some of the potential computational characteristics of the olfactory pathway as it performs its initial information processing. Although we do not claim to model any specific central nervous system regions involved in olfactory processing, our analysis may relate to some global characteristics of this pathway. Our simple model for the sensory array and a single processing unit is depicted in Figure 1. Some of the important characteristics of a real olfactory system that are not taken into account by such a simple model are discussed in section 4.

The model we introduce is also interesting from a mathematical point of view. The problem of linear separability (LS) of points in  $N$ -dimensional space has received considerable attention since Schläfli in 1852 (Kinzel, 1998). In the mathematics literature, Cover (1965) studied the problem of LS of independent dichotomies using combinatorial method. In computer science, the perceptron, introduced by Rosenblatt (1962) and analyzed in detail by Minsky and Papert (1969), gave a major boost to the field of neural networks. More recently, by introducing statistical mechanics techniques, Gardner (1988) and Gardner and Derrida (1988) extended Cover's results to cases where there are correlations between the points that have to be linearly separated.

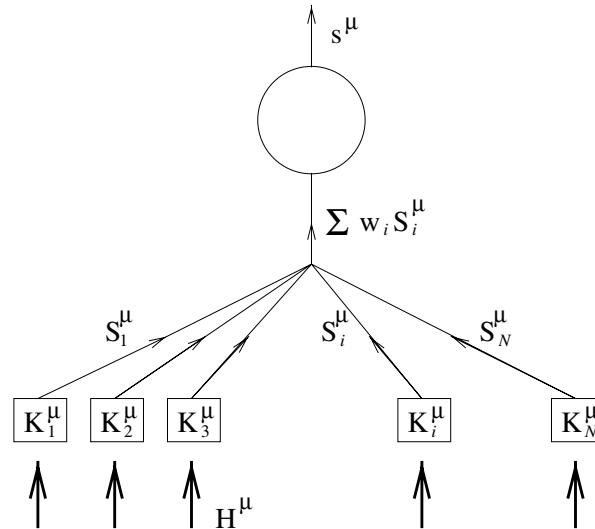


Figure 1: Schematic representation of our model.  $N$  sensory neurons, represented by boxes, are exposed to odorant  $\mu$ , present in concentration  $H^\mu$ . Each sensory neuron  $i$  is characterized by a set of affinities  $K_i^\mu$ ; the sensory input elicits from neuron  $i$  a response  $S_i^\mu$ , as given by equations 2.4 and 2.5. A weighted sum of the  $N$  sensory responses serves as the input of the secondary neuron (circle), whose output  $s^\mu$ , generated in response to odorant  $\mu$ , is given by equation 2.6.

We generalize the problem of separating (zero-dimensional) points to the separability of (one-dimensional) strings or curves, embedded in  $N$ -dimensional space. In the context of our problem, the curves that need to be separated are parameterized continuously by the odorant concentration.

In principle, one can address the separability of curves by placing a discrete set of points on each curve, thereby mapping the problem onto the previously solved one of separating points. One should note, however, that points that lie on the same curve are not independent; in fact, they are correlated in ways that render the previously developed analytical methods unapplicable. Therefore, we present an extensive numerical analysis of the capacity of this special neural network—of  $N$  sensory neurons that provide input to a single processing unit. The capacity we calculate is interpreted as follows. The sensory system is exposed to  $p+1$  odorants, one at a time. One of these is the target; the aim is to distinguish the target from all the other  $p = \alpha N$  odorants that form a noisy olfactory background.

The model, based on a single-layer perceptron, is introduced and discussed in detail in section 2.1. Then we describe the method we have developed in order to determine the capacity numerically. To do this, we adapted and used several different techniques. One of these, a learning algorithm

introduced by Nabutofsky and Domany (1991), is described in section 2.2. This algorithm, like all other perceptron learning rules, finds the separation plane (if the problem is LS); however, unlike other learning algorithms, it provides a rigorous signal to the fact that a sample of examples is *not* LS.

Another technique we adapted to our purposes is finite-size scaling (FSS) analysis of the data. The main results are presented in section 3 as curves of capacity as a function of odorant concentration in the thermodynamic ( $N \rightarrow \infty$ ) limit, obtained by extrapolation, using FSS, from data obtained at a sequence of  $N$  values. This large  $N$  limit is quite natural from both practical and theoretical points of view. In practice, for  $N$  of the order of a few hundred, the results can barely be distinguished numerically from those at the  $N \rightarrow \infty$  limit. As to the theoretical side, the situation in this limit is much cleaner and easier to analyze. The final section contains a critical discussion of the results from a biological point of view.

Our central finding is summarized in Figure 6. If we fix the range of concentrations  $H_{\min} < H < H_{\max}$  in which the system operates and increase the number of background odorants, we will reach a critical number  $p_c$  beyond which the system fails to discriminate the target. This critical number is proportional to the number of sensory neurons  $N$ , that is,  $p_c = \alpha_c N$ , and it decreases when the concentration range increases.

## 2 Computational Model

---

**2.1 Odorant Identification as Linear Separation of Curves in N-Dimensional Space.** The simple neural assembly that is considered here consists of a single secondary neuron that receives inputs from an array of  $N$  units that model the sensory neurons. The single secondary neuron represents a grandmother cell, whose task is to detect one particular target odorant, labeled 0. The sensory scenario we consider allows exposure of the neuronal assembly to a single odorant, which may be either the target odorant or one of  $p = \alpha N$  background odorants. The odorant provides simultaneous stimuli to the  $N$  sensory neurons. The aim of the single secondary neuron is to determine whether the odorant that generated the incoming signal from the sensory array is the target odorant 0. We assume that all odorants, background and target, are presented to the sensory array in concentrations  $H$  that lie within a range

$$H_{\min} < H < H_{\max}. \quad (2.1)$$

We pose the following well-defined quantitative question: What is the maximal number  $p_c$  of different background odorants that our neuron can distinguish from the target, for any concentration within the prescribed range? To sharpen the question, we put it in a more precise mathematical form. Consider  $\mu = 1, 2, \dots, p$  background odorants with respective concentrations  $H^\mu$  in the range 2.1. Odorant  $\mu$  is characterized by the  $i = 1, \dots, N$

affinities  $K_i^\mu$  of the  $N$  receptors. According to the RAD model, these affinities are selected independently from a distribution  $\psi(K)$  (Lancet et al., 1993).

All our numerical results were obtained using for  $\psi(K)$  the form (note:  $K \geq 0$ )

$$\tilde{\psi}(K) = \frac{K}{\sigma^2} \exp\left(-\frac{K^2}{2\sigma^2}\right). \quad (2.2)$$

The average and variance of this distribution are given by

$$\langle K \rangle = \sqrt{\frac{\pi}{2}}\sigma \quad \text{var}(\tilde{\psi}(K)) = 0.65\sigma. \quad (2.3)$$

The distributions suggested by Lancet, et al. (1993) were Poisson and binomial. With regard to the computational limitations of our model, the important idea behind the RAD model lies not in the exact form of the distribution but on the fact that the affinities can be thought of as independent random variables. The main computational features of our model will not be altered as long as the distribution has the following features: it is zero for negative affinities and has finite first and second moments. We have used  $\tilde{\psi}(K)$  since it satisfies the previous constraints and is easier to deal with in analytical calculations.

When receptor  $i$  is exposed to odorant  $\mu$  at concentration  $H^\mu$ , its response is given by

$$S_i^\mu = f(K_i^\mu H^\mu), \quad (2.4)$$

where  $f(x)$  is a sigmoid-shaped function. We use

$$f = x/(1 + x) \quad (2.5)$$

throughout this article. Although there is some evidence for linear scaling of the signals from the glomerular layer with odorant concentrations (Ma & Shepherd, 2000; Wachowiak, Cohen, Zochowski, & Falk, 2000), a general analysis should take into account the inevitable saturation at high concentrations (see equation 2.5), as dictated by the law of mass action (Schild, 1988).

The value taken by the affinity  $K_i^\mu$  sets the particular concentration scale at which odorant  $\mu$  affects the  $i$ th sensory neuron. From this point on, we set

$$\sigma = 1$$

in equation (2.2). This means that the concentrations are measured in inverse units of the parameter  $\sigma$ .

The set of values  $\{S_i^\mu\} = \{S_1^\mu, S_2^\mu, \dots, S_N^\mu\}$  constitutes a vector of signals  $\mathbf{S}^\mu$ , generated by the entire sensory array, when it is exposed to odorant  $\mu$ . The  $\{S_i^\mu\}$  serve as inputs to our secondary neuron, which we model as a linear threshold element or perceptron. Its output signal is given by

$$s^\mu = \text{sign} \left( \sum_i^N w_i S_i^\mu \right) = \text{sign}(\mathbf{w} \cdot \mathbf{S}^\mu). \quad (2.6)$$

The simple neural network described is schematically presented in Figure 1. The sensory neurons are represented by boxes and the secondary neuron by a circle. We require the output of this neuron to differentiate the target odorant from the background, yielding

$$s^\mu = \begin{cases} -1 & \text{for } \mu = 1, \dots, P \text{ (background)} \\ +1 & \text{for } \mu = 0 \text{ (target)} \end{cases} \quad (2.7)$$

for any odorant concentration in the allowed range 2.1.

To understand the geometrical meaning of this requirement, note that when the concentration of odorant  $\mu$  is varied in the allowed range 2.1, the corresponding vector  $\mathbf{S}^\mu$  traces a curve (or string) in the  $N$ -dimensional space of sensory responses. The requirement 2.7 means that there exists a hyperplane, such that the entire curve that corresponds to the target odorant lies on one side of it, while the curves that correspond to all  $p$  background odorants lie on the other side. This explains our statement that the problem we solve deals with the linear separability of curves.

We show that a solution to this classification problem can be found provided  $p < p_{\max} = \alpha_c N$ . We estimate the critical capacity  $\alpha_c$  numerically. This is done by extrapolating results obtained for various values of  $N$  using finite-size scaling techniques, to the limit  $N \rightarrow \infty$ . The value of  $\alpha_c$  is evaluated as a function of the limiting odorant concentrations.

In order to obtain these results using existing methodology, the most natural and straightforward thing to do is to place a discrete set of  $\zeta = 1, 2, \dots, M$  points  $\mathbf{S}^{\mu\zeta}$  on each curve, corresponding to different concentrations, and to require that the  $M$  points that lie on the curve of the target odorant are linearly separable from the  $PM$  points that represent the background. That is, equations 2.7 become

$$s^{\mu\zeta} = \begin{cases} -1 & \text{for } \mu = 1, \dots, P; \zeta = 1, \dots, M \text{ (background)} \\ +1 & \text{for } \mu = 0; \zeta = 1, \dots, M \text{ (target)}. \end{cases} \quad (2.8)$$

This raises the technical question of how many (discrete) representatives of the same odorant should be included in the learning set. We show that while the critical number of odorants  $p_{\max}$  scales linearly with  $N$ , the number of representatives of a single odorant,  $M$ , has to grow at least as fast as  $N^2$ .

This ensures that increasing  $M$  further does not change the results of the calculation (e.g., the value of  $p_{\max}$ ), and hence the  $M$  discrete points indeed correctly represent the continuous curves on which they lie.

Our problem has been turned into one of learning  $M(p + 1)$  patterns that constitute our training set  $\mathcal{L}$ . For technical reasons, it is convenient to introduce and work with normalized patterns,

$$\xi^{\mu\zeta} = \frac{\mathbf{S}^{\mu\zeta} s^{\mu\zeta}}{\sqrt{(\mathbf{S}^{\mu\zeta})^2}}, \quad (2.9)$$

with  $\zeta = 1, \dots, M$  running over the  $M$  discrete concentrations and  $\mu = 0, 1, \dots, p$  over all odorants. Note that we also multiplied each pattern  $\mathbf{S}^{\mu\zeta}$  by its desired output  $s^{\mu\zeta}$ ; after this change of representation, the condition of linear separability, equation 2.8, becomes

$$\text{sign}(\mathbf{w}^* \cdot \xi^{\mu\zeta}) > 0 \text{ for } \mu = 0, 1, 2, \dots, p; \zeta = 1, 2, \dots, M. \quad (2.10)$$

**2.2 The Learning Algorithm.** The question of whether the target odorant can or cannot be distinguished from the background has been reduced to the question, Is there a set of weights  $w_i$ ,  $i = 1, \dots, N$ , for which all  $M(p + 1)$  inequalities 2.10 are satisfied? This problem is of the type studied by Rosenblatt (1962) and is an example of classification by a single-layer perceptron. A solution exists if one can find a weight vector  $\mathbf{w}^*$  (that parameterizes the perceptron) such that for all the patterns  $\xi^{\mu\zeta}$  in the training set  $\mathcal{L}$ , the “field,”

$$h^{\mu\zeta} = \xi^{\mu\zeta} \cdot \mathbf{w}^* > 0, \quad (2.11)$$

that is, the projection of the weight vector  $\mathbf{w}^*$  onto all patterns  $\xi^{\mu\zeta}$ , is positive. We wish to determine the size of the training set  $\mathcal{L}$ —the number of background odorants  $p$  for which a solution  $\mathbf{w}^*$  can be found. This is done by executing a search for a solution  $\mathbf{w}^*$  by means of a learning algorithm. There are several learning algorithms (Rosenblatt, 1962; Abbott and Kepler, 1989) in the literature. All are guaranteed to find such a weight vector in a finite number of steps provided a solution exists. If, however, the problem is not LS and a solution does not exist, most learning algorithms will run ad infinitum. An exception is the algorithm of Nabutovsky and Domany (ND; 1991), which detects, in finite time, that a problem is nonlearnable. This is a batch perceptron learning algorithm, presenting sequentially the entire training set  $\mathcal{L}$  in one sweep and repeating the process until either a solution is found or nonlearnability is established. We found that this algorithm is efficient and convenient to use (see Abbott & Kepler, 1989, for other algorithms that detect non-LS problems).<sup>1</sup>

<sup>1</sup> The fact that we use a learning procedure to establish the boundaries of linear separability does not imply and is unrelated to any possible plasticity of the olfactory bulb. The algorithm is being used only to find a solution or show that it does not exist.

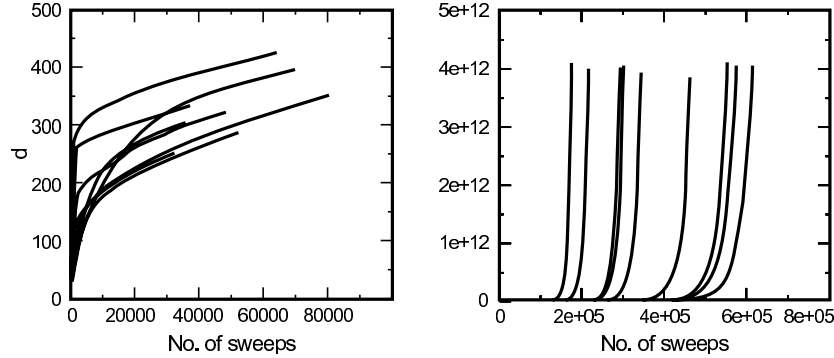


Figure 2: Despair  $d$  as a function of the number of sweeps for  $N = 10$ ,  $H_{\max} = 50$ ,  $H_{\min} = 0.5$ , and  $p = 25$ . (a) Learnable cases. (b) Cases that were not learnable. Note the huge difference in the vertical and horizontal scales.

ND introduced a parameter  $d$ , which they called *despair*, which is calculated in the course of the learning process.  $d$  is bounded if the training set  $\mathcal{L}$  is LS. Since the ND algorithm can be shown either to find a solution  $\mathbf{w}^*$  or transgress the bound for  $d$  in a finite number of learning iterations,  $d$  effectively signals if the learning set  $\mathcal{L}$  fails to be linearly separable. The theorem they proved can be easily extended to the distribution of examples in our problem.<sup>2</sup> We introduced a halting criterion, which is probably more stringent than necessary, since no attempt has been made to determine an optimal lower bound. In Figures 2a and 2b, typical evolutions of the despair are shown for an LS case and a non-LS case, respectively. The behavior of  $d$  is strikingly different in the two cases, showing that  $d$  is a good indicator of learnability. In the learnable cases,  $d$  grows linearly with the number of learning sweeps until a solution is found (and the curves terminate). In the non-LS cases,  $d$  grows exponentially with the number of sweeps and would continue to grow; the process is halted when its value exceeds a known bound that must be satisfied if the problem is LS.

We now describe the ND algorithm used in the simulations. The patterns of the learning set  $\mathcal{L}$  are presented one at a time (one cycle constitutes a sweep). ND have shown that for binary valued patterns ( $\xi_i = \pm 1$ ), that is, patterns on the vertices of a unit hypercube, an upper bound  $d_c$  exists iff the training set is LS. On the other hand, the dynamics is shown to take  $d$  beyond that bound in a finite (linear in  $N$ ) number of iterations unless a solution exists and the algorithm halts. Initialize the process with  $d = 1$ ,  $\mathbf{w} = \xi^1$ .

<sup>2</sup> The original ND algorithm was designed for binary vectors, pointing at the corners of a  $N$ -dimensional hypercube. It can be shown that the theorem can be extended to vectors on the unit sphere.

Go to the next example. If it is correctly classified, do nothing to the current weight vector, and go to the next example. Once a misclassified example  $\xi^{\mu\zeta}$  is found, update the weight vector as well as the parameter  $d$ , according to

$$\mathbf{w}_{new} = \frac{\mathbf{w} + \eta \xi^{\mu\zeta}}{|\mathbf{w} + \eta \xi^{\mu\zeta}|} \quad (2.12)$$

$$d_{new} = \frac{d + \eta}{\sqrt{1 + 2\eta h^{\mu\zeta} + \eta^2}}. \quad (2.13)$$

$\eta$  is not just a learning-rate parameter but an effective modulation function, chosen in order to maximize the increase of the despair as

$$\eta = \frac{1/d - h^{\mu\zeta}}{(1 - h^{\mu\zeta}/d)}. \quad (2.14)$$

The learning dynamics halts if all patterns are correctly classified or, if the value of  $d$  exceeds an upper bound, given by

$$d > d_c = \frac{N^{(N+1)/2}}{2^{N-1}}. \quad (2.15)$$

This is guaranteed to happen in at most  $Nd_c^2$  steps.

### 3 Numerical Experiments

---

Since there is a large number of parameters that are to be varied, we first present a detailed description of the manner in which we deal with them.

There are two random elements in our studies. The first is in the selection of  $M$  concentrations for each odorant, within the range 2.1. The second is the choice of  $K_i^\mu$ , the affinity of receptor  $i$  to odorant  $\mu$ , selected at random from the distribution  $\tilde{\psi}(K)$  of equation 2.2. For every choice of the remaining variables, we generate an ensemble of experiments and average the object we are measuring over these two random elements. We select  $L_a$  times the set of affinities and for each of these perform  $L_c$  times the random selection of concentrations.

The object we wish to estimate numerically is the probability  $P$  that the  $p$  curves described in section 1 are LS. To this end, we place  $M$  points on each curve and measure the corresponding probability  $P(H_{\min}, H_{\max}, p, N; M)$ . As we will see, for large enough values,  $M > M_c$ , this probability becomes independent of  $M$ ; beyond  $M_c$ , the set of  $M$  discrete points represents the corresponding curves faithfully, and hence the limiting value  $P(H_{\min}, H_{\max}, p, N; M > M_c)$  is our estimate for  $P(H_{\min}, H_{\max}, p, N)$ .

Finally, we are interested in this function in the large  $N$  limit, when  $N \rightarrow \infty$  and  $p \rightarrow \infty$ , while  $\alpha = p/N$  is fixed. This limit is obtained by extrapolating our finite  $N$  results using finite-size scaling methods.

Our first task is to determine how  $M_c$  scales with  $N$ , that is, how dense a set of concentrations is to be used so that  $M$  discrete points represent accurately the continuous curves  $S_i^\mu(H^\mu)$  of equation 2.4.

**3.1 Scaling of  $M_c$ .** We choose values for  $N$  (number of receptor cells),  $p$  (number of odorants), and  $H_{\min}, H_{\max}$  (limiting concentrations). We also set some value for  $M$ , the number of concentrations by which every odorant is represented ( $M$  will be varied).

We proceeded according to the following steps:

1. Draw from the distribution  $\tilde{\psi}(K)$  a set of affinities  $K_i^\mu$  for all  $N$  receptors and  $p$  odorants.
2. Generate for each odorant  $M$  concentration values from a uniform distribution in the allowed range  $H_{\min} < H < H_{\max}$ , and construct the set  $\{\xi^{\mu\zeta}\}$  of normalized patterns.
3. Run the ND learning algorithm until it stops. Register whether the set  $\{\xi^{\mu\zeta}\}$  was LS.

Steps 2 and 3 are repeated  $L_c$  times for each set of affinities; the three-step process is repeated for  $L_a$  different sets of affinities. We used  $L_a = 100$ ; increasing it further made no difference. With such a value of  $L_a$ , the results did not depend on  $L_c$ ; having tried  $1 \leq L_c \leq 10$ , we used  $L_c = 1$  in our simulations.

At this point, we have  $L_c \cdot L_a$  experiments, out of which a fraction of  $P(H_{\min}, H_{\max}, p, N, M)$  cases were linearly separable. Keeping  $H_{\min}, H_{\max}, p, N$  fixed, we increase  $M$  and repeat the entire process, obtaining the probability functions  $P(H_{\min}, H_{\max}, p, N, M)$ , which are plotted in Figure 3 versus  $M/N^2$ . Clearly the curves saturate when  $M > M_c \approx N^2$ . From this point on, we have fixed the value of  $M$  at  $M = N^2$ . This numerical result can be estimated by using the analysis of Gardner and Derrida (1988) for the capacity of biased patterns, using for the ‘‘magnetization’’  $m$  the value  $m \propto 1/N$ . This gives, in addition to the leading behavior  $M_c \approx N^2$ , logarithmic corrections as well. We cannot rule out the possibility of such logarithmic corrections to the scaling we found here.

**3.2 Measuring the Probabilities  $P(H_{\max}, p, N)$ .** In all our experiments, we fixed the value of  $H_{\min} = 0.5$ , and hence the dependence of the probability on this variable has been suppressed. This choice is justified by the fact that in odorant concentrations near and above half maximal saturation ( $H = 1$ ), nonlinearity becomes significant. This makes the discrimination task more challenging for a given dynamic range ( $H_{\max}/H_{\min}$ ). For vari-

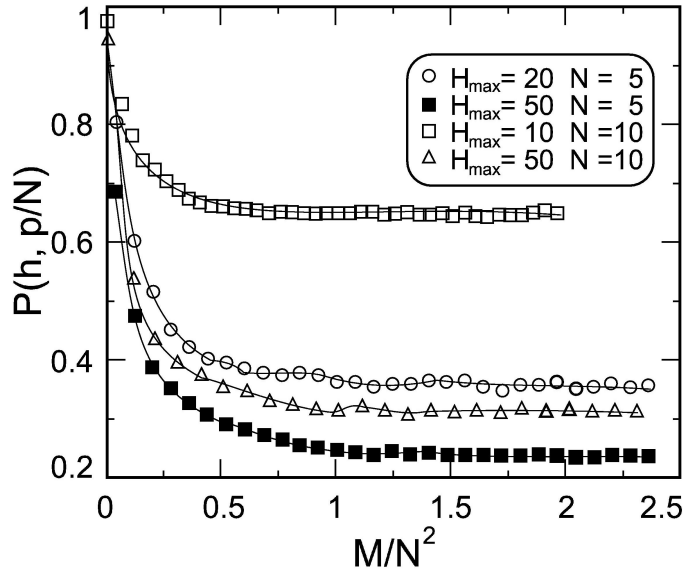


Figure 3: The estimated probabilities saturate when  $M$ , the number of representatives of a given odorant, exceeds  $M_c \approx N^2$ . For brevity, we denote  $P(H_{\min}, H_{\max}, p, N)$  by  $P(h, p/N)$ . In all cases,  $p = 25$  odorants were used, with  $H_{\min} = 0.5$ .

ous values of  $N$ ,  $p$ , and  $H_{\max}$ , we calculate  $P(H_{\max}, p, N)$  in the manner described above. Keeping  $N$  and  $H_{\max}$  fixed, we increase  $p$ . For  $p \ll N$ , we have  $P(H_{\max}, p, N) \approx 1$ , and the probability of LS decreases as  $p$  increases. We stop increasing  $p$  when  $P(H_{\max}, p, N)$  becomes smaller than some  $\varepsilon$ .

The variation of  $P(H_{\max}, p, N)$  versus  $p/N$  is presented, for three values of  $H_{\max}$  and four values of  $N$ , in Figure 4. The results presented in these figures are discussed in the next section.

We should mention here that for large  $N$ , we used a heuristic modification of the ND halting criterion to label a problem as non-LS. Typical evolutions of the despair parameter are shown in Figure 2. Each curve represents the history for a single learning set. Notice the huge difference in scales for the learnable and the unlearnable cases. The wide separation in final values of  $d$  suggests that a more practical (e.g., smaller) upper bound be used. For  $N = 30$  (the largest value treated here), we used a different halting criterion in order to escape from the need to reach an exponentially high upper bound. After a small number of successful trial runs, which produced linear separability, we identified the highest value of the despair  $d_m$  that was reached for a learnable set. This value was used to define our new heuristic halting criterion,  $d_{bound} = N d_m$ .

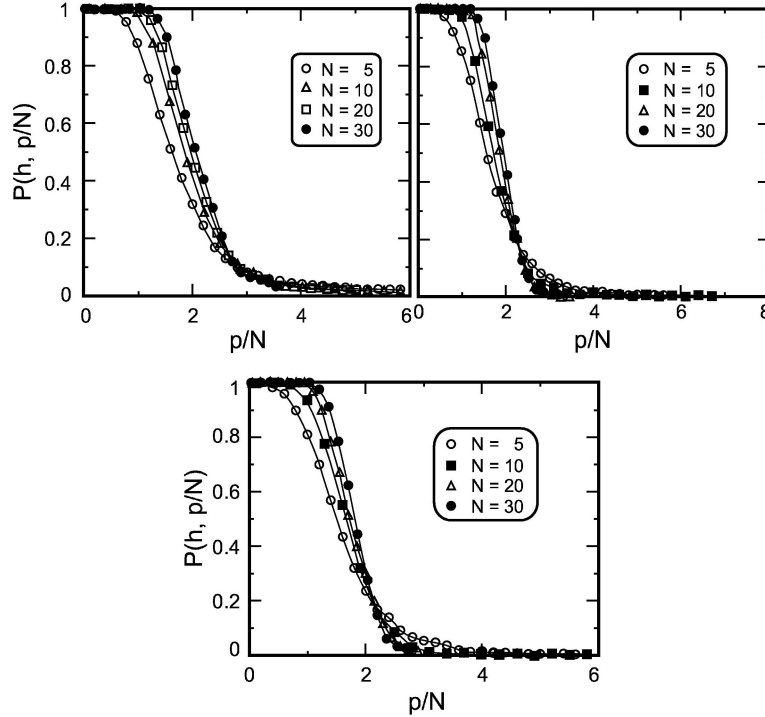


Figure 4: (Top left) Probability that a learning set is LS as a function of  $\alpha = p/N$ , for maximal concentration  $H_{\max} = 50$  and  $H_{\min} = 0.5$ . (Top right)  $H_{\max} = 70$  and  $H_{\min} = 0.5$ . (Bottom)  $H_{\max} = 100$  and  $H_{\min} = 0.5$

**3.3 Finite-Size Scaling Analysis.** As expected, for small  $\alpha = p/N$ , the probability for linear separability is close to 1, and it decreases as  $\alpha$  increases. The curves obtained for fixed  $H_{\max}$  become sharper as  $N$  increases.

Note that curves obtained for different  $N$  values cross at approximately the same value of  $\alpha$ . Similar behavior of the corresponding probability functions has been observed for random uncorrelated patterns (Cover, 1965). Notice, however, that the crossing point is at some probability  $P < 1/2$ . Similar curves obtained for other architectures, such as the parity and committee machines (Nadler & Fink, 1997), crossed at  $P > 1/2$ . If there is a sharp transition in the thermodynamic limit ( $N \rightarrow \infty$ ), these curves should approach a step function, with

$$P(H_{\min}, H_{\max}, \alpha) = \begin{cases} 1 & \alpha < \alpha_c(H_{\min}, H_{\max}) \\ 0 & \text{otherwise} \end{cases}. \quad (3.1)$$

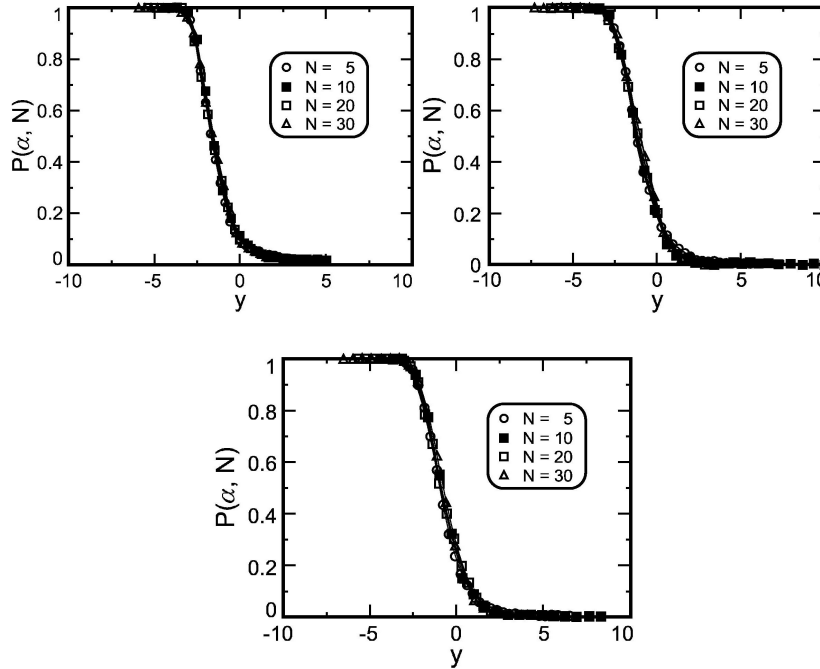


Figure 5: (Top left) Probability that a learning set is LS as a function of  $y = (\alpha - \alpha_c)N^{\frac{1}{\nu}}$ , for maximum concentration  $H_{\max} = 50$  and  $H_{\min} = 0.5$ . (Top right)  $H_{\max} = 70$  and  $H_{\min} = 0.5$ . (Bottom)  $H_{\max} = 100$  and  $H_{\min} = 0.5$ . The result of a least-squares fit are (top left)  $H_{\max} = 50$ ,  $\nu = 4.8$ ,  $\alpha_c = 2.8$ ; (top right)  $H_{\max} = 70$ ,  $\nu = 2.85$ ,  $\alpha_c = 2.25$ ; (bottom)  $H_{\max} = 100$ ,  $\nu = 2.75$ ,  $\alpha_c = 2.1$ .

That is, for  $\alpha$  below a certain  $\alpha_c(H_{\min}, H_{\max})$ , a learning set will be LS with probability one, and conversely, it will be LS with probability zero for  $\alpha > \alpha_c(H_{\min}, H_{\max})$ . The manner in which such a step function is approached as  $N \rightarrow \infty$  can be described by a finite-size scaling analysis (Privman, 1990).

For each value of  $H_{\max}$  (keeping  $H_{\min}$  fixed), we tried a simple rescaling of the  $\alpha$  variable, with two adjustable parameters,  $\alpha_c$  and  $\nu$ :

$$y = (\alpha - \alpha_c)N^{\frac{1}{\nu}}.$$

For the proper choice of  $\alpha_c$  and  $\nu$ , we expect data collapse; that is, curves obtained for different values of  $N$  are expected to fall onto a single function, provided  $P(H_{\max}, \alpha, N)$  is plotted versus the scaled variable  $y$ . As can be seen in Figure 5, this expectation is borne out; the evidently good data collapse indeed substantiates the idea of a sharp transition at  $\alpha_c$ . As  $N$  increases, the function  $P(H_{\max}, \alpha, N)$  becomes increasingly sharper; its width near  $\alpha_c$  decreases at a rate governed by the exponent  $\nu$ . Finally, we present in Figure 6

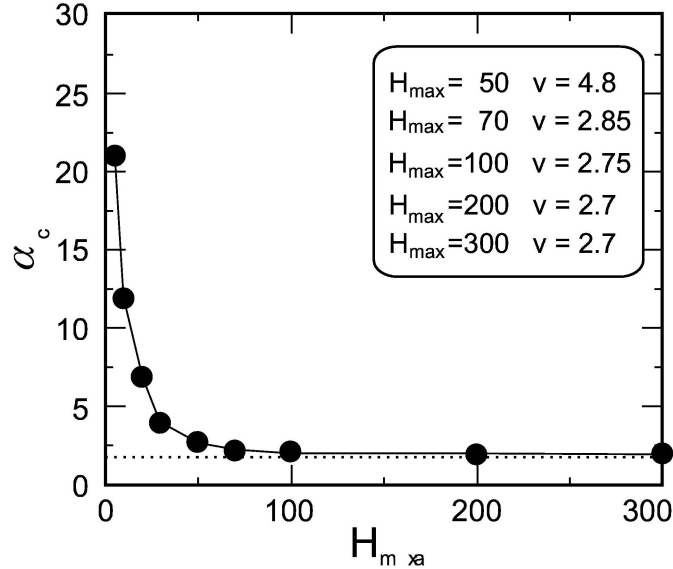


Figure 6: Phase diagram in the  $\alpha$ ,  $H_{\max}$  plane for  $H_{\min} = 0.5$ . The curve separates the  $\alpha$ ,  $H_{\max}$  plane into the regions where the network can distinguish the target odorant (below) and where it cannot (above). Below (above) the curve a learning set is (not) LS with probability one in the thermodynamic limit. The horizontal dotted line is  $\alpha_c = 2$ .

the behavior of  $\alpha_c$  as a function of  $H_{\max}$  (for fixed  $H_{\min}$ ). As  $H_{\max}$  increases, separation of the curves becomes an increasingly difficult task, and hence  $\alpha_c(H_{\max})$  decreases. We find that it saturates at a low value close to  $\alpha_{\min} = 2$ , which is exactly the Cover result. This interesting point is explained in the appendix.

Although we deal here with linear separability of curves, which one would expect to be a more difficult task than separating points, we found that our  $\alpha_c$  exceeds the value derived for points,  $\alpha_c = 2$ . The reason is that this is the critical capacity for separating random, independent points; the curves we are trying to separate are not independent of each other. In fact, by construction, we have  $S_i^{\mu} > 0$  for all the background odors; hence, all these curves lie on one side of an entire family of planes. The target odorant, which also satisfies  $S_i > 0$ , should lie on the other side of the separating plane.

The curve  $\alpha_c(H_{\max})$  is, in effect, a phase boundary; on one side, we have a “phase” in which the problem is LS, while on the other (high  $\alpha$ ) region, it is not. We now present a brief description of the manner in which linear separability breaks down as we cross this phase boundary by increasing  $H_{\max}$  at fixed  $\alpha$ .

**3.4 Breakdown of LS Near-Phase Boundary.** The manner in which LS breaks down as  $H_{\max}$  increases beyond the phase boundary is nicely illustrated by Figures 7 and 8.

Consider the  $p$  curves, in an  $N$ -dimensional space, which represent the odorants, in a linearly separable case. We present in Figure 7 (left) a projection of these curves onto a randomly chosen plane. One of these (indicated by an arrow) is the target odorant; it seems to be entangled with the other curves. The point at which all curves seem to converge corresponds to the maximal concentration. The purpose of the learning dynamics is to find a particular direction  $\mathbf{w}$  along which one is able to separate the target curve from the others. Denote by  $\mathcal{V}$  the hyperplane that passes through the origin and is perpendicular to  $\mathbf{w}$ ; this is the linear manifold that separates the target from all the background curves. Select now any plane  $\mathcal{Q}$  that contains  $\mathbf{w}$ , and project all curves onto  $\mathcal{Q}$ ; this produces Figure 7 (right). The horizontal dotted line shown here is the intersection of the hyperplane  $\mathcal{V}$  with the plane  $\mathcal{Q}$ . The projected background odorant curves lie on one side of this line and the target on the other. The situation depicted here is LS.

Consider now what happens when we turn the problem into non-LS by increasing  $H_{\max}$  beyond the phase boundary. As we increase the maximal concentration, the target odorant's curve penetrates the "wrong" side of hyperplane  $\mathcal{V}$ . A picture of this situation is shown in Figure 8 (left). This is a non-LS problem, which means that no matter how long we run our learning algorithm, we will never find a hyperplane  $\mathcal{V}$  that separates the target from

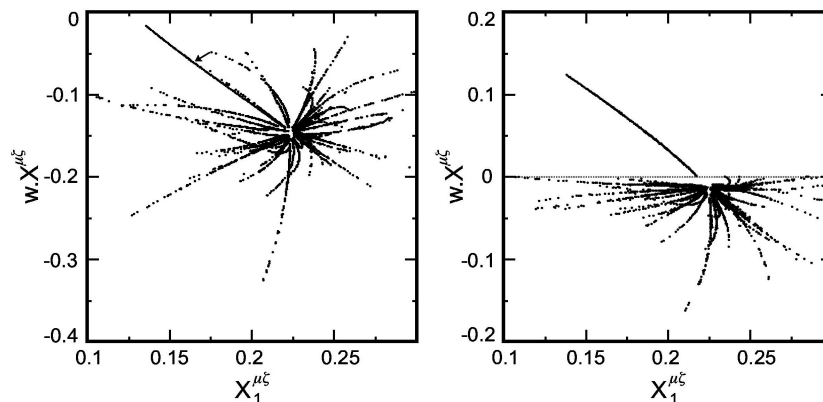


Figure 7: The curves representing the  $p$  background odorants and the target odorant (marked by an arrow) in an LS case,  $\alpha < \alpha_c$ , for  $H_{\max} = 50$ ,  $H_{\min} = 0.5$ ,  $N = 20$ . (Left) The curves are projected onto a randomly selected plane and (right) onto a plane  $\mathcal{Q}$  (see text) that contains the weight vector  $\mathbf{w}^*$ , determined by the learning algorithm. The horizontal dotted line in the right part of the figure demonstrates linear separability of the target from background.

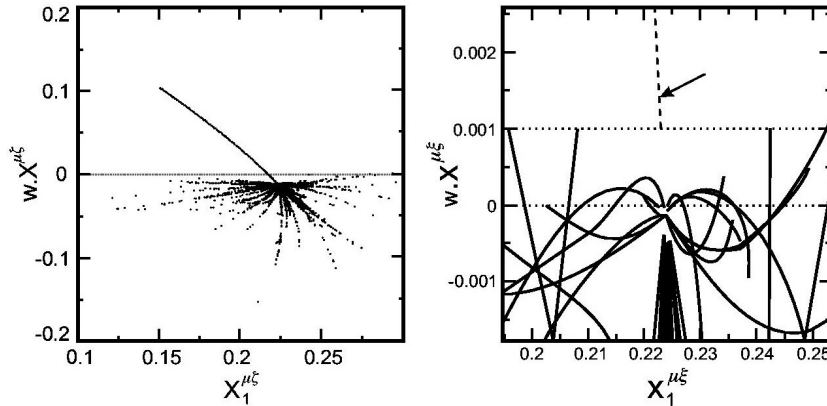


Figure 8: The same situation as in Figure 7, but with  $H_{\max}$  increased beyond the limit of learnability. (Left) Projecting the curves onto the same plane as for the LS case, we see that the target penetrates to the “wrong” side of the broken line. (Right) Further attempts to learn a new  $\mathbf{w}$  will fail.

all the background. If nevertheless we keep running our learning algorithm, the direction of our candidate for  $\mathbf{w}$  will keep changing as we “learn,” but since the critical capacity curve of Figure 6 has been crossed, no amount of further learning will produce a separating plane. The density of points near the high concentration limit is much larger than for low concentrations. Hence, further learning will perhaps be able to separate the target from the background at high concentrations, but then separability breaks down at low concentrations (see Figure 8, right).

#### 4 Summary and Discussion

In the olfactory bulb of most vertebrates, each secondary neuron (mitral or tufted cell) receives input from only one glomerulus, which is innervated, in all likelihood, by axons stemming from olfactory epithelial sensory cells that express the same olfactory receptor protein. Thus, the grandmother cell modeled here may not simply represent a mitral or tufted cell. Moreover, the periglomerular and granule cells constitute a network of recurrent local connections, which are not taken into account in the present (purely feedforward) model.<sup>3</sup> Our analysis may be relevant, in an abstract fashion, to the information processing that takes place at higher olfactory system centers, where broader convergence patterns may exist (Haberly, 2001). We do not

<sup>3</sup> Lack of recurrent connections imposes simple dynamics, excluding observed oscillatory behavior (Dorries & Kauer, 2000; Lam, Cohen, Wachowiek, & Zochowski, 2000).

claim that any neural elements in this pathway may actually be represented by grandmother cells with linear feedforward inputs. However, we have introduced a basic “minimal” model that allows a quantitative estimation of olfactory discrimination capacity.

Several studies analyze neuronal networks for the olfactory system (Hopfield, 1991; Li & Hopfield, 1989; Wilson & Bower, 1992; Li, 1990, 1995; Li & Hertz, 2000). However, none of these was based on a quantitative model for the affinity relationships within the entire olfactory receptor repertoire. Here, we use the receptor affinity distribution (RAD) model, which was developed, based on general biochemical considerations, for receptor repertoires, including that of olfactory receptors. The power of this approach is in using a global knowledge about the repertoire to analyze the fidelity of discrimination among odorants. It has been pointed out in the past that the RAD model may be used to analyze the signal-to-noise ratio in systems in which specific binding to a receptor has to be distinguished from the background of numerous other receptors that constitute nonspecific binding (Lancet, Sadovsky, & Seidman, 1993; Lancet, Horovitz, & Katchalski-Katzir, 1994). Here, we apply a similar concept to an analysis of signal-to-noise discrimination in the case of a neuronal network whose input stems from a receptor repertoire.

The results presented here suggest that for a fixed number of background odorants, there is a maximal odorant concentration beyond which odorant discrimination becomes impossible. This is not surprising, since olfactory receptors are saturable, and at very high concentrations weak affinity receptors as well as high affinity ones will generate comparable signals. However, it is noteworthy that despite the fact that information capacity for odorant discrimination rapidly declines as odorant concentration goes up, the network we analyzed is still capable of discrimination even at concentrations for which  $H\langle K \rangle$  is of the order of a few hundred (where  $\langle K \rangle$  is the average affinity).

The model network consists of  $N$  sensory neurons, each characterized by a set of affinities to a number of odorants. When any particular odorant,  $\mu$ , is present, sensory neuron  $i$  produces a (nonlinear) response,  $S_i^\mu$ . These responses constitute the inputs to a single processing unit (secondary neuron), which performs weighted summation of all the  $N$  inputs. The secondary neuron's output is the sign of this weighted sum. The aim of this single processing unit is to identify one single odorant and separate it from all the others that may be sensed by the system. This secondary neuron plays the role of a grandmother cell for a particular target odorant. An assembly of  $P_0$  such secondary neurons may constitute, together with the sensory neurons, a system that is able to identify the presence of  $P_0$  target odorants from a background of  $P$  odorants.

We posed a well-defined quantitative question: Given that each odorant may appear with a concentration  $H_\mu$  that lies in a certain range,  $H_{\min} < H^\mu < H_{\max}$ , what is the maximal number of background odorants  $P_c = \alpha_c N$

from which a single target can be separated with probability 1? The answer is summarized in Figure 5, where  $\alpha_c$ , the critical capacity, is plotted versus  $H_{\max}$ . The result is obtained in the limit of large  $N$  (i.e., many sensory neurons; in fact, for  $N = 100$ , this result should already give excellent precision). For a dynamic range of  $H_{\max}/H_{\min}$  of about 100, we find  $\alpha_c \approx 2.5$ . That is, for, say,  $N = 300$  sensory neurons, we can distinguish the target from about 750 background odorants. Hence, if we assemble 750 odorants and appoint a grandmother cell for each, we will be able to identify them one by one.

In order to get this quantitative answer, we had to generalize an old problem of linear separability of  $P$  points on an  $N - 1$  dimensional hypersphere to the new problem of linearly separating  $P$  curves that lie on the same hypersphere. We have shown that in order to represent a curve by discrete points that lie on it, we have to place  $M \propto N^2$  points on each curve. The results were obtained by a perceptron learning algorithm that signals when a problem is unlearnable (i.e., nonlinearly separable).

Results obtained at various values of  $N$  were shown to collapse when plotted as functions of properly defined scaled variables, which allowed easy extrapolation to large values of  $N$ .

## Appendix

---

The behavior of the phase boundary for large concentrations  $\alpha_c \rightarrow 2$  (see Figure 5) is quite surprising since the network may be expected to enter a totally confused state due to the saturation of the nonlinear sensory neurons. This could be expected to lead instead to  $\alpha_c \rightarrow 0$ . That the Cover result ( $\alpha_c = 2$ ) is recovered in the high concentration regime can be in fact be understood by the following argument.

We first calculate the probability  $P(S)$  that a sensory unit gives a response  $S$  to the presentation of an odorant in the range 2.1 by

$$P(S) = \langle \delta(S - f(HK)) \rangle, \quad (\text{A.1})$$

where the average is taken over possible concentrations  $H$  uniformly distributed in range 2.1 and according to the RAD model, over the affinities,  $\tilde{\psi}(K)$  of equation 2.2.  $f$  is given by equation 2.6. The integrals lead to

$$P(S) = \frac{\sqrt{\pi}(1-S)^{-2}}{\sigma(H_{\max} - H_{\min})} \left( \text{Erfc} \left( \frac{S}{\sigma H_{\max}(1-S)} \right) - \text{Erfc} \left( \frac{S}{\sigma H_{\min}(1-S)} \right) \right), \quad (\text{A.2})$$

where  $\text{Erfc}(x) = \int_0^x \exp(-u^2) du / \sqrt{\pi}$  is the complementary error function. This probability has one peak that sharpens and moves to higher values of  $S$  as  $H_{\max}$  grows. However, at the very ends of the interval,  $S = 1$  or  $0$ ,

the probability is zero. That  $P(1) = 0$  for every  $H_{\max}$  is the source of the surprise. The peak that concentrates all the probability gets arbitrarily close to  $S = 1$  as the concentration increases, but never makes it to the extreme of the interval. In fact,  $S_{\text{peak}} \approx 1 - c/H_{\max}$ . Therefore, the components of the vectors  $\mathbf{S}^{\mu\zeta}$  will be with overwhelming probability at the peak position, which can be written as  $S_i^\mu = 1 - \varepsilon_i^\mu$  with all  $\varepsilon_i^\mu(H_{\max}, K_i^\mu)$  small but strictly positive. Neglecting second-order terms in  $\varepsilon$  the normalized patterns will then be:

$$\begin{aligned}\tilde{S}_i^\mu &\equiv \frac{S_i^\mu}{|\mathbf{S}^\mu|} = \frac{(1 - \varepsilon_i^\mu)}{\sqrt{N}(1 - 2\bar{\varepsilon}^\mu)^{\frac{1}{2}}} \\ &= (1 - \varepsilon_i^\mu + \bar{\varepsilon}^\mu)/\sqrt{N}.\end{aligned}$$

Therefore, the  $\tilde{\mathbf{S}}^{\mu\zeta}$  vectors are unbiasedly distributed around  $(1, 1, \dots, 1)/\sqrt{N}$ . We are taken back to the original Cover-Gardner problem of separating  $p$  unbiased patterns with a hyperplane, and the result  $\alpha_c = 2$  is no longer a surprise. This argument does not deal with the asymptotic behavior of the capacity in the presence of any kind of noise. In that case, the naive expectations that  $\alpha_c \rightarrow 0$  for  $H_{\max} \rightarrow \infty$  are probably borne out.

### Acknowledgments

---

We thank Ido Kanter for most useful discussions. The research of E.D. was supported by the Germany-Israel Science Foundation, the Minerva Foundation, and the U.S.-Israel Binational Science Foundation. The work reported here was initiated during visits of N.C. to the Weizmann Institute, which were supported by grants from the São Paulo Society of Friends of the Weizmann Institute and the Gorodesky Foundation. J.E.P.T.'s research was supported by a graduate fellowship of the Fundação de Amparo à Pesquisa do Estado de São Paulo. N.C. received partial support from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

### References

---

- Abbott, L. F., & Kepler, T. B. (1989). Optimal learning in neural network memories. *Journal of Physics A, Math. and Gen.*, 22(14), L711–717.
- Cover, T. (1965). Geometric and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Tran. Elect. Comput.*, 14, 326.
- Dorries, K. M., & Kauer, J. S. (2000). Relationships between odor-elicited oscillations in the salamander olfactory epithelium and olfactory bulb. *J. Neurophysiology*, 83(2), 754–765.
- Gardner, E. (1988). The space of interactions in neural network models. *Journal of Physics A, Math. and Gen.*, 21, 257.

- Gardner, E., & Derrida, B. (1988). Optimal storage properties of neural network models. *Journal of Physics A, Math. and Gen.*, *21*, 271.
- Haberly, L. B. (2001). Parallel-distributed processing in olfactory cortex: New insights from morphological and physiological analysis of neuronal circuitry. *Chem. Senses*, *26*(5), 551–576.
- Hopfield, J. J. (1991). Olfactory computation and object perception. *Proceedings of the National Academy of Sciences USA*, *88*, 6462.
- Kinzel, W. (1998). Phase transitions of neural networks. *Phil. Mag.* **B**, *77*, 1455.
- Lam, Y. W., Cohen, L. B., Wachowiak, M., & Zochowski, M. R. (2000). Odors elicit three different oscillations in the turtle olfactory bulb. *J. Neuroscience*, *20*(2), 749–762.
- Lancet, D., Horovitz, A., & Katchalski-Katzir, E. (1994). Models for analysis of protein-ligand interactions. In J. P. Behr (Ed.), *Molecular recognition in biology* (pp. 25–71). New York: Wiley.
- Lancet, D., Sadovsky, E., & Seidman, E. (1993). Probability model for molecular recognition in biological receptor repertoires—significance to the olfaction system. *Proceedings of the National Academy of Sciences USA*, *90*, 3715.
- Li, Z. (1990). A model of olfactory adaptation and sensitivity enhancement in the olfactory bulb. *Biol. Cybern.*, *62*, 349.
- Li, Z. (1995). Modeling the sensory computations of the olfactory bulb. In E. Domany, J. L. van Hemmen, & K. Schulten (Eds.), *Models of neural networks*. New York: Springer-Verlag.
- Li, Z., & Hertz, J. (2000). Odour recognition and segmentation by a model olfactory bulb and cortex. *Network: Computation in Neural Systems*, *11*, 83.
- Li, Z., & Hopfield, J. J. (1989). Modeling the olfactory bulb and its neural oscillatory processings. *Biol. Cybern.*, *61*, 379.
- Ma, M., & Shepherd, G. M. (2000). Functional mosaic organization of mouse olfactory receptor neurons. *Proceedings of the National Academy of Sciences USA*, *97*(23), 12869–12874.
- Minsky, M., & Papert, S. A. (1969). *Perceptrons*. Cambridge, MA: MIT Press.
- Nabutovsky, D., & Domany, E. (1991). Learning the unlearnable. *Neural Computation*, *3*, 604.
- Nadler, W., & Fink, W. (1997). Finite size scaling in neural networks. *Physical Review Letters*, *78*, 555.
- Privman, V. (Ed.). (1990). *Finite-size scaling and numerical simulations of statistical systems*. Singapore: World Scientific.
- Rosenblatt, F. (1962). *Principles of neurodynamics*. New York: Spartan Books.
- Schild, D. (1988). Principles of odor coding and a neural network for odor discrimination. *Biophysica Journal*, *54*(6), 1011.
- Wachowiak, M., Cohen, L. B., Zochowski, M., & Falk, C. X. (2000). The spatial representation of odors by olfactory receptor neuron input to the olfactory bulb is concentration invariant. *Biological Bull.*, *199*(2), 162–163.
- Wilson, M. A., & Bower, J. M. (1992). Cortical oscillations and temporal interactions in a computer-simulation of piriform cortex. *J. Neurophysiol.*, *67*, 981.