

# The influence of tuning width and dynamic range on the acuity of population codes

M. Bethge  
D. Rotermund  
K. Pawelzik

Institute of Theoretical Physics, University Bremen, Kufsteiner Str., D-28334 Germany

The question whether and how the acuity of visual perception can be related to the receptive field size of sensory neurons has triggered several theoretical studies that seek to determine optimal tuning widths with respect to coding efficiency. While an early result by Hinton for binary tuning functions suggests that a small tuning width is never advantageous, a more recent work claims that sharp tuning is optimal if the stimulus space is of dimension  $D = 1$ . Here we solve this paradox by separating the width from the smoothness of a tuning function. In particular, we find that broad tuning with small dynamic range is optimal for all  $D$  under the usual assumptions. On the other hand, broad tuning is not compatible with energy constraints and for the efficiency of visual representations the tuning width turns out to be relevant only in case of rather artificial dot stimuli. In contrast, a small dynamic range is advantageous for arbitrary stimuli whenever the time scale of decoding is small.

## 1 Introduction

The relevance of population coding has been impressively demonstrated in [9] for neuronal representations of saccadic eye movements in the deep layers of the superior colliculus. By reversible inactivation of small subsets of collicular neurons they showed that it is not the most active cell that exclusively controls the eye movement (as it has been hypothesized before), but rather the average preferred direction weighted by the activities of a large population of coarsely tuned neurons. More generally, population coding is suspected to play an important role in many systems due to the frequent finding “*that the accuracy with which primates are able to perform perceptual or motor tasks is much better than expected from the tuning width of single cells that are presumed to be involved in these tasks*” [12].

For population codes, the tuning width  $w$  clearly does not provide a lower bound for spatial resolution and an early theoretical study by Hinton [7, 8] demonstrates even a superiority of coarse coding: for *binary* radial symmetric tuning functions distributed uniformly over a  $D$  dimensional stimulus space and *vanishing neuronal response variability*, the minimal decoding error scales according to

$$\sigma \propto w^{1-D} \quad . \quad (1)$$

This means that sharp and broad tuning are equally good in case of  $D = 1$ , while broad tuning is optimal for all  $D \geq 2$ .

Subsequently several theoretical studies investigated the effect of the tuning width on the acuity of population codes. Eurich & Schweigler [6] confirmed the result given by Eq. 1 analyzing the

same model as in [7] apart from the infinitely sized stimulus space, which was replaced with a  $D$  dimensional sphere. Another very different argument for the advantage of coarse coding was presented by Baldi & Heiligenberg [2], who considered the approximation error of a particular radial basis function network.

On the other hand Snippe & Koenderink [11] argued against the square nature of the sensitivity profiles used in [7] and through their analysis using Gaussian receptive fields they found that coarse coding is not always optimal, but sharp tuning is better in the case of  $D = 1$ . While this finding is also in contradiction to the result of [2], which was based on Gaussian tuning curves, too, they claimed that their analysis was more accurate. More recently, the conclusion of [11] was supported by a work of Zhang & Sejnowski [13], who derived an equivalent scaling rule on the basis of Fisher information

$$\sigma^2 \propto w_Z^{2-D}, \quad (2)$$

which they claimed to be “universal”. In particular, they concluded for *any noise model* and *all* radial symmetric tuning functions distributed uniformly over a  $D$  dimensional stimulus space that sharp tuning is optimal in case of  $D = 1$ , sharp and broad tuning are equally good in case of  $D = 2$ , and broad tuning is optimal in case of  $D \geq 3$ .

Obviously, however, this “universal scaling rule” does not account for the particular case studied by Hinton. Since Fisher information is particularly relevant in case of small noise [4], the absence of noise cannot explain this contradiction. In fact, we will see below that for  $D = 1$  a small tuning width becomes even worse if noise is increased. Furthermore, we will show that tuning functions exist, which are not binary, but exhibit the same scaling as predicted by Hinton’s model. Hence, the latter does not constitute an irrelevant exception, but a universal scaling rule cannot match Eq. 2.

Taken together, the literature on optimal tuning width does not provide a coherent picture today. In particular it lacks a thorough discussion of the assumptions used leading to divergent conclusions. In the following section, we present a new scaling rule, according to which Eq. 1 and Eq. 2 can be understood as special cases depending on a particular choice for the dynamic range of the tuning functions. In section 3, we investigate how conclusions about the optimal tuning width change by refining the model with respect to the demand on the visual system to represent high-dimensional data. While this question has already been addressed in [14], we here present a new approach to this problem, which does not rely on assumptions about which stimulus parameters are actually encoded. Finally, in section 4 we illustrate why a small dynamic range or even binary tuning [5, 3] is likely to be advantageous for sensory coding.

## 2 What determines the optimal tuning width?

The contradiction between both scaling rules, Eq. 1 and Eq. 2, can be resolved by recognizing that the ansatz for the variation of the tuning width used in [13] implies a scaling of the dynamic range of the tuning function at the same time. In order to decouple the tuning width from the dynamic range of a tuning it is necessary to use a slightly different ansatz, which we will first demonstrate by a simple example, where the tuning function of each single neuron  $f(\vec{x}) = f_{max}\phi(|\vec{x} - \vec{c}|)$  also depends only on the Euclidean distance to the center  $\vec{c}$

$$\phi(z) = \begin{cases} 1 & , \quad z < a \\ 1 - \frac{z-a}{b-a} & , \quad a < z < b \\ 0 & , \quad z > b \end{cases} \quad (3)$$

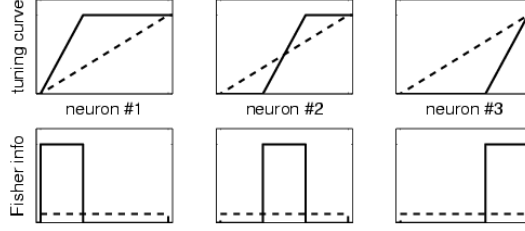


Figure 1: Why a small dynamic range increases total Fisher information. Coding scheme 1 (solid) with small dynamic ranges is compared with coding scheme 2 (dashed) with large dynamic ranges. The corresponding tuning curves are shown in the upper panel and the resulting Fisher information for each tuning function in case of additive Gaussian noise is shown in the lower panel. The total Fisher information of scheme 1 is three times larger than the total Fisher information of scheme 2

but now can be adjusted by two independent parameters  $a$  and  $b$  instead of a single scaling parameter only. Following [13], we assume an (improper) uniform distribution of the tuning function centers  $\vec{c}$  over the entire stimulus space by which the average tuning function array becomes isotropic and one may consider the reconstruction error w.r.t. an arbitrary direction, say  $\vec{e}_1$ , only. Furthermore, it suffices to consider the tuning function at  $\vec{c} = 0$ . In case of additive Gaussian noise with variance  $v$  the corresponding Fisher information component  $J_1$  yields

$$J_1[f(\vec{x})] = f_{max}^2 \begin{cases} 0 & , \quad |\vec{x}| < a \\ \frac{1}{v} \cdot \frac{1}{(b-a)^2} \left( \frac{x_1}{|\vec{x}|} \right)^2 & , \quad a < |\vec{x}| < b \\ 0 & , \quad |\vec{x}| > b \end{cases} . \quad (4)$$

The total Fisher information in the uniform density approximation as used in [13] is proportional to the average over  $J_1$ :

$$\bar{J} = \int J_1[f(\vec{x})] d\vec{x} . \quad (5)$$

In case of  $D = 1$  we obtain

$$\bar{J} = 2 \cdot \frac{f_{max}^2}{v} \cdot \frac{1}{b-a} = 2 \cdot \frac{f_{max}^2}{v} \cdot \frac{1}{d} , \quad (6)$$

where we defined  $d := b - a$  as the dynamic range (i.e. the length of the region with positive Fisher information). In case of  $D = 2$  we obtain

$$\bar{J} = \frac{\pi}{2} \cdot \frac{f_{max}^2}{v} \cdot \frac{b+a}{b-a} = \frac{\pi}{2} \cdot \frac{f_{max}^2}{v} \cdot \frac{w}{d} , \quad (7)$$

where we defined  $w := a + b$  as the tuning width. In case of  $D = 2$  we obtain

$$\bar{J} = \frac{4\pi}{9} \cdot \frac{f_{max}^2}{v} \cdot \frac{w^2}{d} , \quad (8)$$

and for arbitrary  $D$  it is straightforward to prove the following scaling rule:

$$\bar{J} \propto \frac{f_{max}^2}{v} \cdot \frac{w^{D-1}}{d} , \quad (9)$$

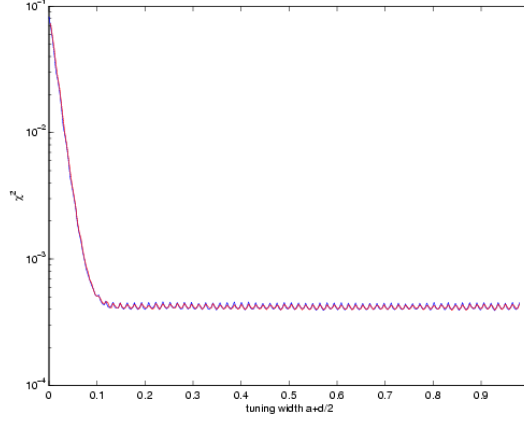


Figure 2: Minimum mean squared error as a function of the tuning width in case of  $D = 1$  and a Poissonian spike count distribution (maximum mean spike count equal to one). The corresponding model is given by equidistantly spaced box tuning functions and a uniform prior distribution with a support of unit length.

This scaling rule does not rely on the noise model as long as the amount of noise does not systematically depend on  $w$  or  $d$  and it can simply be generalized to other tuning functions, where the ramp function describing the radial component of the tuning curves is replaced with some other decay profile (e.g. any sigmoidal function). As explained in [4], however, the Fisher information  $J_1(|\vec{x} - \vec{c}|)$  is tightly related to the minimum mean squared error only if the tuning functions are sufficiently regularized.

For a constant dynamic range Eq. 9 reproduces Eq. 1, while Eq. 2 is obtained under the special assumption  $d \propto w$ . In other words, Fisher information of radial symmetric tuning functions does not only depend on the width, but also on the dynamic range. The dependence on the width is well described by Hinton’s rule, and has the same simple geometrical explanation, namely that the surface of a  $D$ -dimensional sphere with diameter  $w$  is proportional to  $w^{D-1}$ . The additional dependence on the dynamic range is due to the fact that Fisher information is proportional to the squared derivative of the tuning function. Therefore, the contribution of a single tuning function to the total Fisher information decreases quadratically with an increase of the dynamic range so that the net effect for the total Fisher information is negative (see Fig. 1).

Hitherto, we can summarize that large receptive fields appear to be advantageous for all  $D$ , while in case of  $D = 1$  the advantage is not very pronounced and cannot be shown on the basis of Fisher information, but requires a more accurate analysis (see Fig. 2). Nevertheless, the optimal shape of tuning clearly depends on the sort of constraints imposed on the tuning functions. E.g. one should note that the average firing rate of a neuron is roughly given by the product  $f_{max}w^D$  and should be limited due to energy consumption and dynamic stability. Above such a limit the maximum firing rate  $f_{max}$  has to be reduced proportional to  $w^{-D}$ , which leads to a small optimal tuning width for all  $D$  in striking contrast to the model considered above. In conclusion, it is not possible to decide the question of optimal tuning width independent from the problem at hand, but the choice of constraints has to be justified carefully in order to allow for meaningful conclusions.

### 3 Multi-dimensional encoding in the visual system

The original work by Hinton was motivated by seeking for theoretical determinants of the coding efficiency of sensory neurons in the visual system. Accordingly, his study is related to the task of reconstructing the location of a single light dot in the two dimensional visual field from the neuronal responses. While Hinton already remarked that his scaling rule holds true only under the assumption that not more than one light dot is located within the same receptive field simultaneously, the more general setting has rarely been investigated so far.

In order to describe arbitrary stimuli, we use a scalar intensity field  $I : [-b_x, b_x] \times [-b_y, b_y] \rightarrow [0, 1]$ ,  $(x, y) \mapsto I(x, y)$ . Furthermore, we assume that the rate response of neuron  $j = 1, \dots, N$  may be described by

$$z_j[I] = \langle F_j, I \rangle = \int_{-b_x}^{b_x} \int_{-b_y}^{b_y} F_j(x, y) I(x, y) dy dx \quad (10)$$

$$f_j[I] = g_j(z_j[I]) \quad (11)$$

where the first equation is a linear scalar product and  $g_j(z_j)$  is some nonlinear gain function with limited output range (i.e.  $f_{min} \leq g_j(z) \leq f_{max}$ ). In case of  $I$  being a delta function this model becomes equivalent to those analyzed in previous studies discussed above.

The advantage of this ansatz is that it allows to separate the problem of function approximation from the problem of noise due to spike count variability. In fact, this is a crucial step, because the optimal set of basis functions  $\{F_j : j = 1, \dots, N\}$  is mainly determined by the shape of the prior distribution  $P(I)$ , while the optimal gain function can be considered independently to some extent.

In the absence of signal-to-noise issues (as raised by spike count variability and a finite output range), the optimal set of basis functions  $\{F_j : j = 1, \dots, N\}$  with respect to a given prior is not unique, because there are always infinitely many other basis function sets that span the same linear space as it is known from simple linear algebra. Therefore, the tuning width cannot play a significant role for the quality of function approximation<sup>1</sup>. Furthermore, the assumption of radial symmetry, which is essential for the scaling rules considered above, is irrelevant for most stimuli, because this symmetry is conserved under the mapping Eq. 10 only in case when the intensity field  $I$  is a delta function. It is important to note that the visual system typically deals with very high-dimensional data (not two dimensional) and hence, it appears natural to assume the number of neurons to be equal to the number of dimensions a population encodes for. In other words, we here take the point of view that it makes sense to ask for the precision with which the  $N$ -dimensional vector  $(z_1, \dots, z_N)$  can be reconstructed from variable neuronal responses with mean  $(g_1(z_1), \dots, g_N(z_N))$ , rather than to ask for the precision with which an intensity field  $I$  can be reconstructed under the assumption of a particular prior distribution  $P(I)$ . A different way to motivate this point of view is to refer to the hypothesis that neuronal responses may reflect independent components of natural scenes [1].

### 4 To steepen or smoothen?

Under the assumption that the  $z_j$  are mutually independent (i.e.  $P(z_1, \dots, z_N) = \prod P(z_j)$ ), and that all dimensions are equivalent with respect to the variance of their prior distributions as well

---

<sup>1</sup>Note that function approximation is related to the question in how far the limited computational power of individual neurons may restrict the functionality of neuronal representations, which is not taken into account within the ideal observer paradigm considered above.

as to their relevance for further processing, it is optimal to encode each dimension exclusively by one neuron [4]. Therefore, the assumption given by Eq. 11 does not imply a loss in coding efficiency. While the tuning width turned out to be weakly related to the efficiency of sensory representations in the visual system, we find that the dynamic range of the gain function quite generally plays a significant role for population coding. In particular, it turns out that a small dynamic range or even binary encoding is advantageous whenever the time scale of decoding is small [5, 3]. The relevance of binary coding will be discussed along experimental data of the H1 neuron of the blowfly [10].

## References

- [1] J.J. Atick. Could information theory provide an ecological theory of sensory processing? *Network*, 3:213–251, 1992.
- [2] P. Baldi and W. Heiligenberg. How sensory maps could enhance resolution through ordered arrangements of broadly tuned receivers. *Biol. Cybern.*, 59:313–318, 1988.
- [3] M. Bethge, D. Rotermund, and K. Pawelzik. Optimal neural rate coding leads to bimodal firing rate distributions. *Network: Comput. Neural Syst.*, 2002. in press.
- [4] M. Bethge, D. Rotermund, and K. Pawelzik. Optimal short-term population coding: when fisher information fails. *Neural Comput.*, 14:2317–2351, 2002.
- [5] M. Bethge, D. Rotermund, and K. Pawelzik. A second order phase transition in neural rate coding: binary encoding is optimal for rapid signal transmission. *Physical Review Letters*, 2002. in press.
- [6] C.W. Eurich and H. Schwegler. Coarse coding: calculation of the resolution achieved by a population of large receptive fields. *Biol. Cybern.*, 76:357–363, 1997.
- [7] G.E. Hinton. Shape representation in parallel systems. *Biophys. J.*, 4:41–68, 1981.
- [8] G.E. Hinton, J.L. McClelland, and D.E. Rumelhart. Distributed representations. In D.E. Rumelhart and J.L. McClelland, editors, *Parallel Distributed Processing*, volume 1, pages 77–109. MIT Press, Cambridge, Massachusetts, 1986.
- [9] C. Lee, W.H. Rohrer, and D.L. Sparks. Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature*, 332:357–360, 1988.
- [10] G.D. Lewen, W. Bialek, and R.R. de Ruyter van Steveninck. Neural coding of naturalistic motion stimuli. *Network: Comput. Neural Syst.*, 12:317–329, 2001.
- [11] H.P. Snippe and J.J. Koenderink. Discrimination thresholds for channel-coded systems. *Biol. Cybern.*, 66:543–551, 1992.
- [12] R. Vogels. Population coding of stimulus orientation by striate cortical cells. *Biol. Cybern.*, 64:25–31, 1990.
- [13] K. Zhang and T.J. Sejnowski. Neuronal tuning: to sharpen or broaden. *Neural Comput.*, 11:75–84, 1999.
- [14] E. Zohary. Population coding of visual stimuli by cortical neurons tuned to more than one dimension. *Biol. Cybern.*, 66:265–272, 1992.