

Performance breakdown in optimal stimulus decoding

Lubomir Kostal,* Petr Lansky, and Stevan Pilarski

Institute of Physiology of the Czech Academy of Sciences, Videnska 1083, 14220 Prague 4, Czech Republic

One of the primary goals of neuroscience is to understand how neurons encode and process information about their environment. The problem is often approached indirectly by examining the degree to which the neuronal response reflects the stimulus feature of interest. In this context, the methods of signal estimation and detection theory provide the theoretical limits on the decoding accuracy with which the stimulus can be identified. The Cramér-Rao lower bound on the decoding precision is widely used, since it can be evaluated easily once the mathematical model of the stimulus-response relationship is determined. However, little is known about the behavior of different decoding schemes with respect to the bound if the neuronal population size is limited. We show that under broad conditions the optimal decoding displays a threshold-like shift in performance in dependence on the population size. The onset of the threshold determines a critical range where a small increment in size, signal-to-noise ratio or observation time yields a dramatic gain in the decoding precision. We demonstrate the existence of such threshold regions in early auditory and olfactory information coding. We discuss the origin of the threshold effect and its impact on the design of effective coding approaches in terms of relevant population size.

Keywords: Fisher information, Cramér-Rao bound, Neuronal coding, Threshold effect

INTRODUCTION

A satisfying description of the principles governing the *neuronal code*, i.e., the way information is represented in neuronal activity, is still not known [8, 41, 47]. A substantial volume of literature approaches the problem by analyzing the degree to which a selected aspect of neuronal activity, denoted as the *response*, reflects some quantity of interest, often denoted as the *stimulus*. The degree of such stimulus-response dependence is quantified by the coding *precision*, or accuracy, with which the stimulus can be ultimately decoded (estimated) from the observed responses. Comparing the coding precision for different types of responses (e.g., counts of action potentials or intervals in between them) then yields important inference on the hypothetical structure of the true neuronal code.

The actual *methodology* for determining the coding precision is provided by the theory of estimation where the mean square error (MSE) is employed as the *measure* of accuracy [20, 49]. Exact answers can be given in only a handful of cases though. In more general situations, including the typically nonlinear models of neural activity, the *asymptotic* theory of infinite populations becomes useful. A major result in this area is that for a wide range of problems there exists an estimator with the variance of errors attaining the *Cramér-Rao* (CR) bound – the theoretical minimum [51]. The CR bound is given in terms of *Fisher information*, a quantity which can be calculated once the dependence of the response distribution on the stimulus value is known. The CR bound can be also stated as the *lower* bound on estimator variance, even without discussing its asymptotic attainment. However, the power of such an approach is significantly weaker [51], e.g., the estimator *bias* or tightness of the bound might become a concern. Seen from a different perspective, one needs results on *both* the estimator and the bound on

possible accuracy. Having only the estimator without knowing the bound on efficiency, or only the bound without a clue on its achievability, is of lesser importance.

The convenience and apparent scope of the Fisher information concept has led to its common acceptance as a useful tool in a variety of problems, in both Bayesian and frequentist contexts [48], including fruitful applications to computational neuroscience [4, 7, 12, 13, 27, 38, 40, 45, 53, 55–57]. On the other hand, though the methods of asymptotic optimality theory are practically useful due to their relative simplicity, their *asymptotic* character must be pointed out. Most of the conclusions are strictly *limit* results as the neuronal population size tends to infinity. Whether such results can be used as approximations under given conditions is not obvious and a careful case-by-case examination is generally recommended [51].

While it is known that the CR bound may not be achievable under certain circumstances [4, 20], the actual dependence of the ultimate MSE on the population size is much less frequently investigated. We focus on the striking behavior of optimal decoding characterized by large and abrupt changes in decoding accuracy – the *threshold effect* [29, 32, 43, 49], which is not captured by the CR bound. In his pioneering study in the field of computational neuroscience, Xie [53] described the threshold effect in a homogeneous population of generic neurons with bell-shaped tuning curves. In addition, he provided a method to estimate the critical population size under the Gaussian setting. In this paper we extend Xie’s work by focusing on specific neural sensory systems (auditory and olfactory), thus showing the biological relevance of the threshold phenomenon. For these systems we demonstrate the complicated dependence of the decoding performance on the stimulus intensity (sound pressure or odorant concentration). We discuss the origins of the threshold effect and refer to the state-of-the-art literature on the subject in the mathematical engineering community. In order to understand the rapid MSE transitions we devise a simple linear toy model with non-Gaussian noise, and derive an approximate expression for the optimal decoding performance

* E-mail: kostal@biomed.cas.cz

in dependence on the sample size.

The threshold effect challenges the common assumption of the MSE approaching the CR bound in an unremarkable way as the population size increases. The key message is that there exists a *critical* range of population sizes where a small increment may yield a dramatic improvement in the decoding performance.

METHODS

Typically, an electrophysiological experiment consists of repeated trials in which the stimulus quantified by θ (e.g., intensity, position, color, etc.) is presented, and the neuronal response r is recorded [8, 11, 17]. Such an approach is used to construct models in which many neurons, possibly independent or even identical, respond at the same time. The response can be, e.g., the number of spikes counted in a certain time window, time to first spike or membrane potential. It is well known that the response varies randomly across trials [42, 47]. Thus by repeating the experiment it is possible, at least in theory, to collect enough data to estimate the probability that a particular response occurs. The neuronal *model* (the stimulus-response relationship) is then fully described in terms of the probability density function parameterized by θ , $f(r; \theta)$, or in terms of the probability mass function. Thus, the response is modeled as a random variable R distributed according to $f(r; \theta)$. In computational neuroscience, theoretical methods based on observed data and biophysical reasoning are frequently used to obtain $f(r; \theta)$ directly [8, 50].

As argued in the introduction, the key question is: assuming full knowledge of the model, $f(r; \theta)$, how precisely may one estimate the fixed stimulus value θ based on the observed response r ? The *estimator* $\hat{\theta}$, which is a function of the response random variable R , is employed to infer the true stimulus value. Ideally, the value $\hat{\theta}(r)$ for some observed response $R = r$ should be as close as possible to θ . Denote the mean value of the estimator as $m(\theta)$, i.e.,

$$m(\theta) = \int \hat{\theta}(r) f(r; \theta) dr. \quad (1)$$

Under regularity conditions on $f(r; \theta)$ [20, 51] the CR bound on the *estimator variance* can be established,

$$\text{Var } \hat{\theta}(R) \geq \frac{m'(\theta)^2}{J(\theta)}, \quad (2)$$

where $J(\theta)$ is the *Fisher information*,

$$J(\theta) = \int \left[\frac{\partial \log f(r; \theta)}{\partial \theta} \right]^2 f(r; \theta) dr. \quad (3)$$

In general, $m(\theta) \neq \theta$ and the estimator is *biased*. The estimator variance can be formally decomposed as

$$\text{Var } \hat{\theta}(R) = \varepsilon^2(\theta) - b^2(\theta), \quad (4)$$

where $\varepsilon^2(\theta) = \int [\hat{\theta}(r) - \theta]^2 f(r; \theta) dr$ is the *mean square error* (MSE) and

$$b(\theta) = m(\theta) - \theta \quad (5)$$

is the bias of the estimator. The actual bias dependence on θ is not known beforehand in a typical estimation problem, making Eq. (2) difficult to employ in practice. The CR bound is thus often stated for *unbiased* estimators by requiring $m(\theta) = \theta$ (at least in the infinitesimal neighbourhood of θ), yielding

$$\varepsilon^2(\theta) \geq \frac{1}{J(\theta)}. \quad (6)$$

Better decoding performance is expected as the population size n increases. In the simplest case the estimator is a function of n independent and identically distributed observations, $\hat{\theta}_n \equiv \hat{\theta}(r_1, \dots, r_n)$, obtained from the product joint density

$$f(r_1, \dots, r_n; \theta) = \prod_{i=1}^n f(r_i; \theta). \quad (7)$$

The estimator mean, $m_n(\theta)$ (as well as bias and MSE), is then generally a function of n . By substituting Eq. (7) into Eq. (3) and by repeating the derivation leading to Eq. (2), the sample size-dependent version of the CR bound is obtained [20],

$$\text{Var } \hat{\theta}(R_1, \dots, R_n) \geq \frac{m'_n(\theta)^2}{nJ(\theta)}, \quad (8)$$

where the expectations are taken with respect to the product density in (7). Similarly to Eq. (6) it holds

$$\varepsilon_n^2(\theta) \geq \frac{1}{nJ(\theta)} \quad (9)$$

for the unbiased version.

Since estimation bias is usually present, and the restriction to unbiased estimators is somewhat artificial and often even impossible, the evaluation of the CR bound and its interpretation might be problematic for finite population sizes. Classically, the bound plays a clear-cut role in the *asymptotic* setting of infinite sample sizes, which is in practice identified (approximately) with the large sample size or high signal-to-noise ratio situations [49, 51]. The theory of asymptotic statistics guarantees that under certain regularity conditions there *exists* an estimator such that as n increases, the probability distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ converges to a normal distribution with zero mean and variance equal to $1/J(\theta)$. In other words, there exists an estimator which is *asymptotically unbiased* and attains the CR bound. The popular *maximum likelihood estimator*,

$$\hat{\theta}_n = \arg \max_{\theta} \sum_{i=1}^n \log f(r_i; \theta), \quad (10)$$

is optimal in the above mentioned sense. An alternative, and much simpler method, is to employ the *moment estimator* defined by

$$\hat{\theta}_n = \mu^{-1}(\bar{r}), \quad (11)$$

where $\bar{r} = \sum_{i=1}^n r_i/n$ is the sample mean and $\mu^{-1}(r)$ is the inverse to the tuning curve defined by

$$\mu(\theta) = \int r f(r; \theta) dr. \quad (12)$$

Note that the moment estimator is generally not asymptotically efficient [51].

RESULTS

Depending on the nonlinearity of the neuronal stimulus-response relationship and the structure of noise, Fisher information might not give a good approximation to decoding precision for practically relevant sample sizes. We point to the discrepancy known as the *threshold effect* [32, 49] between the theoretically achievable and the actual MSE. Although the degradation of estimation performance with decreasing sample size is certainly not surprising, the rate of this degradation might be.

For the purpose of illustration we investigate two models of information coding in auditory and olfactory sensory systems based on experimental measurements.

Auditory-nerve rate response to tone intensity

Winslow and Sachs [52] studied the response of cat auditory nerve fibers to a varying sound pressure level of a pure tone. Ensembles of fibers best responding to 8 kHz sound frequency were investigated. The study distinguishes three fiber types according to their spontaneous activity: the *low* type forms 10% of the ensemble, the *medium* type accounts for 30% and the *high* type for the remaining 60%. The tuning curve of each fiber type follows the model by Sachs and Abbas [35] that describes the relationship between nerve fibers firing rate and basilar membrane displacement in response to a pure tone stimulus,

$$\mu(\theta) = \frac{10^{c\theta/20} r_m}{10^{t_E/20} (1 + 10^{-t_I/20} 10^{\theta/10})^{c/3} + 10^{c\theta/20}} + r_{sp}. \quad (13)$$

The stimulus θ is the sound pressure level expressed in decibels relative to an arbitrary reference and the response is given in spikes per second. The parameter t_E determines the response threshold, $t_I = 100$ dB is a parameter related to the two-tone suppression model [35], r_m is the maximum rate change which can be observed in response to a pure tone stimulus, r_{sp} is the spontaneous discharge rate and $c = 1.77$ is a constant. The values of t_E, r_m, r_{sp} for each fiber type are summarized in Table 1. There is some natural variation in the parameter values. We account for this variation by increasing the response variance of each neuronal type with respect to the originally reported sub-Poisson relationship. In accordance with Winslow and Sachs [52] we assume that the constant tone duration is sufficiently long so that the distribution of spike counts is well approximated by a Gaussian distribution with mean given by Eq. (13) and a Poisson-like variance, $\sigma(\theta)^2 = \mu(\theta)$. The potential negativity of responses is of no importance to what follows.

We may conveniently consider the ensemble response vector to be identically and *independently* [52] distributed across ensembles according to the model

$$f(r; \theta) = \sum_{k=1}^3 \frac{w_k}{\sqrt{2\pi\mu_k(\theta)}} \exp\left(-\frac{[r - \mu_k(\theta)]^2}{2\mu_k(\theta)}\right), \quad (14)$$

	low	medium	high
w	0.1	0.3	0.6
t_E [dB]	89.4	65.7	31.9
r_m [spike/s]	135.1	183.6	90.6
r_{sp} [spike/s]	0.5	5.4	68.4

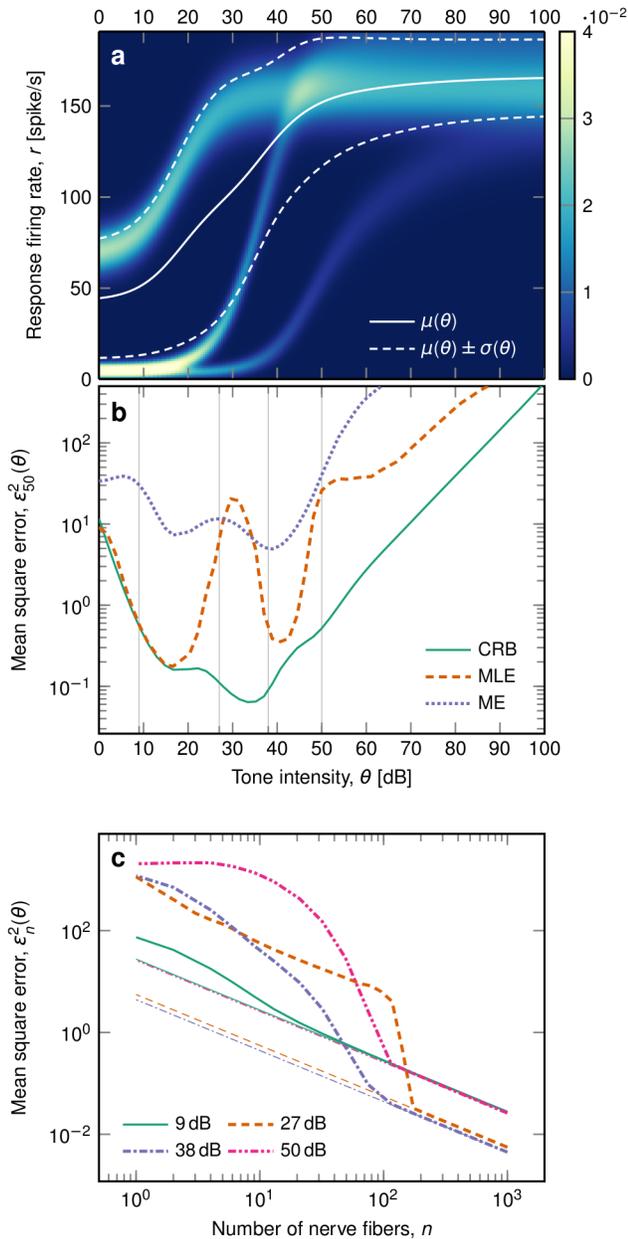
Table 1. Parameters of auditory nerve fibers. Three types of fibers are distinguished with respect to their spontaneous activity (low, medium, high). The parameters t_E, r_m, r_{sp} determine the tuning curve in Eq. (13), the factor w specifies the fraction of each type in the population (data taken from Winslow and Sachs [52]).

where the index $k = 1, 2, 3$ indexes the parameter sets of each neuronal type (low, medium, high), see Table 1. The stimulus-response model given by Eq. (14) is shown in Fig. 1a, accompanied by the ensemble tuning curve, $\mu(\theta) = \sum_k w_k \mu_k(\theta)$, and by the standard deviation of the ensemble response.

The model in Eq. (14) satisfies the conditions for the asymptotic efficiency of the maximum likelihood estimator [51]. However, the convergence of the estimator towards the bound deserves a closer look. Fig. 1b shows the CR bound evaluated according to Eq. (9) for $n = 50$ nerve fibers and tone intensities in the range from 0 to 100 dB. For a comparison, the MSEs of the maximum likelihood estimator given by Eq. (10) and the moment estimator given by Eq. (11) are shown. The MSE of the moment estimator is significantly larger than the CR bound for most of investigated tone intensities, with local minima that do not coincide with minimal CR bound values (maximal Fisher information). The behavior of the maximum likelihood estimator is striking in the sense that it performs poorly in the neighbourhood of smallest CR bound values – precisely the opposite of what one would expect. There is even a region of intensities where the MSE of the maximum likelihood estimator is larger than that of the moment estimator. In other words, for finite sample sizes the CR bound might not provide even a crude approximation to the estimation MSE. Furthermore, the maximal Fisher information value and the smallest MSE do not necessarily coincide for some fixed value of n . The reason lies in the fact that for distinct stimuli values the threshold region occurs at different values of n .

The approach of the maximum likelihood estimator towards the CR bound as the number of nerve fibers increases is shown in Fig. 1c for four selected values of $\theta = 9, 27, 38$ and 50 dB (also marked in Fig. 1b). The logarithmic scale of both axes is convenient since the CR bound dependence on n is linear, all lines having the same slope with vertical offset equal to $1/J(\theta)$ [49]. The maximum likelihood estimator for $\theta = 27$ and 50 dB displays the *threshold effect* characterized by a rapid deterioration of performance with respect to the CR bound as the sample size decreases below certain critical, *threshold* region. For example, if the population of 180 nerve fibers is reduced by 20%, the decoding precision at 27 dB drops by more than one order of magnitude. On the other hand, increasing the population by 20% past the threshold region improves the precision 1.2 times only.

Threshold effects make the interpretation of $J(\theta)$ difficult for finite sample sizes. For example, even though $J(9) \approx$



$J(50)$ in Fig. 1c, the actual MSE in estimating $\theta = 50$ dB is almost two orders of magnitude larger than that of $\theta = 9$ dB below the threshold region. Moreover the estimation of $\theta = 27$ dB or $\theta = 38$ dB is much less accurate than that of $\theta = 9$ dB up to $n = 50$, contrary to the expectations based on Fisher information only. The asymptotic character of CR bound is prominent here: although $\theta = 27$ dB is in theory better identifiable than $\theta = 9$ dB, it takes a larger sample size to do so. Fig. 1c also demonstrates that the sufficiently high n that guarantees estimator efficiency generally does not depend on the value of Fisher information, but on the more detailed probabilistic properties of the model.

The maximum likelihood estimator is generally biased, but in the model given by Eq. (14) the bias is small, decreasing rapidly with n so that the MSE shown in Fig. 1b, c is dominated

Figure 1. Estimation of tone intensity from the response firing rate of the auditory nerve. (a) Stimulus-response model of a population of auditory nerve fibers best responding to 8 kHz sound frequency in a cat. The color indicates the probability density of observing a particular firing rate given a pure tone of sound pressure level θ (in decibels). Three types of auditory nerve fibers can be distinguished in the fibre ensemble [52], the average tuning curve of the ensemble (solid white line) and its standard deviation (dashed) are indicated. (b) Comparison of Cramér-Rao bound (CRB) with the mean square error (MSE) of maximum likelihood (MLE) and moment (ME) estimators for $n = 50$ nerve fibers. The maximum likelihood estimator attains the CRB for a range of intensities up to 17 dB. Note that the theoretically best identifiable range between cca. 20 and 40 dB actually results in the worst performance, demonstrating that the value of Fisher information may be poorly related to the actual MSE for finite sample sizes. Four values of intensity are marked by grey vertical lines in (b) and their MSEs are shown in (c) in dependence on the number of fibers in the population. The convergence of the maximum likelihood estimator MSE to the CRB, especially for $\theta = 27$ and 50 dB, shows the pronounced *threshold effect*, where the region of poor performance is followed by a rapid transition towards the CRB. The threshold region determines the critical population size where a small change in the number of neurons results in a dramatic change of decoding performance.

by estimation variance. (The MSE of the moment estimator is not shown in Fig. 1c since it does not achieve CR bound with increasing n .) Frequently, the moment estimator also exhibits a threshold-like effect, although much smaller in magnitude. The threshold effect itself is not a particular pathological property of numerical maximization of Eq. (10), though. Theoretical considerations show that the abrupt deviation of the estimator MSE from the CR bound is inevitable, at least in certain cases, as argued in the Discussion.

The analysis of optimal decoding performance presented above allows us to make some additional observations on the results presented in Winslow and Sachs [52]. Originally, the population model of nerve fibers was used to determine the *just-noticeable difference* in the perception of tone intensity by means of an optimal (close to maximum likelihood) decision rule. Despite slightly different methodology, the conclusions of Winslow and Sachs [52] based on a population of $n = 260$ fibers are qualitatively consistent with the CR bound shown in Fig. 1b. In particular, the region of tone levels between 30 and 40 dB is shown to be the best identifiable, with performance rapidly decreasing for both smaller and larger values. The explanation of the correspondence lies in the fact that the square root of the CR bound is approximately proportional to the just noticeable difference [8] and, more importantly, the population size is *sufficiently high* for the CR bound attainment (Fig. 1c). We do not have enough anatomical data to extend the results of Winslow and Sachs [52] and we wish to restrain from speculations. Rather, the message of this example lies in the notion of estimator threshold performance and the fact that even in simple models, Fisher information might be a misleading measure of decoding accuracy. Together, these aspects might be used to put *non-trivial bounds* on the proper population sizes.

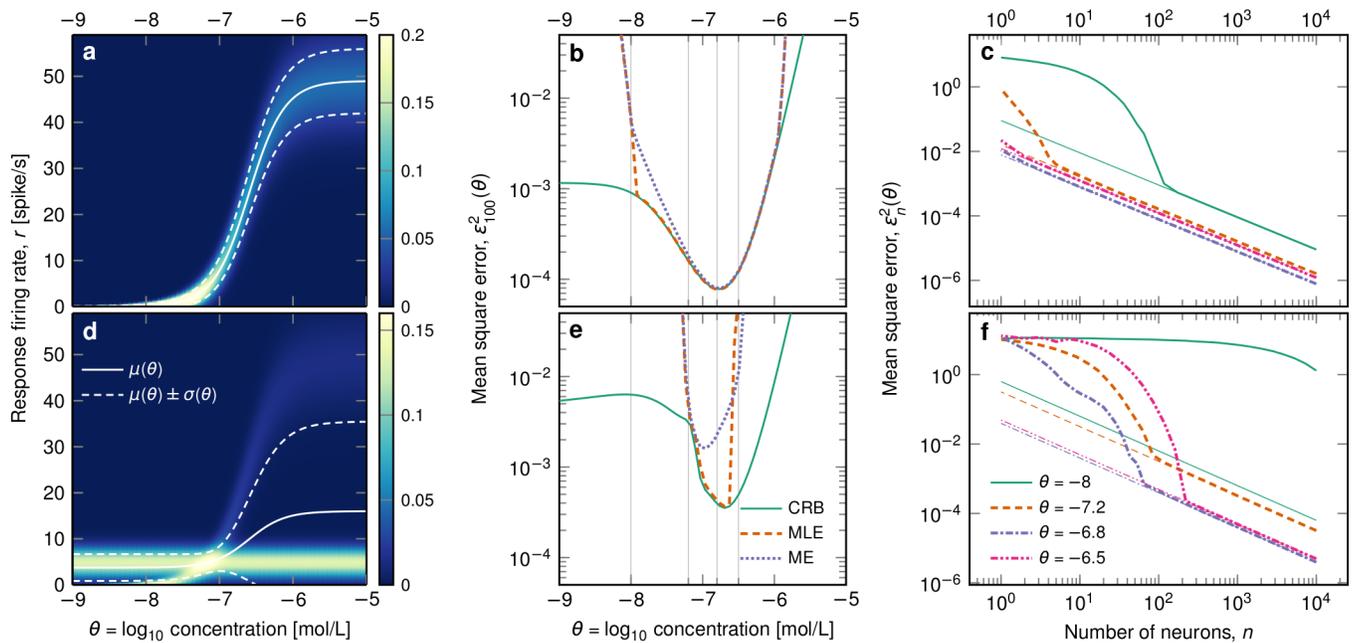


Figure 2. Estimation of a single odorant concentration in the absence (top row) and presence (bottom row) of spontaneous activity. The stimulus-response model is based on electrophysiological recordings of rat olfactory receptor neurons [33]. (a) Probability density function of the response firing rate for each value of concentration. The tuning curve (white solid line) is described by the Hill function. (b) Cramér-Rao bound and MSEs for the maximum likelihood and moment estimators are compared for the population size $n = 100$. The maximum likelihood estimator is efficient for (almost) the entire range of considered concentrations. Note the rapid deterioration of performance for $\theta < -8$ as discussed in the main text. (c) The behavior of the maximum likelihood estimator MSE for selected concentrations, $\theta = -8, -7.2, -6.8, -6.5$, is further investigated in dependence on the sample size. Except for $\theta = -8$ no threshold effect is observed. In the presence of background spontaneous activity the overall response of the neuronal ensemble is described by the mixture stimulus-response model (d) with decreased coding range. Qualitatively, the CR bound is not changed too much (b, e), but a rapid increase in the maximum likelihood estimator MSE is observed for mid- to high-concentrations as well as for very small ones (e). The maximum likelihood estimation under spontaneous background activity is plagued by the emergence of threshold effects (f), which are not captured by the decreased Fisher information with respect to (c).

Early neural coding of single odorant concentration

An increase of noise in the stimulus-response relationship results in a poorer decoding performance, as reflected by a lowering of Fisher information. However, the decrease in the information value may not capture the actual loss of decoding precision due to the emergence of threshold regions, as demonstrated by the following model.

Many sensory neurons are naturally exposed to stimuli which are relatively weak when compared to the noise of either environmental or biological origin. A particular instance of such internal noise is the spontaneous activity of neurons, which is usually highly irregular and unpredictable *in vivo*, as is the case of rat olfactory receptor neurons (ORN) [10]. Stimulus estimation under such circumstances is complicated by the adjustment of the evoked response against the spontaneous activity [5, 18]. In the case of naturally intermittent stimulation [19, 24, 25] the number of actually responding neurons in the ensemble might be significantly smaller than the number of spontaneously active ones, as investigated in the sparse-coding setup [14, 26].

Rospars *et al.* [33] studied dose-response relationships in rat ORNs based on electrophysiological recordings for various pure

odorants and their mixtures. Each ORN expresses only one type of olfactory receptor – though each receptor type can recognize different odorants and, conversely, a single odorant is usually recognized by multiple receptor types. In what follows we focus on the stimulation by a single odorant only; see Rospars *et al.* [33] for details on the experimental procedures. The ORN tuning curve relating the odorant concentration and the evoked firing rate follows from the fact that the conductance change in the sensory outer segment of the membrane of the olfactory receptor neuron is described as an amplified version of the odorant-receptor interaction [34]. Assuming that the dependence of voltage on conductance and the dependence of frequency on voltage are both linear, it follows that the tuning curve corresponds to the Hill function (see Rospars *et al.* [33] for more details)

$$\mu(\theta) = \frac{F_M}{1 + 10^{(\log_{10} K - \theta)N}}. \quad (15)$$

The response $\mu(\theta)$ is expressed in spikes/second, the stimulus θ is the decadic logarithm of odorant concentration in mol/L, F_M is the maximum asymptotic firing rate, K is the odorant concentration that evokes response rate equal to $F_M/2$ and N is the Hill coefficient. The values of the model parameters

depend on the odorant type, here we choose the typical values $F_M = 49$ spikes/s, $N = 1.8$ and $K = 2.5 \cdot 10^{-7}$ mol/L [33]. For convenience we assume again that the response spike counts are normally distributed with a Poisson-like variance, as done commonly [3, 28, 56]. The exact form of the response distribution or variance-mean dependence does not affect the qualitative aspects of our conclusions. Then the stimulus-response model of the ORN is given by the Gaussian density

$$f(r; \theta) = \frac{1}{\sqrt{2\pi\mu(\theta)}} \exp\left(-\frac{[r - \mu(\theta)]^2}{2\mu(\theta)}\right). \quad (16)$$

The model is shown in Fig. 2a, together with the tuning curve given by Eq. (15) and the corresponding standard deviation of responses. Fig. 2b shows the CR bound for $n = 100$ identical neurons together with MSEs of both maximum likelihood and moment estimators. The maximum likelihood estimator quite precisely attains the CR bound for the whole range of relevant concentrations. The performance of the moment estimator is almost the same as that of the maximum likelihood estimator except for a region of very low concentrations. The first conclusion is that the maximum likelihood estimator behaves optimally, as desired and usually expected. Note, however, that $\theta \doteq -8$ is the “critical” stimulus value below which the maximum likelihood estimator MSE rapidly deviates from the CR bound. Since the ORN model is CR bound-admissible and also satisfies the conditions for maximum likelihood estimator asymptotic efficiency, the value of the critical point decreases with increasing n . Fisher information is not changing rapidly in the vicinity of the critical point, in fact it approaches a constant value equal to $N^2 \log(10)^2/2$ as the concentration decreases, but the internal properties of the model do change significantly. The finiteness of Fisher information results from the subtle and exact mathematical balance between the decrease of noise variance (pushing $J(\theta)$ to infinity) and the fact that $f(r; \theta)$ becomes less and less θ -dependent (pushing $J(\theta)$ to zero). Obviously, one cannot expect the CR bound to provide information on the actual MSE for small sample sizes under such circumstances. The outlying results of the maximization in Eq. (10) pollute the maximum likelihood estimator variance rapidly, and effectively cause the performance breakdown [6].

Fig. 2c shows the dependence of the MSE on n for four selected concentration values, $\theta = -8, -7.2, -6.8, -6.5$. There is no visible threshold effect requiring further attention, except for $\theta = -8$ where the concentration is too small and where the effect is explained in the paragraph above.

As mentioned, not all ORNs in the ensemble respond to the stimulus at the same time (either because of their receptor type or because of the spatio-temporal intermittence of the stimulation) and a majority of ORNs are spontaneously active. Assume that only some neurons in the population are responding to the stimulus. The stimulus-response model may be expressed as

$$f(r; \theta) = \frac{w_S}{\sqrt{2\pi F_S}} \exp\left(-\frac{(r - F_S)^2}{2F_S}\right) + \frac{1 - w_S}{\sqrt{2\pi\mu(\theta)}} \exp\left(-\frac{[r - \mu(\theta)]^2}{2\mu(\theta)}\right), \quad (17)$$

where $0 < w_S < 1$ is the fraction of spontaneously active neurons, $F_S = 5$ spikes/sec is the average ORN spontaneous activity rate [33] and Poisson variability is assumed.

Fig. 2d–f shows results obtained on model from Eq. (17) for $w_S = 3/4$ and the total population size $n = 100$, the scale corresponds to panels a–c. The coding range of the tuning curve in Fig. 2d is reduced when compared to Fig. 2a due to the presence of background spontaneous activity, and the CR bound is correspondingly increased (e). The global shapes of the CR bound in (b) and (e) are similar and the minimum is approximately at the same position in both cases. However, the presence of spontaneous activity limits the actual estimation reliability to only a part of the theoretical coding range (e). There is a rapid deterioration in the maximum likelihood estimator performance for θ exceeding half of the coding range after $\theta \doteq -6.6$, mainly due to the increased response standard deviation. Also note that the moment estimator performs comparably to the maximum likelihood estimator only for a range of low concentrations around $\theta = -7.2$, and for $\theta > -6.6$ there is even a region where the moment estimator performs better.

The MSE of the maximum likelihood estimator exhibits the typical threshold behavior (Fig. 2f). The threshold region onset for $\theta = -8$ exceeds the sample size of $n = 10^4$. Similarly to the auditory model (Fig. 1c) the value of Fisher information is not related to the required sample size. For example, $J(-6.5) > J(-7.2)$ but $\theta = -6.5$ is harder to estimate up to $n \doteq 200$. There is a huge difference in the estimation performance for the two concentrations with equal Fisher information values, $\theta = -6.5$ and $\theta = -6.8$.

Threshold effect and maximum likelihood decoding: an illustration

The following model is not related to any neuroscientific problem, but it allows certain approximate calculations, which are impossible for the models presented above. In addition, we demonstrate that the threshold effect is not related to the existence of bias. We consider the maximum likelihood decoding and we show its relationship to the smallest achievable MSE. Our example belongs to the simplest class of estimation problems, the case of *linear* estimation. The “response” random variable R is related to some particular “stimulus” value θ as

$$R = \theta + X, \quad (18)$$

where X is the *noise* random variable independent of θ , and θ can be any real number. We consider X to be a mixture of two Gaussian distributions, $X \sim (1 - p)\mathcal{N}(0, \sigma_1^2) + p\mathcal{N}(0, \sigma_2^2)$, so that the “stimulus-response” model is

$$f(r; \theta) = \frac{1 - p}{\sigma_1 \sqrt{2\pi}} \exp\left(-\frac{(r - \theta)^2}{2\sigma_1^2}\right) + \frac{p}{\sigma_2 \sqrt{2\pi}} \exp\left(-\frac{(r - \theta)^2}{2\sigma_2^2}\right), \quad (19)$$

and $0 < p < 1$. The maximum likelihood estimator is *unbiased* for all sample sizes due to symmetry and linearity of the model

in Eq. (19), therefore the CR bound in Eq. (9) is valid for all n . Furthermore, let

$$\sigma_1 \gg \sigma_2. \quad (20)$$

The MSE of the maximum likelihood estimator in dependence on the sample size can rarely be given in a closed form. However, for the model at hand we can approximately express the MSE as follows. Assume that all n samples are from the $\mathcal{N}(\theta, \sigma_2^2)$ distribution. Such event occurs with probability p^n . The MSE of the maximum likelihood estimator equals the CR bound, σ_2^2/n , and hence it is optimal [20]. Note that the modes of both Gaussians in Eq. (19) coincide, and due to Eq. (20) the precision of the maximum likelihood estimator is not affected by the presence of $\mathcal{N}(\theta, \sigma_1^2)$ in the likelihood function. Let generally $1 \leq k \leq n$ samples be drawn from the $\mathcal{N}(\theta, \sigma_2^2)$ distribution so that the maximum likelihood MSE equals σ_2^2/k . The remaining $n - k$ samples from $\mathcal{N}(\theta, \sigma_1^2)$ will not affect this MSE unless they are clustered within a short interval, say, $3\sigma_2$. The inequality (20) guarantees that such clustering is rare. Since the probability of observing the k samples in n trials is given by the binomial distribution, the average MSE can be written as

$$\varepsilon_n^2(\theta) \doteq \frac{(1-p)^n \sigma_1^2}{n} + \sum_{k=1}^n \binom{n}{k} p^k (1-p)^{n-k} \frac{\sigma_2^2}{k}, \quad (21)$$

where the first term results from $k = 0$ (all samples come from $\mathcal{N}(\theta, \sigma_1^2)$).

The Eq. (21) represents not only the maximum likelihood MSE but also the smallest *achievable* MSE since all terms are based on the CR bound and thus cannot be made smaller. As p decreases the threshold region is postponed to higher sample sizes, since the factor $(1-p)^n$ in Eq. (21) decreases less rapidly. The approximation predicts a significant threshold effect and a correct $1/n$ -like scaling in the asymptotic region (Fig. 3). For the chosen values $p = 0.1, \sigma_1 = 1$ and $\sigma_2 = 0.001$ the Eq. (21) already provides an excellent approximation to numerical simulation.

The expected curvature of the likelihood function at its maximum is dominated by the σ_2 -component of the mixture, due to Eq. (20). Hence the CR bound predicts a relatively small MSE (Fig. 3). Below the threshold, however, the major contributor to the MSE is the σ_1 -component as follows from Eq. (21). These “large” errors are often denoted as *non-local* in the threshold-effect literature (see the next section) because they generally cannot be inferred from the local behavior of the likelihood function in the neighbourhood of its maximum.

DISCUSSION

The decoding MSE often deviates from the CR bound abruptly as the sample or population size decreases below certain critical region. The region where the CR bound approximates the MSE of maximum likelihood (or any other) decoding is noticeably separated from the region of poor compliance by a transitional zone of rapid convergence towards the bound. This

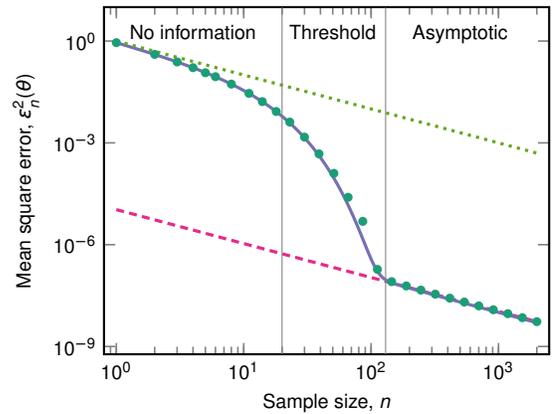


Figure 3. Illustration of the threshold effect. The parametric model is given by Eq. (19) with $\theta = 0, p = 0.1, \sigma_1 = 1$ and $\sigma_2 = 0.001$. The mean square error (MSE) of the maximum likelihood decoding is averaged over 50 000 trials (dots) and shows a prominent threshold effect, which is not captured by the Cramér-Rao (CR) bound (dashed) even though the decoding is unbiased. The analytic approximation to the smallest achievable MSE (solid, Eq. (21)) is in excellent agreement with numerical results and demonstrates that the maximum likelihood decoder is optimal for this model. The MSE for small sample sizes is effectively given by the CR bound of the broad Gaussian component (σ_1^2/n , dotted) and the information from the specific (narrow) Gaussian component increases non-linearly with sample size. Three regions of decoding performance (no information, threshold, asymptotic) are typically distinguished (see the Discussion).

phenomenon, denoted as the *threshold effect*, was historically first noted in the pulse code and frequency modulation systems in dependence on the signal-to-noise ratio [29, 32, 37, 43] and discussed in the context of classical estimation theory by van Trees [49]. The threshold effect occurs under broad conditions, it has been studied in deterministic as well as Bayesian settings including non-Gaussian priors (see the overview in [48]). To our best knowledge, the only computational neuroscience study mentioning the threshold effect is by Xie [53], who describes a threshold behavior of the maximum likelihood estimator in a population of generic neurons with bell-shaped tuning curves and relates its onset to the signal-to-noise ratio of the system. Here we emphasize the generality and potential importance of the threshold effect (as perceived in the field of mathematical engineering [48]) for specific biological systems and discuss the phenomenon thoroughly.

When the MSE vs. the sample size is visualised on the log-log scale (Figs. 1c and 2c, f), three regions are heuristically distinguished as n increases (Fig. 3) [1, 31, 49]:

1. The initial *no information* region where the MSE is dominated by non-local errors (see below). Typical decoding errors are several orders of magnitude larger than predicted by the CR bound. The estimate is often distributed uniformly over the range of searched values and thus the MSE may not initially improve with n (Fig. 2f).
2. The transitional *threshold* region is characterized by the rapid (faster than $1/n$) decrease of the MSE towards the

CR bound (a pronounced case is shown in Fig. 1c for $\theta = 27$ dB). Man-made devices are often required to operate above the threshold region, which sets a non-trivial bound on the required number of observations [48].

3. The final *asymptotic* region where the MSE scales as $1/[nJ(\theta)]$. The CR bound is only useful in this region since it cannot account for the threshold phenomenon.

The logarithmic scaling ensures that individual samples or neurons in the population are not the actual units of interest. Rather we are interested in the relative increment in the size. For example, if the population size is increased by 10 % in the threshold region then the MSE does not decrease 1.1 times only, as predicted by the CR bound in Eq. (8), but often by several orders of magnitude.

It is important to note that the rapid increase in the decoding precision is not due to correlations, synchronization or general dependence in the neuronal activity. The threshold effect is a statistical phenomenon occurring even under completely independent and identically distributed responses. The origin of the threshold phenomenon and its relationship to the CR bound are heuristically explained in the case of the maximum likelihood decoding [37, 49]. Fisher information in Eq. (3) is a *local* quantity, equal to the expected curvature of the main peak of the likelihood function, or equivalently, to the expected curvature of the likelihood at the true value of the parameter. The main peak is guaranteed to dominate the likelihood function asymptotically, as the sample size tends to infinity. However, for finite sample sizes the likelihood function may develop peaks well separated from the true stimulus value, which occasionally return the maximum likelihood. Such an error is denoted as *non-local*. The global shape of the likelihood function, by definition, is not included in the Fisher information and therefore the CR bound cannot predict the MSE whenever the non-local errors matter. The threshold effect, in essence, depends on the frequency of the non-local errors.

In his seminal work on unbiased estimation, Barankin [2] provided the largest lower bound on MSE that is *achievable* (not necessarily by the maximum likelihood approach), although no method is known to find such optimal estimators. His bound is very tedious to evaluate in practice but the existence of threshold effects can be demonstrated analytically [23, 37]. The maximum likelihood MSE in the threshold region closely follows the Barankin bound in certain models [1, 6], justifying heuristically the decoder optimality in finite sample sizes. Various other bounds that capture the threshold effect in biased estimation have been proposed in the literature for both deterministic and Bayesian scenarios, see an exhaustive overview in van Trees and Bell [48] or a concise list of references in Renaux [30]. In particular the Bayesian approach is of interest to the neuroscience community [3, 54]. Unfortunately, the bounds usually apply to particular decoding schemes, models, or are restrictive in other manner. For example, Xie [53] derived an approximate expression for the threshold region in a homogeneous population of Gaussian neurons. The expression is not suitable for

more complicated (e.g., multimodal) noise distributions since it depends on the Taylor expansion of the likelihood function [48]. The actual amplitude of the threshold effect is also of interest but its estimation would require a separate methodology. A further discussion of various bounds and their potential applicability to neuroscientific models is one of the key points that needs to be addressed in a future research. The analysis of the threshold effect under the correlated response [16, 21], where the Fisher information does not scale in proportion to the number of samples or neurons [7, 46], represents an additional challenge.

CONCLUSIONS

We showed that optimal decoding procedures may be plagued by threshold effects, making direct usage of Fisher information problematic in finite sample (or population) sizes. The threshold effect delays (in terms of the required sample size or signal-to-noise ratio) the onset of the asymptotic regime. More properly, Fisher information is a relevant quantity when the estimation is performed *above* the threshold region [6].

We demonstrated that the threshold effect occurs under broad conditions, depending on the subtle probabilistic properties of the model. The investigated auditory and olfactory models are represented by a mixture of distributions. However, the threshold effect may occur in a variety of models, in homogeneous or non-homogeneous populations, or in single models with Gaussian noise (such as in the olfactory neuron for low concentrations, Fig. 2c). There is currently no simple and reliable indicator of the threshold effect strength that could be obtained without performing unpractically time-consuming numerical calculations, although the topic is an active research area (see van Trees and Bell [48] and references therein).

We focused on two decoding schemes throughout this paper. The maximum likelihood estimation is used as the optimal procedure frequently [15, 22, 28, 39, 40, 53] and is also reported to be biologically plausible in some cases [9]. Nonetheless, there are disadvantages associated with the maximum likelihood approach as well. For example, its performance depends on the form of $f(r; \theta)$ and consequently it is quite sensitive to the mismatch of the model. The moment estimator is often easy to calculate but does not generally attain the CR bound asymptotically [51]. However, we believe that especially in the non-asymptotic setting of real neuronal systems, the computational complexity or energetic expenses of decoding operations should be taken into account [36, 44]. In other words, the concept of asymptotic efficiency should not provide the ultimate guideline for the choice of decoder.

ACKNOWLEDGMENTS

This work was supported by the Institute of Physiology RVO:67985823 and by the Czech Science Foundation project 15-08066S.

- [1] Athley, F., “Threshold region performance of maximum likelihood direction of arrival estimators,” *IEEE Trans. on Signal Process.* **53**, 1359–1373 (2005).
- [2] Barankin, E. W., “Locally best unbiased estimates,” *Ann. Math. Stat.* **20**, 477–501 (1949).
- [3] Berens, P., Ecker, A. S., Gerwin, S., Tolias, A. S., and Bethge, M., “Reassessing optimal neural population codes with neuro-metric functions,” *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4423–4428 (2011).
- [4] Bethge, M., Rotermund, D., and Pawelzik, K., “Optimal short-term population coding: when Fisher information fails,” *Neural Comput.* **14**, 2317–2351 (2002).
- [5] Chacron, M. J., Longtin, A., and Maler, L., “The effects of spontaneous activity, background noise, and the stimulus ensemble on information transfer in neurons,” *Netw. Comput. Neural Syst.* **14**, 803–824 (2003).
- [6] Chaumette, E., Galy, J., Quinlan, A., and Larzabal, P., “A New Barankin Bound Approximation for the Prediction of the Threshold Region Performance of Maximum Likelihood Estimators,” *IEEE Trans. Sig. Proc.* **56**, 5319–5333 (2008).
- [7] Dayan, P. and Abbott, L. F., “The Effect of Correlated Variability on the Accuracy of a Population Code,” *Neural Comput.* **11**, 91–101 (1999).
- [8] Dayan, P. and Abbott, L. F., *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (MIT Press, Cambridge, 2001).
- [9] Deneve, S., Latham, P. E., and Pouget, A., “Reading population codes: a neural implementation of ideal observers,” *Nat. Neurosci.* **2**, 740–745 (1999).
- [10] Duchamp-Viret, P., Kostal, L., Chaput, M., Lansky, P., and Rospars, J.-P., “Patterns of spontaneous activity in single rat olfactory receptor neurons are different in normally breathing and tracheotomized animals,” *J. Neurobiol.* **65**, 97–114 (2005).
- [11] Gerstner, W. and Kistler, W. M., *Spiking Neuron Models: Single Neurons, Populations, Plasticity* (Cambridge University Press, Cambridge, 2002).
- [12] Greenwood, P. E., Ward, L. M., Russell, D. F., Neiman, A., and Moss, F., “Stochastic resonance enhances the electrosensory information available to paddlefish for prey capture,” *Phys. Rev. Lett.* **84**, 4773–4776 (2000).
- [13] Harper, N. S. and McAlpine, D., “Optimal neural population coding of an auditory spatial cue,” *Nature* **430**, 682–686 (2004).
- [14] Hromádka, T., DeWeese, M. R., and Zador, A. M., “Sparse Representation of Sounds in the Unanesthetized Auditory Cortex,” *PLoS Biol.* **6**, e16 (2008).
- [15] Iolov, A., Ditlevsen, S., and Longtin, A., “Stochastic optimal control of single neuron spike trains,” *J. Neural Eng.* **11**, 046004 (2014).
- [16] Jackson, B. S. and Carney, L. H., “The spontaneous-rate histogram of the auditory nerve can be explained by only two or three spontaneous rates and long-range dependence,” *JARO* **6**, 148–159 (2005).
- [17] Kandel, E. R., Schwartz, J. H., and Jessel, T. M., *Principles of neural science* (Elsevier, New York, 1991).
- [18] Kostal, L., Lansky, P., and Rospars, J.-P., “Review: neuronal coding and spiking randomness,” *Eur. J. Neurosci.* **26**, 2693–2701 (2007).
- [19] Kostal, L., Lansky, P., and Rospars, J.-P., “Efficient olfactory coding in the pheromone receptor neuron of a moth,” *PLoS Comput. Biol.* **4**, e1000053 (2008).
- [20] Lehmann, E. L. and Casella, G., *Theory of point estimation* (Springer Verlag, New York, 1998).
- [21] Lowen, S. B. and Teich, M. C., “Auditory-nerve action potentials form a nonrenewal point process over short as well as long time scales,” *J. Acoustic. Soc. Am.* **92**, 803–806 (1992).
- [22] Ludwig, K. A., Miriani, R. M., Langhals, N. B., Marzullo, T. C., and Kipke, D. R., “Use of a Bayesian maximum-likelihood classifier to generate training data for brain-machine interfaces,” *J. Neural Eng.* **8**, 046009 (2011).
- [23] McAulay, R. J. and Seidman, L. P., “A useful form of the barankin lower bound and its application to ppm threshold analysis,” *IEEE Trans. Inf. Theory* **15**, 273–279 (1969).
- [24] Murlis, J., “Odor plumes and the signal they provide,” in *Insect Pheromone Research: New Directions*, edited by R. Carde and A. Minks (Chapman and Hall, New York, 1996) pp. 221–231.
- [25] Mylne, K. R. and Mason, P. J., “Concentration fluctuation measurements in a dispersing plume at a range of up to 1000m,” *Q. J. Roy. Meteor. Soc.* **117**, 177–206 (1991).
- [26] Olshausen, B. A. and Field, D. J., “Sparse coding of sensory inputs,” *Curr. Opin. Neurobiol.* **14**, 481–487 (2004).
- [27] Paradiso, M. A., “A theory for the use of visual orientation information which exploits the columnar structure of striate cortex,” *Biol. Cybern.* **58**, 35–49 (1988).
- [28] Pouget, A., Dayan, P., and Zemel, R. S., “Inference and computation with population codes,” *Annu. Rev. Neurosci.* **26**, 381–410 (2003).
- [29] Quinn, B. G. and Kootsookos, P. J., “Threshold Behaviour of the Maximum Likelihood Estimator of Frequency,” *IEEE Trans. Sig. Proc.* **42**, 3291–3294 (1994).
- [30] Renaux, A., “Weiss-Weinstein bound for data-aided carrier estimation,” *IEEE Trans. Sig. Proc.* **14**, 283–286 (2007).
- [31] Richmond, C. D., “Mean-squared error and threshold SNR prediction of maximum-likelihood signal parameter estimation with estimated colored noise covariances,” *IEEE Trans. Inf. Theory* **52**, 2146–2164 (2006).
- [32] Rife, D. C. and Boorstyn, R. R., “Single-Tone Parameter Estimation from Discrete-Time Observations,” *IEEE Trans. Inf. Theory* **20**, 591–598 (1974).
- [33] Rospars, J.-P., Lansky, P., Chaput, M., and Viret, P., “Competitive and noncompetitive odorant interaction in the early neural coding of odorant mixtures,” *J. Neurosci.* **28**, 2659–2666 (2008).
- [34] Rospars, J.-P., Lansky, P., Tuckwell, H. C., and Vermeulen, A., “Coding of odor intensity in a steady-state deterministic model of an olfactory receptor neuron,” *J. Comput. Neurosci.* **3**, 51–72 (1996).
- [35] Sachs, M. B. and Abbas, P. J., “Phenomenological model for two-tone suppression,” *J. Acoust. Soc. Am.* **60**, 1157–1163 (1976).
- [36] Sarpeshkar, R., “Analog Versus Digital: Extrapolating from Electronics to Neurobiology,” *Neural Comput.* **10**, 1601–1638 (1998).
- [37] Seidman, L. P., “Performance limitations and error calculations for parameter estimation,” *Proc. IEEE* **58**, 644–652 (1970).
- [38] Seriès, P., Latham, P. E., and Pouget, A., “Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations,” *Nat. Neurosci.* **7**, 1129–1135 (2004).
- [39] Seriès, P., Stocker, A. A., and Simoncelli, E. P., “Is the homunculus “aware” of sensory adaptation?” *Neural Comput.* **21**, 3271–3304 (2009).
- [40] Seung, H. S. and Sompolinsky, H., “Simple models for reading neuronal population codes,” *Proc. Natl. Acad. Sci. U.S.A.* **90**, 749–753 (1993).

- [41] Shadlen, M. N. and Newsome, W. T., “Noise, neural codes and cortical organization,” *Curr. Opin. Neurobiol.* **4**, 569–579 (1994).
- [42] Shadlen, M. N. and Newsome, W. T., “The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding,” *J. Neurosci.* **18**, 3870–3896 (1998).
- [43] Shannon, C. E., “Communication in the presence of noise,” *Proc. IRE* **37**, 10–21 (1949).
- [44] Silver, R. A., “Neuronal arithmetic,” *Nat. Rev. Neurosci.* **11**, 474–489 (2010).
- [45] So, K., Koralek, A. C., Ganguly, K., Gastpar, M. C., and Carmena, J. M., “Assessing functional connectivity of neural ensembles using directed information,” *J. Neural Eng.* **9**, 026004 (2012).
- [46] Sompolinsky, H., Yoon, H., Kang, K., and Shamir, M., “Population coding in neuronal systems with correlated noise,” *Phys. Rev. E* **64**, 051904 (2001).
- [47] Stein, R. B., Gossen, E. R., and Jones, K. E., “Neuronal variability: noise or part of the signal?” *Nat. Rev. Neurosci.* **6**, 389–397 (2005).
- [48] van Trees, H. L. and Bell, K. L., *Bayesian bounds for Parameter Estimation and Nonlinear Filtering/Tracking* (John Wiley and Sons, New York, 2007).
- [49] van Trees, H. L. and Bell, K. L., *Detection, Estimation and Modulation Theory, Part I* (John Wiley and Sons, New York, 2013).
- [50] Tuckwell, H. C., *Introduction to Theoretical Neurobiology, Vol. 2* (Cambridge University Press, New York, 1988).
- [51] van der Vaart, A. W., *Asymptotic statistics* (Cambridge University Press, Cambridge, UK, 2000).
- [52] Winslow, R. L. and Sachs, M. B., “Single-tone intensity discrimination based on auditory-nerve rate responses in background of quiet, noise, and with stimulation of the crossed olivocochlear bundle,” *Hearing Res.* **35**, 165–190 (1988).
- [53] Xie, X., “Threshold behaviour of the maximum likelihood method in population decoding,” *Netw. Comp. Neural Syst.* **13**, 447–456 (2002).
- [54] Yaeli, S. and Meir, R., “Error-based analysis of optimal tuning functions explains phenomena observed in sensory neurons,” *Frontiers in Comput. Neurosci.* **4**, 130 (2010).
- [55] Yarrow, S., Challis, E., and Seriès, P., “Fisher and Shannon information in finite neural populations,” *Neural Comput.* **24**, 1740–1780 (2012).
- [56] Zhang, K., Ginzburg, I., McNaughton, B. L., and Sejnowski, T. J., “Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells,” *J. Neurophysiol.* **79**, 1017–1044 (1998).
- [57] Zhang, K. and Sejnowski, T. J., “Neuronal tuning: to sharpen or broaden?” *Neural Comput.* **11**, 75–84 (1999).