RESEARCH ARTICLE

Efficient population coding depends on stimulus convergence and source of noise

Kai Röth [1,2], **Shuai Shao**[1,3], **Julijana Gjorgjieva**[1,2] *

1 Computation in Neural Circuits Group, Max Planck Institute for Brain Research, Frankfurt, Germany,
2 School of Life Sciences, Technical University of Munich, Freising, Germany, **3** Donders Institute and Faculty of Science, Radboud University, Nijmegen, Netherlands

* gjorgjieva@brain.mpg.de

Abstract

Sensory organs transmit information to downstream brain circuits using a neural code comprised of spikes from multiple neurons. According to the prominent efficient coding framework, the properties of sensory populations have evolved to encode maximum information about stimuli given biophysical constraints. How information coding depends on the way sensory signals from multiple channels converge downstream is still unknown, especially in the presence of noise which corrupts the signal at different points along the pathway. Here, we calculated the optimal information transfer of a population of nonlinear neurons under two scenarios. First, a lumped-coding channel where the information from different inputs converges to a single channel, thus reducing the number of neurons. Second, an independent-coding channel when different inputs contribute independent information without convergence. In each case, we investigated information loss when the sensory signal was corrupted by two sources of noise. We determined critical noise levels at which the optimal number of distinct thresholds of individual neurons in the population changes. Comparing our system to classical physical systems, these changes correspond to first- or second-order phase transitions for the lumped- or the independent-coding channel, respectively. We relate our theoretical predictions to coding in a population of auditory nerve fibers recorded experimentally, and find signatures of efficient coding. Our results yield important insights into the diverse coding strategies used by neural populations to optimally integrate sensory stimuli in the presence of distinct sources of noise.

Author summary

Information about the external environment is processed by populations of sensory neurons using a combinatorial neural code comprised of spikes from multiple neurons. In many cases, this information has to be compressed before reaching downstream circuits due to limitations on axonal transmission properties and the metabolic cost of firing. A prominent theoretical framework known as efficient coding postulates that sensory populations transmit maximal information about sensory stimuli subject to metabolic constraints. Although used to successfully predict the shape of receptive fields in the insect

and mammalian retina, this theory has often considered linear neurons or a single source of noise. Here, we derived a framework for information transfer in populations of noisy nonlinear neurons, specifically addressing the implications of the location of the noise source, and the way the sensory stimulus converges downstream, on the efficiency of information transfer. Our theoretical predictions support efficient coding in experimentally recorded populations of auditory nerve fibers coding for sound intensity.

Introduction

Neurons in sensory organs encode information about the environment and transmit it to downstream circuits in the brain. In many sensory systems, the sensory signal is not coded merely by individual neurons but rather by the joint activity of populations of neurons, which likely coordinate their responses to represent the stimulus as efficiently as possible. One signature of this efficient parallel coding might be the remarkably diverse response properties exhibited by many first-order sensory neurons. In the visual pathway, for example, the first-order sensory neurons are the retinal ganglion cells (RGCs) which send information to the thalamus through the optic nerve. There exist around thirty different RGC types which encode different visual features as characterized by the cells' spatiotemporal receptive fields and nonlinear computations [1–3]. Yet, there are also RGC types which in parallel encode a single stimulus feature differing in their firing thresholds [4–6], and hence provide parallel information streams. Another example is the first synapse level of the auditory pathway, where each inner hair cell transmits information about sound intensity to approximately ten to thirty different auditory nerve fibers (ANFs) [7]. ANFs differ in several aspects of their responses, including spontaneous rates and firing thresholds [8]. However, each fiber receives exclusive input from only a single inner hair cell. As in the retina, this results in a highly parallelized stream of sensory information. Similarly, this parallel encoding of a single stimulus feature with a population of neurons with different thresholds has been shown in olfactory receptor neurons [9], in mammalian touch receptors [10], and electro receptors of electric fish [11].

We asked whether the diverse response properties of a population of neurons encoding a single stimulus feature are a consequence of the evolutionary pressure of the sensory system to efficiently encode sensory stimuli. A powerful theoretical framework to address this question is the efficient coding. This framework postulates that during evolution sensory systems have optimized information encoding given biophysical and metabolic constraints. Predictions from efficient coding are consistent with many properties of primary sensory neurons, including center-surround receptive fields [12] and a split into ON and OFF pathways in the retina [13, 14], as well as the input-output functions of neurons [15] and sensory adaptation to changing stimulus statistics in the insect retina [16]. Applying the efficient coding framework requires determining a set of constraints that are relevant for the sensory system in question. Rather than investigating efficient coding in a specific sensory system, we sought to derive a general theoretical framework that applies to multiple sensory systems focusing on two questions: first, how the source and size of noise affects the accuracy of information coding, and second, how downstream signal convergence influences the optimality of information transfer.

Noise is a ubiquitous phenomenon in biological information processing and corrupts signal transmission at different processing stages (reviewed in [17]). The size [18–24] and source [25, 26] of noise can have distinct effects on signal encoding. For instance, previous studies have shown that neural populations adopt a strategy of independent coding [5, 14, 25, 27–30] or

decorrelation [12, 31–37] in conditions of low noise, and a strategy of redundant coding in the presence of high noise [5, 12, 14, 25, 27–35, 37, 38]. For populations of neurons, redundant coding can be interpreted as multiple neurons in the population, which acquire the same response thresholds to average out uncertainties in stimulus representation due to noise. When noise is negligible, the individual thresholds are expected to be distinct from each other, which is in agreement with experimental data [5, 28, 30]. Therefore, the source and size of noise have nontrivial influences on the encoding of sensory information.

Besides the source of noise, a second factor when maximizing information between stimulus and response is how the stimulus converges downstream after it is encoded by the neural population. For instance, it might be advantageous to compress sensory information before reaching downstream circuits due to axonal transmission limitations and the metabolic cost of firing of multiple neurons. Previous studies have assumed a framework in which the spiking output of the neurons converges, or is lumped, into one single output variable [29, 30]. In contrast, other works have assumed a framework without signal convergence, i.e. where the signal is encoded by the independent spiking output of each neuron in the population [5, 14, 25, 39]. Therefore, signal convergence also fundamentally influences optimal population coding.

Here, we investigated efficient stimulus coding in populations of more than two neurons as a function of the source and size of noise, and the type of stimulus convergence, and generated novel predictions about how these two aspects affect the optimal coding strategies of the neural populations. In particular, we maximized Shannon's mutual information between a one-dimensional stimulus (a single stimulus feature) and the population's response. Typically, the efficient coding framework has been applied to populations of linear neurons, where the contribution of the noise entropy term to the mutual information has been ignored, resulting in degeneracies in the optimal solutions [12, 13, 37, 40]. We considered a nonlinear version of efficient coding that includes two types of noise: additive input noise which corrupts the stimulus before it enters a neuron's nonlinearity and output noise implemented as spike generating noise which affects the output of the nonlinearity. We found that the exact implementation and source of noise can have fundamental implications for the conclusions arising from efficient coding.

For biologically realistic intermediate sources of noise, the particular downstream convergence of the sensory signal, either through lumping or independence, determines the number of distinct population thresholds. In agreement with previous studies [5, 14, 25, 27–30], we found that for low noise levels the optimal population thresholds are all distinct, while for high noise levels all thresholds become identical. However, unlike other studies we found surprising transitions from all thresholds being distinct at low noise to all thresholds being equal at high noise, which happen through a set of bifurcations at critical noise levels. These bifurcations resemble first- and second-order phase transitions in the case of the lumped and independent output variables, respectively. We related these phase transitions to the curvature of the information landscape, giving us insights into the optimal coding solutions and their relationship to well-understood critical systems in physics. We also compared our theoretical predictions to optimal coding of experimentally recorded auditory nerve fibers where we found signatures of optimality.

Results

Theoretical framework

We studied a population of spiking neurons encoding a sensory stimulus under different noise scenarios. A population of N neurons encodes a static, one-dimensional stimulus s drawn from a stimulus distribution $P(s)$ through the spike counts $\{k_1, \ldots k_N\} \equiv \vec{k}$ emitted in a coding
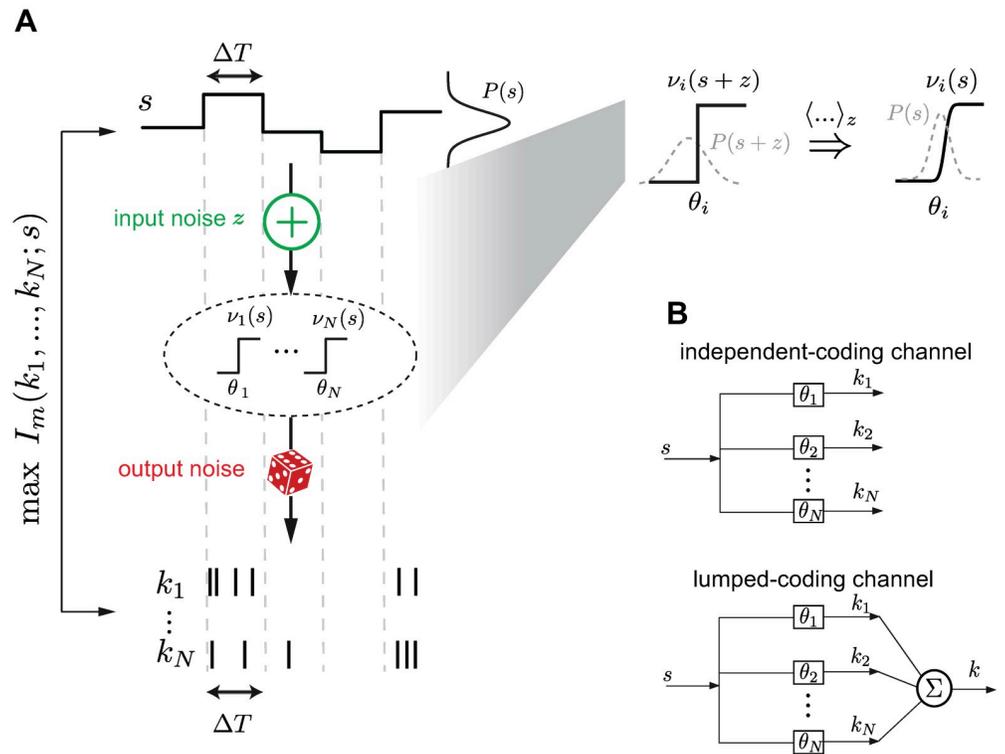
Fig 1. Stimulus encoding with a population of neurons in the presence of input and output noise. A. Framework: A static stimulus s (top) is encoded by a population of spike counts $\{k_1, \ldots k_N\}$ (bottom) in a coding time window ΔT. The stimulus is first corrupted by additive input noise z and then processed by a population of N binary nonlinearities $\{\nu_1, \ldots \nu_N\}$. Stochastic spike generation based on Poisson output noise corrupts the signal again. Thresholds $\{\theta_1, \ldots, \theta_N\}$ of the nonlinearities are optimized such that the mutual information $I_m(k_1, \ldots, k_N; s)$ between stimulus and spike counts is maximized. Inset: Introducing additive input noise and a binary nonlinearity can be interpreted as having a sigmoidal nonlinearity after the input noise is averaged, $\langle \ldots \rangle_z$. Shallower nonlinearities result from higher input noise levels. **B**. Two different scenarios of information transmission: In the independent-coding channel each neuron contributes with its spike count to the coding of the stimulus, while in the lumped-coding channel all spike counts are added into one scalar output variable that codes for the stimulus.

time window ΔT (Fig 1A). We considered different stimulus distributions, parametrized by the generalized normal distribution [41], but here we primarily discuss the case of a Gaussian stimulus distribution (see Methods). The mapping from the stimulus value s to the spike count vector $\vec{k}$ happens through a set of N nonlinear functions (tuning curves) $\{\nu_1(s), \ldots \nu_N(s)\}$, where $\nu_i(s)$ denotes the firing rate of the respective neuron i. Such a mapping can be implemented by a variety of sensory systems, for instance, the retina which processes various visual stimulus attributes, such as light intensity or contrast [42], the olfactory receptor neurons which process a range of concentrations of a single odor [9, 39, 43, 44], or the auditory nerve fibers (ANFs) which transmit information about sound pressure levels [45]. Here, we only focus on optimizing one stage of this transformation, namely the nonlinearity which takes a filtered stimulus s as input and converts it into action potentials $\vec{k}$. For example, if we assume that our sensory system of interest is the population of ANFs, which are highly nonlinear processing units [46], then s represents the stimulus value following preprocessing by the cochlea and the inner hair cells.

We modeled the neurons' tuning curves as binary, described by two firing rate levels $\{0, \nu_{\max}\}$ with an individual threshold θ_i separating the stimuli into two firing rates. Thus, the

input-output functions of each neuron can be represented by $v_i(x) = v_{max}\Theta(\theta_i - x)$, where Θ is the Heaviside function. This simplification is justified by the fact that many sensory neurons have been described with steep tuning curves that resemble binary neurons [16, 27, 36], and it makes the problem mathematically traceable. We derived the number and values of distinct thresholds in the population when the signal is corrupted by two sources of noise: input noise, which affects the signal before the nonlinearity, and output noise, which affects the neuronal outputs after the nonlinearity.

Input noise.   Before being processed by the nonlinearity, the stimulus s is corrupted by additive noise z drawn from a distribution $P(z)$. The size of input noise can be quantified by the ratio of its variance $\langle z^2 \rangle \equiv \sigma^2$ to the stimulus variance $\langle s^2 \rangle \equiv \sigma_s^2$. Without loss of generality, we set $\sigma_s^2 = 1$ and thus σ^2 alone stands for the size of input noise. The noise affects the stimulus independently for each nonlinearity. We did not consider correlated noise since previous work has shown that the case of correlated noise can be reduced to independent noise with lower σ^2 for a model of two neurons and Gaussian stimuli [25]. Higher-order correlations in the case of more than two neurons might not be trivially reduced to non-correlated input noise with smaller σ^2; however, such investigations are beyond the scope of our paper. Similarly to the stimulus distribution, we primarily examined the case with the noise drawn from a Gaussian distribution, $z \sim N(0, \sigma^2)$, but we also considered other distributions (see Methods). Since the input to the nonlinearities is $x = s + z$, the effective tuning curves, $v_i(s)$, can be described to have sigmoidal shape (Fig 1A, inset). A larger input noise size, determined by the variance of the noise σ^2, corresponds to a shallower slope. In the remainder of the text, we use the standard deviation σ to refer to the size of input noise.

Output noise.   Output noise was implemented by generating output spikes stochastically; here, each of the spike counts k_i in a coding window ΔT was Poisson distributed. Large output noise corresponds to the case when the product of v_{max} and ΔT is small; in this case the output of a given cell i is often $k_i = 0$ making it more difficult to distinguish whether the underlying firing rate for that neuron is 0 and thus the stimulus is smaller than the threshold θ_i, or whether the firing rate is v_{max} and the stimulus greater than θ_i. The output noise size can thus be quantified by the expected spike count for maximum firing rate, $R := v_{max}\Delta T$, where small R means high noise.

Within this framework, we maximized the mutual information between stimulus and output spike counts, and optimized the number and values of distinct thresholds, $\{\theta_i\}$, of the neuronal nonlinearities, while varying the size of input and output noise (see Methods). We chose the mutual information as the objective function to quantify the optimality of the encoding because it does not rely on any specific assumptions of how this information should be decoded, and presents an upper bound for any other efficiency measure [47]. In addition to two noise sources, we further distinguish between two different scenarios previously considered in the literature for how the sensory signal converges after being processed by the population of neurons (Fig 1B): (1) an *independent-coding channel* where a vector of spike counts $\vec{k} = \{k_i\}$ generates a population code of the stimulus where each spike count independently contributes to the total information [5, 14, 25], and (2) a *lumped-coding channel* where a scalar output variable $k = \Sigma_i k_i$, obtained by summing the individual spike counts k_i, codes for the stimulus [29, 30].

The independent-coding channel transmits more information than the lumped-coding channel

To understand how stimulus convergence influences information transmission in larger neural populations of more than two neurons, we compared the mutual information and optimal

Spike count			Estimation probab. $P(s \in \Delta_j | k, \vec{k})$		
k	k_1	k_2	$P(s \in \Delta_1)$	$P(s \in \Delta_2)$	$P(s \in \Delta_3)$
0	0	0	1	0	0
R	R	0	0	1	0
$2R$	R	R	0	0	1
1	1	0	0	α	$1 - \alpha$
1	0	1	0	0	1
1	?	?	0	α'	$1 - \alpha'$
0	0	0	β	-	$1 - \beta$
1	1	0	0	-	1
1	0	1	0	-	1

A vanishing output noise $(R \to \infty)$ — independent and lumped

B intermediate output noise assume $k = 1$ — independent / lumped

C high output noise $(R \to 0)$ — independent and lumped

— independent
— lumped
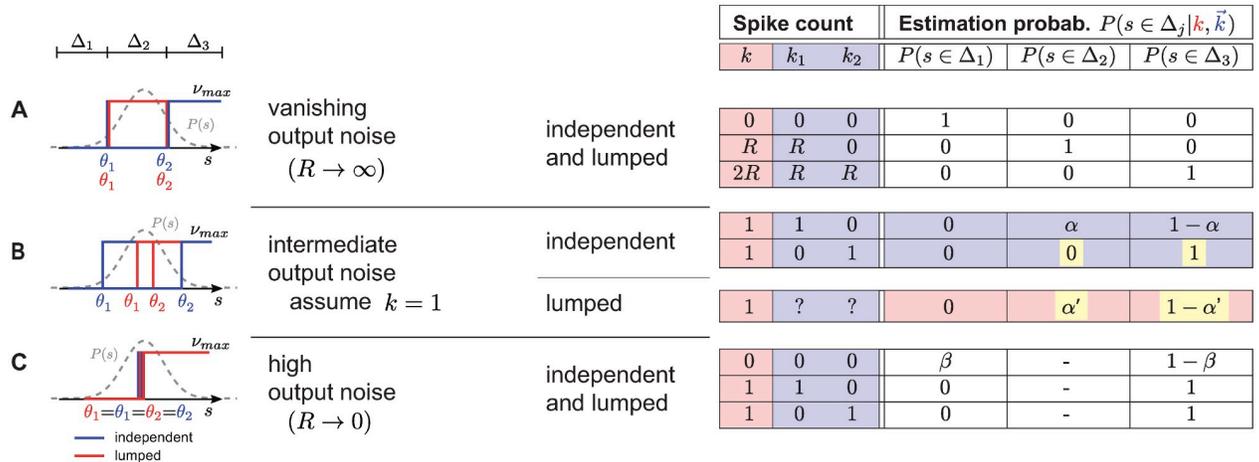
Fig 2. Schematic illustrating the dominance in information of the independent- over the lumped-coding channel. Here, we treat the case of $N = 2$ cells, vanishing input noise ($\sigma = 0$) and **A**. vanishing output noise ($R \to \infty$), **B**. intermediate output noise when the total number of spikes $k = k_1 + k_2 = 1$, and **C**. high output noise ($R \to 0$). Left: The relative positions of optimal thresholds of both the independent- (blue) and lumped-coding (red) channels. Right: The stimulus "estimation probabilities" $P(s \in \Delta_j | \vec{k})$ for the two different channels. Yellow shading shows where the noise entropy is higher in the lumped-coding channel. α, α', and β denote non-zero probability values (see text).

thresholds between the lumped- and independent-coding channel scenarios in the presence of two noise sources.

To first gain intuition, we illustrate the case with vanishing input noise ($\sigma = 0$) and a population with two neurons with thresholds $\theta_1 < \theta_2$, which divide the entire stimulus distribution into three regions: Δ_1: $s < \theta_1$, Δ_2: $\theta_1 \leq s < \theta_2$ and Δ_3: $s \geq \theta_2$ (Fig 2, left). Here, we computed all possible spike counts and corresponding "estimation probabilities," $P(s \in \Delta_i | \vec{k})$, which describe the probability of the stimulus being in each of the three regions $\{\Delta_i\}_{i=\{1,2,3\}}$ for a given spike count $\vec{k}$ (Fig 2). These vary as a function of output noise, and we considered three cases: high, intermediate, and negligible output noise. First, in the limit of vanishing output noise where $R = v_{max}\Delta T$ is very large, the information encoded by both channels is identical because with optimal thresholds both reach capacity and transmit $\log_2(3)$ bits of information (Fig 2A). In particular, whenever the stimulus is larger than the threshold of a given cell, that cell will on average fire R spikes. Since $R \to \infty$, for that given cell the probability of having 0 spikes is infinitesimal. This unambiguously determines the stimulus region $\{\Delta_i\}$ in which the stimulus occurs. Hence, the estimation probabilities all become either 0 or 1, leading to identical output entropy for both coding channels, and consequently identical mutual information with zero noise entropy.

For intermediate output noise, the independent- and the lumped-coding channels have distinct estimation probabilities. Although in principle the number of emitted spikes can be anything, let us consider the example where the total number of spikes is 1 ($k_1 + k_2 = 1$, Fig 2B). We demonstrate that the lumped-coding channel loses information because knowledge about the identity of which individual cell spiked is lost. For example, if the cell with higher threshold θ_2 fires a spike, this implies with certainty that the stimulus is greater than θ_2. The lumped-coding channel fails to encode this information since in principle the spike could have been emitted by the cell with lower threshold θ_1. Thus, the estimation probabilities α' and $1 - \alpha'$ for the stimulus being below or above θ_2, respectively, are nonzero. For the independent-coding channel, however, the corresponding estimation probabilities α and $1 - \alpha$ are nonzero if the cell with the lower threshold θ_1 fires a spike. Therefore, for the independent-coding channel

there are more cases in which the uncertainty is resolved, leading to higher mutual information. As an example, for output noise of $R = 2.5$, the mutual information for the independent- and lumped-coding channel is 1.30 and 1.01 bits, respectively (when the thresholds are optimized).

For very high output noise, $R \rightarrow 0$, the expected spike count of either of the cells is very small, even when the stimulus is larger than the respective threshold with the resulting firing rate $v_{\max}$ (Fig 2C). This means that most of the time the observed spike count of each cell is 0, rarely 1, and never 2—the probability of observing more than one spike is infinitesimal. In this high noise regime, the optimal solution for both the independent- and lumped-coding channels is to make both thresholds identical, i.e. $\theta_1 = \theta_2$. Therefore, the intermediate regime Δ_2 does not exist and the two possibilities of having a spike from either cell are equivalent. Thus, if the observed spike count is 1, then there is no possibility of error for either channel. Similarly, if the observed spike count is 0, the two different estimation probabilities are the same for both channels, namely β and $1 - \beta$ for the stimulus being above or below $\theta_{1,2}$, respectively. This results in identical mutual information between stimulus and response for both channels.

Equipped with this intuition, we computed the maximal mutual information by optimizing thresholds for the lumped- and independent-coding channels for a population of three neurons where we varied both the input and output noise continuously (Fig 3A and 3B). We again found that the independent-coding channel overall transmits more information than the
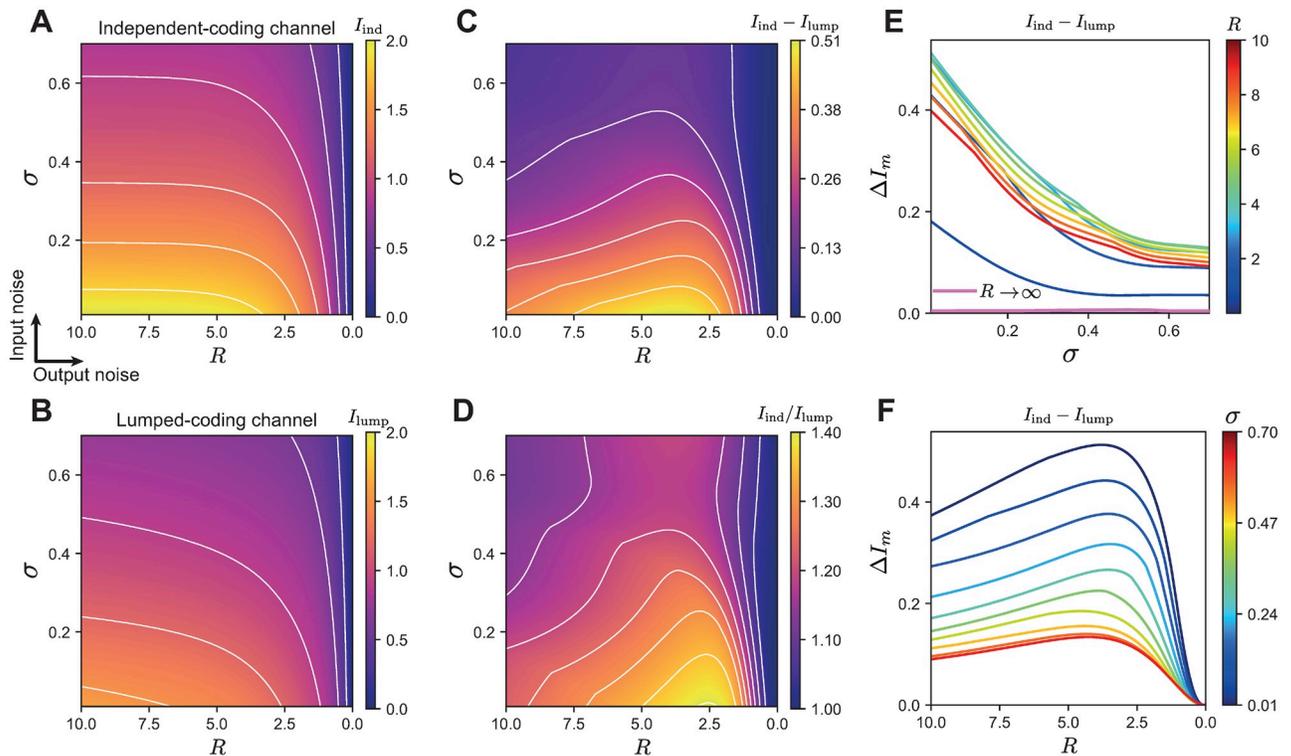
Fig 3. Maximized mutual information for the lumped- and independent-coding channels for a population of three neurons. A. Information of the independent-coding channel for different combinations of output noise R and input noise σ. Contours indicate constant information. **B**. Information of the lumped-coding channel. **C**. Absolute information difference between the two coding channels. **D**. Information ratio between the two coding channels. Both C and D show a region of intermediate output noise where the independent-coding channel substantially outperforms the lumped-coding channel. **E**. Information difference depending on input noise σ for various levels of output noise R, corresponding to vertical slices from C. We also include the special case of zero output noise, $R \rightarrow \infty$. **F**. Information difference depending on output noise R for various levels of input noise σ, corresponding to horizontal slices from C.

lumped-coding channel. Additionally, we found that the increase of information with smaller output noise (higher R) saturates faster in the case of the independent-coding channel, as can be seen by the flattening of the contour lines as R increases (compare Fig 3A and 3B). The optimal thresholds that lead to these values of maximal mutual information are shown in Fig 4 and described in the next section.

We quantified the ratio and the absolute difference in information transmission between the two channels (Fig 3C and 3D). For all finite input and output noise levels, the independent-coding channel outperforms the lumped-coding channel since the contribution of each neuron to the overall spike count provides additional information about the stimulus that is lost by summing all the spike counts through lumping. The information loss is the largest at intermediate levels of output noise and low levels of input noise; for instance, at $R \approx 2.5$ and $\sigma \approx 0$ the independent-coding channel transmits up to 40% more information than the lumped-coding channel (Fig 3D).

To best visualize these differences, we fixed one source of noise and varied the other. In the special case of zero output noise ($R \to \infty$), the two channels transmit almost the same information. For finite output noise R, the information loss in the lumped-coding channel relative to the independent-coding channel monotonically decreases as a function of the input noise, σ (Fig 3E). The difference in information transmitted by the independent- and lumped-coding channels as a function of the output noise R for fixed input noise σ demonstrates that the information loss due to lumping is a non-monotonic function of output noise R (Fig 3F), with the largest loss occurring in the biologically realistic range of intermediate noise [36, 48]. This non-monotonicity can be explained by the fact that in the limit of very large or very small output noise the lumped-coding channel transmits as much information as the independent-coding channel (Figs 2, 3E and 3F).

In summary, we found that in the presence of both input and output noise, the lumped-coding channel transmits less information than the independent-coding channel, and we can intuitively understand these trade-offs in a small population of two neurons.

Optimal thresholds for the independent- and lumped-coding channels

We computed the optimal population thresholds at which the spiking output of the populations achieves maximal information about the stimulus. We first discuss the case with three neurons. For both the independent- and lumped-coding channel, the optimal number of distinct thresholds in the population depends on the source and level of noise (Fig 4). When both sources of noise are negligible, the optimal number of thresholds is three, representing a fully diverse population where all thresholds are distinct. However, when both input and output noise are high, the optimal number of thresholds in the population is one, representing a fully redundant population where all thresholds are identical. The most interesting cases arise at intermediate input and output noise levels, where we found two distinct optimal thresholds. To gain a better understanding of the transition between different threshold regimes as a function of noise, we fixed one level of noise and examined the thresholds as a function of the other noise level.

We found that the number of distinct thresholds in the population generally decreases with increasing input or output noise through a set of bifurcations. We call the noise levels at which these bifurcations in the thresholds appear *critical* noise levels. We found that for the lumped-coding channel threshold bifurcations occur at lower noise levels compared to the independent-coding channel. This result makes intuitive sense because lumping multiple information pathways into a single coding channel reduces the possible values of the encoding variable and increases the noise entropy, and therefore acts like an additional noise source.
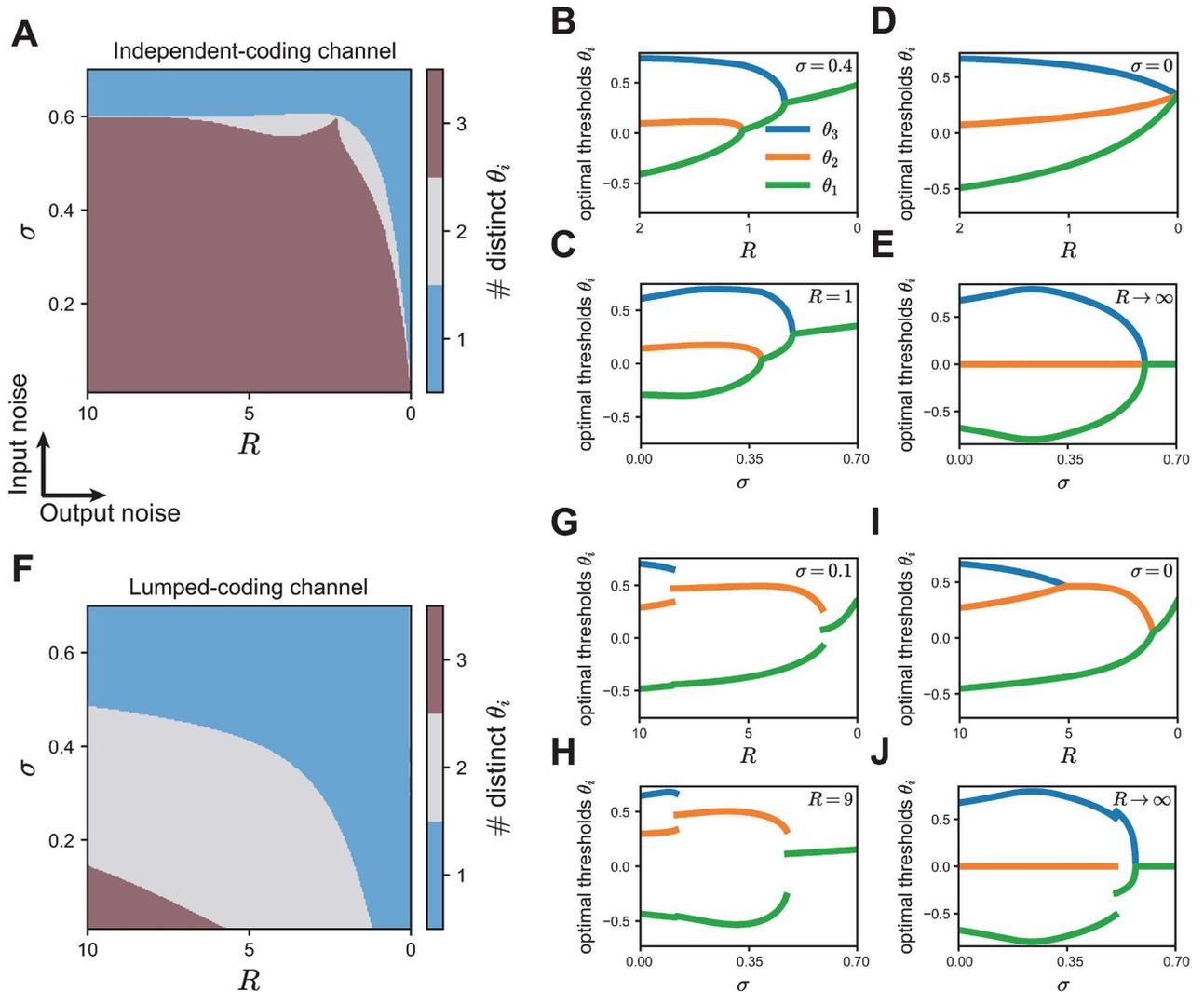
Fig 4. Optimal thresholds for the independent- and lumped-coding channels. Optimal thresholds for the independent-coding channel (A-E) compared to the lumped-coding channel (F-J) for a population of three neurons. **A**. The optimal number of distinct thresholds depends on input noise σ and output noise R. **B**. The optimal thresholds as a function of output noise for a fixed value of input noise ($\sigma = 0.4$). **C**. The optimal thresholds as a function of input noise for a fixed value of output noise ($R = 1$). **D**. The optimal thresholds as a function of output noise in the limit of no input noise ($\sigma = 0$). **E**. The optimal thresholds as a function of input noise in the limit of vanishing output noise ($R \to \infty$). **F-J**. As (A-E) but for the lumped-coding channel. Intermediate noise levels where bifurcations occur in (G,H) take smaller values of R and σ in the lumped- that in the independent-coding channel since lumping itself acts like a source of noise (G: $\sigma = 0.1$, H: $R = 9$).

For the independent-coding channel (Fig 4A), the thresholds become distinct from each other gradually, in the sense that the differences between the optimal thresholds change continuously, both as a function of output noise when the input noise level is fixed (Fig 4B) and also as a function of input noise when the output noise level is fixed (Fig 4C). In the case when one source of noise is zero, these bifurcations represent the transition from all optimal thresholds being distinct directly to the state where all optimal thresholds are identical, without an intermediate state where two thresholds are the same (Fig 4D and 4E). For instance, in the absence of input noise ($\sigma = 0$), the population's thresholds are all distinct from each other for all finite ranges of output noise except when $R \to 0$ (Fig 4D). In the absence of output noise ($R \to \infty$), there is a critical value $\sigma_{\text{crit}} > 0$ at which the population transitions directly from all

thresholds being distinct to all thresholds being equal (Fig 4E). Note that for all these bifurcations the threshold differences change continuously, i.e. there are no jumps of optimal threshold values with varying noise.

Surprisingly, we found a small range of input noise, $0.54 < \sigma < 0.6$, for which we observed a non-monotonic change in the number of optimal thresholds when varying the output noise R (S1 Fig). A similar result has been observed when optimizing Fisher information—a different, local measure for information—in a population of bell-shaped tuning curves in a model of optimal coding of interaural time differences in the auditory brain stem [28].

In comparison, for the lumped-coding channel, the bifurcations occur as the threshold differences at critical noise values change abruptly, or discontinuously, when one noise source varies and the other remains fixed (Fig 4G and 4H). Here, the system has an intermediate number of thresholds for a large range of noise values, and the transition from one to three distinct thresholds is not simultaneous as either noise vanishes. Rather, the discontinuous threshold jumps at each bifurcation become continuous in the absence of input noise (Fig 4I), as normally seen for the independent-coding channel, or partly continuous in the absence of output noise (Fig 4J). These two scenarios agree with two previous studies, where a lumped-coding channel was studied with only output noise [30], or with only input noise [29]. Our results are also consistent with previous studies for small populations of two neurons and only one source of noise [5, 14, 27], large populations with only output noise [39] and two-neuron populations with multiple noise sources [25]. We show that similar patterns of how the number of distinct thresholds evolve as a function of two different noise sources for the independent-lumped-coding channels also hold for larger neural populations (S2 Fig).

Taken together, our theory derives different configurations of optimal thresholds in populations of more than two noisy neurons that depend on how the sensory stimulus is combined to produce spiking output and the location of the noise that corrupts the signal.

Optimal threshold differences represent order parameters in phase transitions

The characteristic bifurcations of the optimal thresholds at critical noise levels suggest the occurrence of phase transitions encountered in a variety of physical systems. In physics, a phase transition is defined by non-analytic behavior of the free energy—usually a discontinuity of its first or second derivative—and can be characterized by an order parameter [49]. For example, a phase transition occurs when the order parameter—which could among others be the density difference at the liquid-vapor critical point, or magnetization of a ferromagnetic material—changes abruptly from zero to non-zero values with an external parameter, such as pressure or temperature. Similarly, in chemistry, the order parameter that changes abruptly from zero to non-zero with temperature is the solubility of liquid mixtures.

Guided by this characterization, we sought to relate the qualitative differences in optimal thresholds of the independent- vs. lumped-coding channel with two noise sources to phase transition phenomena (Fig 4). We illustrate the results for a population with three neurons, and thus have two order parameters which are the two threshold differences, $\theta_2 - \theta_1$ and $\theta_3 - \theta_2$. To determine whether a phase transition occurs, we computed the first and second derivatives of the mutual information with respect to a given noise parameter (Fig 5). Using the Ehrenfest classification of phase transitions [50], a discontinuity in the first (second) derivative with respect to the noise implies a first- (second-) order phase transition.

We found that the orders of the phase transitions always correspond to the discontinuity of the threshold differences—being the order parameters—when noise varied. For continuous threshold bifurcations, there was a discontinuity in the second derivative with respect to
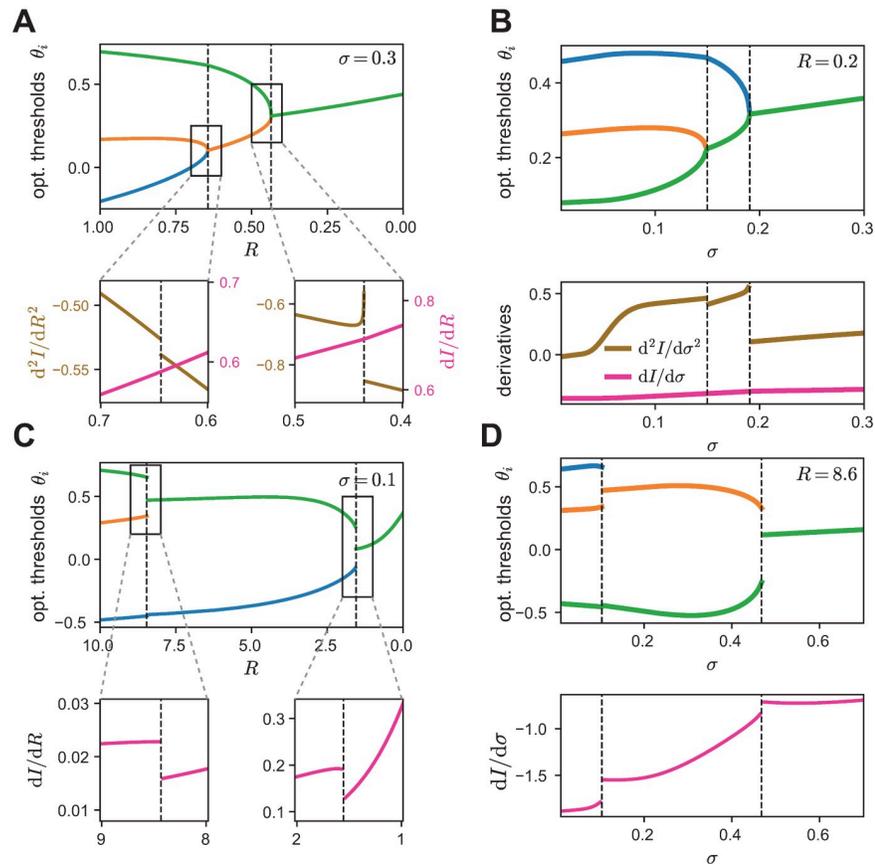
Fig 5. Threshold differences as phase transitions with respect to both noise sources. A. Optimal thresholds for the independent-coding channel depending on output noise R. Insets: The first derivative of the mutual information as a function of noise is continuous, while the second derivative is discontinuous at the critical noise values where the thresholds separate, implying a second-order phase transition. **B**. As in A, but with respect to input noise σ. **C**. Optimal thresholds as in A but for the lumped-coding channel. The first derivative is discontinuous at the critical noise values where the thresholds separate, implying a first-order phase transition. **D**. As in C but with respect to input noise σ.

output noise, thus corresponding to a second-order phase transition (Fig 5A). All phase transitions for the independent-coding channel were continuous and thus of second-order also with respect to input noise (Fig 5B). This result is in agreement with a previous study which also found a second-order phase transition in a population of two neurons in the presence of only input noise [5]; we extended this result to populations of more than two neurons and with more than one noise source. We next investigated phase transitions in the lumped-coding channel.

For discontinuous threshold bifurcations we observed a discontinuity in the first derivative with respect to output noise and thus a phase transition of first-order (Fig 5C). This is almost always the case for the lumped-coding channel, also with respect to input noise (Fig 5D). An exception to this is when one noise source vanishes, e.g. input noise (Fig 4I) or output noise (Fig 4J), for which the phase transitions are of second-order (S3 Fig).

Continuous phase transitions from different physical systems often behave very similarly around critical points (e.g. the Ising model at critical temperature or the liquid-gas transition at the critical point in the temperature-pressure plane [49]). This phenomenon is known as *universality* and the universality class to which a system belongs can be characterized by critical exponents [49]. For example, the critical exponent β classically describes how the the order

Table 1. Critical exponents as a function of the two noise sources. Critical exponents are obtained by fitting a monomial to the threshold differences or eigenvalues near the critical noise values (l denotes the left, and r the right side) as a function of each noise source (see S4 Fig).

Crit. exp.	Definition	Fitted value (mean ± SEM)
β_σ	$\Delta\theta \propto \|(\sigma - \sigma_c)/\sigma_c\|^{\beta_\sigma}, \sigma < \sigma_c$	0.5027 ± 0.0018
β_R	$\Delta\theta \propto \|(R - R_c)/R_c\|^{\beta_R}, R > R_c$	0.5018 ± 0.0023
$\phi_{\sigma,l}$	$\|\lambda\| \propto \|(\sigma - \sigma_c)/\sigma_c\|^{\phi_{\sigma,l}}, \sigma < \sigma_c$	1.0034 ± 0.0019
$\phi_{\sigma,r}$	$\|\lambda\| \propto ((\sigma - \sigma_c)/\sigma_c)^{\phi_{\sigma,r}}, \sigma > \sigma_c$	0.9977 ± 0.0005
$\phi_{R,l}$	$\|\lambda\| \propto \|(R - R_c)/R_c\|^{\phi_{R,l}}, R < R_c$	0.9967 ± 0.0015
$\phi_{R,r}$	$\|\lambda\| \propto ((R - R_c)/R_c)^{\phi_{R,r}}, R > R_c$	1.0023 ± 0.0025

https://doi.org/10.1371/journal.pcbi.1008897.t001

parameter behaves for small temperature changes close to (but below) the critical temperature. In our system with mutual information and noise, β describes the behavior of threshold differences for noise values slightly smaller than critical noise values σ_c, R_c, i.e. $\Delta\theta \propto |(\sigma - \sigma_c)/\sigma_c|^{\beta_\sigma}$ or $\Delta\theta \propto |(R - R_c)/R_c|^{\beta_R}$, respectively (S4(A) and S4(B) Fig). We obtained critical exponents for both noises sources by fitting a monomial to the positive part of the threshold differences depending on one noise value, while treating the other noise value as a parameter that we varied (Table 1). Similarly, we fitted the critical exponents for the eigenvalues which approach zero at critical noise values. Since the eigenvalues have finite values on both sides around the critical noise values, we separately fitted critical exponents for each side; for e.g. for the output noise R, we fitted $|\lambda| \propto |(R - R_c)/R_c|^{\phi_{R,l}}$ for $R < R_c$ and $|\lambda| \propto ((R - R_c)/R_c)^{\phi_{R,r}}$ for $R > R_c$ (S4(C) and S4(D) Fig; Table 1, l denotes the left, and r the right side). Hence, the critical exponent for the eigenvalues is approximately 1, while for the threshold differences as order parameters it is approximately 0.5, the value predicted by the mean field theory for all continuous phase transitions [51]. Since mean-field theory ignores statistical fluctuations, in most physical systems the measured exponents are different than the ones predicted by theory, and are referred to as "anomalous" exponents [49]. In our model, the mutual information already takes into account statistical fluctuations, and appears to be an analytic function of the thresholds (see Fig 6B and 6C). Therefore, we do not expect an analogous mechanism that would lead to anomalous scaling exponents. Our results extend previous theoretical work which considered a population of two neurons with only input noise and already reported a critical exponent close to 0.5 [5].

Together this shows that the threshold differences in the population of neurons represent order parameters and determine the order of the observed phase transitions: discontinuous threshold differences correspond to first-order phase transitions while continuous threshold differences correspond to second-order phase transitions. We provide an extensive comparison between bifurcations of the optimal threshold values and phase transitions observed in statistical mechanical models in the Discussion.

Characteristic shape of the information landscape at critical noise levels

To gain a better understanding of the information landscape, especially at the critical noise values at which threshold bifurcations appear, we examined the Hessian matrix of the mutual information, I_m, with respect to the thresholds, $\partial^2 I_m/(\partial\theta_i\partial\theta_j)$. The Hessian can be understood as an extension of the second derivative to higher-dimensional functions. The eigenvalues of the Hessian quantify the curvature of the information landscape in the direction of the respective eigenvectors, which themselves stand for the directions of principal curvatures in the
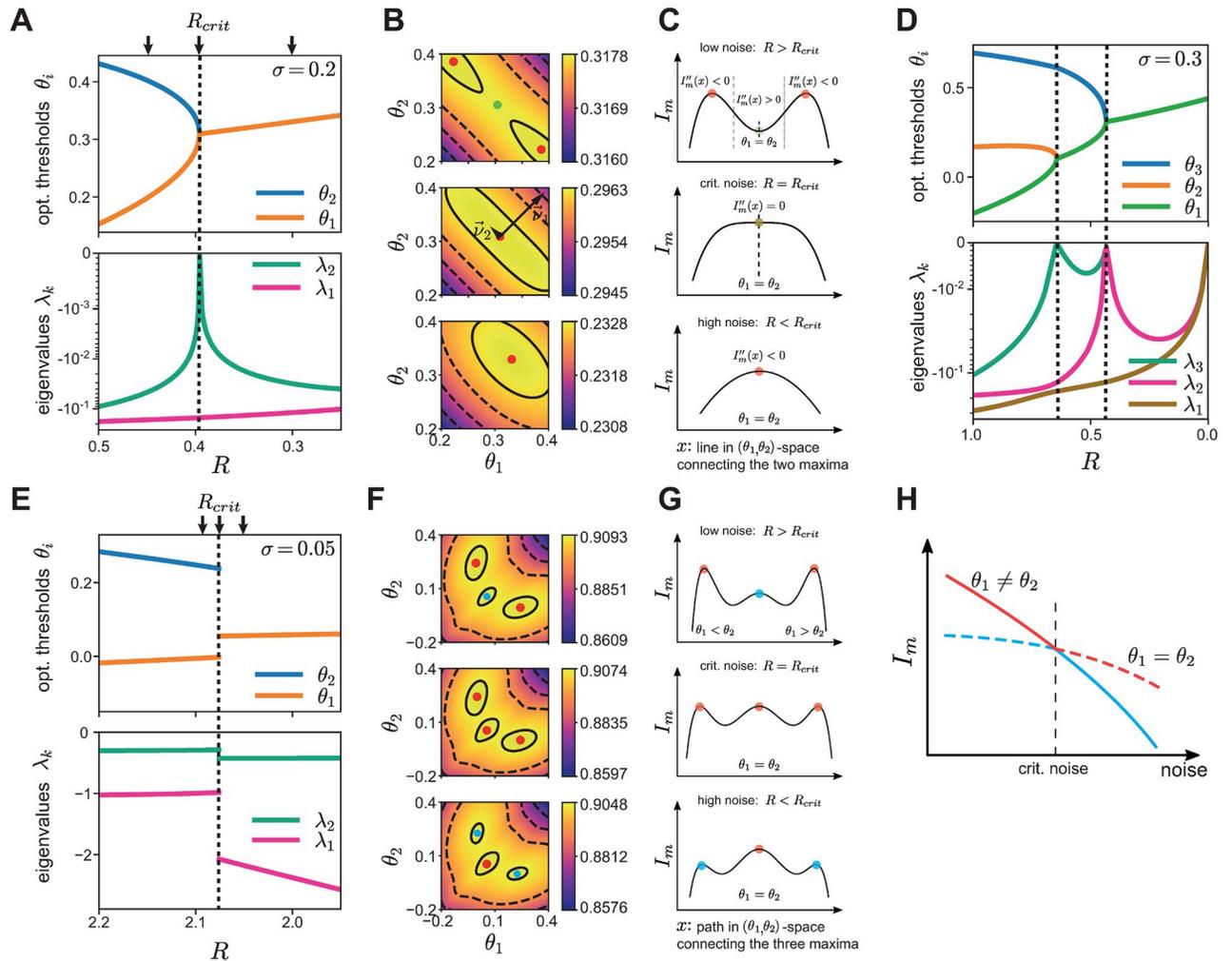
Fig 6. Information landscape for the independent- and lumped-coding channels undergoes different phase transitions around critical noise levels. A. Top: Optimal thresholds of the independent-coding channel for a population of two neurons as a function of output noise R. Bottom: Corresponding eigenvalues of the Hessian of the information landscape with respect to thresholds. At the critical noise value $R_{crit} \approx 0.396$ at which the threshold bifurcation occurs (vertical dashed line) one eigenvalue approaches zero. **B**. Information landscape $I_m(\theta_1, \theta_2)$ for the three output noise levels R indicated by arrows in A. Top: For $R > R_{crit}$, there are two equal global maxima. Middle: At $R = R_{crit}$, the eigenvectors of the Hessian are shown and scaled by the corresponding eigenvalue (the eigenvector with the smaller eigenvalue, $\vec{v}_2$, was artificially lengthened to show its direction). At the critical noise value the information landscape locally takes the form of a ridge. Bottom: For $R < R_{crit}$, there is one global maximum, meaning that the optimal thresholds are equal (bottom). **C**. The mutual information as a function of the line x in (θ_1, θ_2) space connecting the two maxima in B. Top: For $R > R_{crit}$ (low noise), there are two inflection points (dashed vertical lines) with zero curvature along the line x. The point with equal thresholds corresponds to a local minimum. Middle: At $R = R_{crit}$, the two maxima, the minimum, and the two inflection points merge into one point, thus the curvature is zero. Bottom: For $R < R_{crit}$, there is a single global maximum with a negative curvature. **D**. As in A but for a population with $N = 3$ neurons. **E**. As in A but for the lumped-coding channel. Both the optimal thresholds and the eigenvalues show a discontinuity at the critical noise level. **F**. Information landscape as in B for the lumped-coding channel and noise values indicated by arrows in E. Local optima are shown in cyan, global ones in red. **G**. Similar to C for the lumped-coding channel. Here the abscissa denotes the (non-straight) path connecting the three optima in F. **H**. Illustration of discontinuous threshold bifurcations, where the global maximum at $\theta_1 \neq \theta_2$ at low noise (red, solid) becomes a local maximum for high noise (cyan, solid), while $\theta_1 = \theta_2$ (dashed) becomes global. As their respective derivatives are different, there is a discontinuity in the first derivative when only taking the global maximum into account (red lines), corresponding to a first-order phase transition.

space defined by the thresholds. To gain intuition about the differences of the information landscape between the independent- and lumped-coding channels, we considered a population of two cells for which the landscape can be easily portrayed in two dimensions. However, the theory extends naturally to populations with more neurons and information landscapes in higher dimensions.

We first considered the independent-coding channel for a fixed level of input noise, while varying the output noise. At the critical noise level, R_{crit}, where the thresholds bifurcate, one eigenvalue of the Hessian decreases to zero (Fig 6A). The information landscape undergoes a transformation around the critical noise levels, from one with two distinct maxima separated by a local minimum at low noise, $R > R_{\mathrm{crit}}$ (Fig 6B, top), where the population thresholds are distinct, to one where there is a unique maximum at high noise, $R < R_{\mathrm{crit}}$, where the population thresholds are identical (Fig 6B, bottom). For $R > R_{\mathrm{crit}}$, there are two inflection points (Fig 6C, top), resulting in two different curvatures along the line that connects the two maxima. At the critical noise, $R = R_{\mathrm{crit}}$, the two maxima converge at the bifurcation point and the two inflection points fuse together such that the curvature becomes zero (Fig 6C, middle). At this point of convergence, the information landscape locally resembles a ridge, which extends along one principal direction of curvature (Fig 6B, middle). The ridge is perpendicular to the other principal direction, which stands for the direction of largest curvature. Finally, for $R < R_{\mathrm{crit}}$, the information landscape has a single maximum with a negative curvature (Fig 6B and 6C, bottom).

We then examined the eigenvalues of the Hessian matrix for a larger population of size $N>2$. We found that at each critical noise level where the thresholds bifurcate, at least one eigenvalue of the Hessian matrix approaches zero. The number of zero eigenvalues—denoting the number of dimensions along which the information does not change locally—is equal to the number of thresholds participating in a bifurcation minus one. For $N = 3$, for example, there are two critical noise values at which the thresholds bifurcate (Fig 6D, top). At one of these critical values, three thresholds are involved and thus the number of eigenvalues approaching zero is two, while at the other critical value only two thresholds are involved, and thus the number of eigenvalues approaching zero is one (Fig 6D, bottom). Locally, threshold combinations along the ridge of the information landscape achieve almost the same information. This ridge is a manifold of dimension $M - 1$, where M is the number of thresholds involved in the bifurcation. The manifold is locally given by

$$\sum_{\left\{ \substack{i \mid \theta_i \text{ involved} \\ \text{in the bifurcation}} \right\}} \theta_i = \text{constant.} \tag{1}$$

As a result, the ridge is oriented at exactly 45˚ with respect to all of the θ-directions participating in the bifurcation. For example, for $M = 2$ this manifold is a line, while for $M = 3$ it is a plane. Following the same argument as for the population with $N = 2$ neurons (Fig 6C), it can be shown that the curvature of the information landscape has to be zero in $M - 1$ principal directions, thus $M - 1$ eigenvalues of the Hessian have to be zero when M thresholds participate in a bifurcation of continuous manner.

For the lumped-coding channel, the eigenvalues of the Hessian do not approach zero at the critical noise levels where the thresholds split (Fig 6E). This is in agreement with the fact that threshold bifurcations are in general discontinuous for the lumped-coding channel (see also Fig 4G and 4H). An exception to this is the limiting case when one noise level is zero, where the lumped-coding channel shows continuous bifurcations and thus second-order phase transitions (S3 Fig). At low output noise, $R > R_{\mathrm{crit}}$ (Fig 6F and 6G, top), the information landscape has two distinct global maxima corresponding to the optimal thresholds, θ_1 and θ_2. However, the information landscape also has a local maximum at $\theta_1 = \theta_2$. As noise increases, this local maximum decreases more slowly compared to the two global maxima, until at the critical noise level R_{crit} the three maxima become equal (Fig 6F and 6G, middle). As noise increases further, $R < R_{\mathrm{crit}}$, the maximum at $\theta_1 = \theta_2$ becomes the single global maximum (Fig 6F and 6G, bottom). Therefore, the phase transition happens at the noise level where the local

maximum becomes the global one. This is a first-order phase transition since at this critical noise level the decrease of maximum information with noise changes abruptly, resulting in a discontinuity in the first derivative (Fig 6H).

Our results show, that for finite noise, the shape of the information landscape for the independent- and the lumped-coding channels can be uniquely related to the nature of the threshold bifurcations (continuous for the independent-coding and discontinuous for the lumped-coding channel), and thus to the order of the phase transition. The information landscape takes a qualitatively different shape at the threshold bifurcations in each case, demonstrating the emergence of a new threshold through splitting either through a gradual "breaking" of the information ridge (Fig 6B and 6C), or through a discrete switching from a local information maximum to the global maximum (Fig 6F–6H).

Thresholds in an auditory nerve fiber population resemble predictions from optimal coding

Next, we sought to compare our theoretical predictions of optimal thresholds to experimentally recorded thresholds of sensory populations to determine whether they are consistent with optimal coding. Specifically, we considered recordings of auditory nerve fibers (ANFs) which code for sound frequency and sound intensity. At the first synapse level of the auditory pathway, each inner hair cell of the cochlea transmits information about sound intensity to approximately ten to thirty different ANFs [7]. ANFs differ in several aspects of their responses, including spontaneous rates and thresholds, with each ANF receiving input exclusively from only a single inner hair cell [8]. We investigated the properties of experimentally recorded ANF tuning curves in the mouse for the frequency that corresponds to the lowest threshold, where the ANF is most sensitive [8]. Therefore, we 'projected' the neuronal code onto the dimension of sound intensity and hence could build a model for the coding of sound intensities based on spike counts. As a population, ANF tuning curves resemble a sigmoid which increases with sound intensity; the sigmoid can be described by a threshold, a dynamic coding range also referred to as a gain, a spontaneous firing rate and a maximal firing rate (see Methods, Fig 7A). Interestingly, ANF response curves with higher spontaneous firing rate have been shown to have narrower dynamic ranges and higher thresholds [8] (Fig 7B). Given the lack of convergence on stimulus channels, we investigated whether our theoretical framework of the independent-coding (rather than the lumped-coding) channel with two sources of noise before and after a nonlinearity can be applied to explain this relationship between spontaneous firing rate, dynamic range and firing threshold, testing the hypothesis that ANFs have optimized their response properties to encode maximal information about the stimulus under biological constraints.

To apply our population coding framework to this type of data would require measurements of the entire population of ANFs. In the absence of such data, we decided to apply the framework to a population of two representative neurons where each neuron can be described by a sigmoidal response function, one with a high and the other with a lower spontaneous firing rate. This is computationally tractable and consistent with previous literature [8, 52]. To obtain the two representative neurons, we proceeded as follows: first, we pooled the measured tuning curves from the same ANF, and fitted each with our sigmoidal function (see Methods, Fig 7A). This resulted in 148 tuning curves from 24 animals. Indeed, we confirmed that the dynamic range is negatively correlated with the normalized spontaneous firing rate, as well as positively correlated with the thresholds (S5 Fig). Then, we divided all the tuning curves into two types based on their normalized spontaneous firing rate and dynamic range (see Methods) [8, 52]. In particular, the 'Type 1' neuron had a higher spontaneous rate, a smaller dynamic
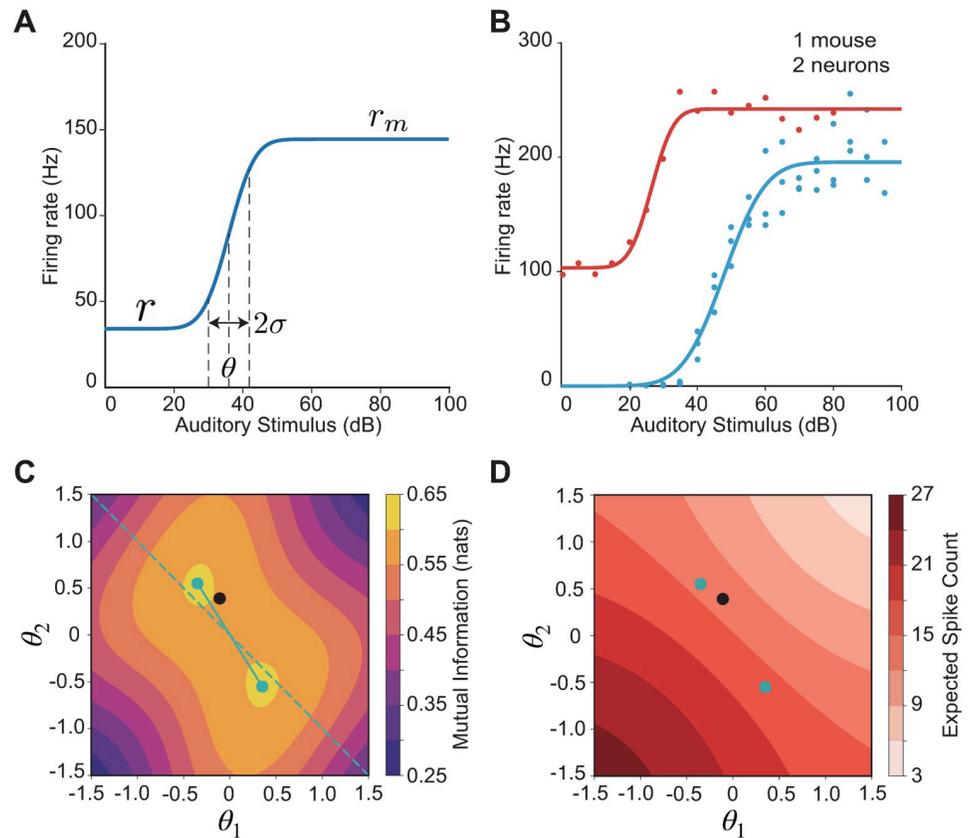
Fig 7. The tuning of auditory nerve fibers (ANFs) match predictions from optimal coding. A. The sigmoidal function that we use to fit ANFs tuning curves. Spontaneous firing rate (r), maximal firing rate (r_m), firing threshold (θ), and the dynamic range (σ) are labelled on the curve. **B**. An example showing original data from ref. [8] and fitted tuning curves. These two tuning curves come from the same mouse. **C**. Optimal configuration (cyan dots) in the contour plot of mutual information. Black dot denotes the fitted thresholds from the data. **D**. Optimal configuration (cyan dots) in the contour plot of average firing rate. Black dot denotes the fitted thresholds from the data.

range and a lower firing threshold, while the 'Type 2' neuron had a lower spontaneous rate, a broader dynamic range and a higher firing threshold (Fig 7B and S5 Fig).

For this two-neuron population where one neuron had a nonzero and the other zero spontaneous rate, we optimized the neuronal thresholds while maximizing the mutual information between the stimulus and the population response. We used the level of input noise to modulate the neurons' dynamic range (see Methods). Using dynamic ranges for the two neurons chosen to match the two neuron types found in the data, we evaluated the mutual information between stimulus and response for a range of firing thresholds (θ_1 and θ_2). We found that the mutual information landscape is centrosymmetric, with a maximal value of 0.603 nats achieved for two pairs of thresholds: $(\theta_1, \theta_2) = \{(-0.35, 0.55), (0.35, -0.55)\}$ (Fig 7C, cyan symbols). Of the two pairs, the second pair yields ∼20% higher mean firing rate (Fig 7D, cyan symbols). Therefore, the information per spike is higher for the first pair (0.0425 vs. 0.0374 nats/spike). When we overlaid the fitted thresholds from the data, we found that they lie remarkably close to the optimal thresholds obtained from the theoretical analysis (Fig 7C and 7D, black symbol). In particular, the maximum mutual information in our model is 0.603 nats, which is only one percent higher from that achieved in the data, 0.597 nats. This suggests that the ANF

population might be configured to maximize information per spike about a distribution of frequencies.

We further explored how sensitive the optimization is to the chosen parameters of the sigmoids of the two types of neurons, more specifically, the dynamic range and the spontaneous rate. When both neurons are binary (no input noise) so that the sigmoids are infinitely steep and have a narrow dynamic range, and the spontaneous rate in both neurons is zero, the contour of the mutual information is both axisymmetric and centrosymmetric but the two pairs of optimal thresholds yield identical mean firing rate (S6(A) Fig). Even when the neurons acquire a finite gain (by increasing input noise) so that the dynamic range is broadened, as long as the gain is the same, the information per spike for the two pairs of optimal thresholds remains identical (S6(B) Fig). Either changing the spontaneous rate or the gain of one of the neurons can break the symmetry in the information landscape, such that it loses its axisymmetry but preserves its centrosymmetry, which is due to the symmetric distribution of stimuli s and the low output noise. The symmetry breaking effect is much stronger when the gains of the two neuron types are different (S6(C) Fig vs. S6(D) Fig) and combining both different gains and non-zero spontaneous rates as in the data preserves the strong symmetry breaking (S6(E)–S6(G) Fig). Intuitively, the neuron with the larger input noise and hence, lower gain, has a threshold with a larger absolute value so that it is farther away from the mean of the stimulus distribution and thus is less affected by the input noise (S6(H) Fig), just like what we found in the data. Therefore, our model with binary neurons and two sources of noise can explain the relationship among spontaneous rate, dynamic range, and thresholds in two types of ANFs in the mouse by maximizing information per spike.

We also extended our framework to a population of three neurons to see if predictions from this model are consistent with the ANF data. We now divided all the tuning curves into three types based on their normalized spontaneous firing rate and dynamic range (see Methods and S7 Fig). Using this three-neuron population model, we computed the mutual information between stimulus and response for a range of firing thresholds (θ_1, θ_2 and θ_3) using dynamic ranges and spontaneous rates for the three neurons chosen to match the three neuron types found in the data (see Methods). We found that the maximum mutual information is 0.751 nats, which is only a few percent higher from that achieved when we further used the thresholds extracted from the data, 0.730 nats. This suggests that the predictions of the extended three-neuron model about the optimality of information transmission are also consistent with the ANF data.

Discussion

We maximized the mutual information between stimulus and responses of a population of neurons which encode a one-dimensional stimulus with a binary nonlinearity corrupted by two different noise sources, specifically, additive input noise before the nonlinearity and Poisson output noise after the nonlinearity. We compared two frameworks for stimulus convergence commonly used in previous studies, specifically, encoding the stimulus with independent transmission channels [5, 14, 25, 39] or lumping the channels into one effective channel [29, 30]. In each scenario, we calculated the optimal thresholds of the population (Fig 4).

Lumping of information channels as a coding strategy with low cost

Unsurprisingly, increasing either input or output noise in the population, decreases the total amount of transmitted information; but the independent-coding channel always encodes more information than the lumped-coding channel, especially for biologically realistic, intermediate output noise values (Fig 3). This occurs because lumping multiple information

pathways into a single coding channel reduces the possible values of the encoding variable and increases the noise entropy, thus introducing additional noise. Therefore, threshold bifurcations in the lumped-coding channel occur at significantly lower critical noise levels compared to the independent-coding channel (Fig 4).

Why would a biological system lump information transmission channels? A biological upside of combining information from multiple streams into one effective channel could be the reduction of neurons needed for information transmission, thus saving space and energy. For example, the optic nerve has a strong incentive to reduce its total diameter since it crosses through the retina and thus causes a blind spot. On the other hand, for a given constraint on space and energy, it is favorable to have many thin, low-rate axons over fewer thick, high-rate axons [53, 54], thus arguing against convergence. However, at least for the retina, an intermediate degree of convergence is probably the optimal solution. One would expect that this degree of convergence depends on the location at the retina. At the fovea of a primate retina, there is minimal convergence from photoreceptors to retinal ganglion cells compared to the periphery [55]. This implies that a higher visual acuity is achieved by increasing information transmission at the cost of energy and space. In contrast, there does not seem to be any convergence in the early auditory pathway: At the first stage of the neural signaling process, one inner hair cell diverges to 10 to 35 auditory nerve fibers [7]. This lack of convergence might be due to the fact that, contrary to the retina, there is no pressure of having a thin ganglion. A recent theoretical study suggests that convergence can compensate the information loss due to a non-linear tuning curve with a small number of output states [56].

We only treated the extreme cases of full convergence—where all neurons are lumped into a single channel—and no convergence. In principle, different combinations of partial convergence, e.g. lumping three outputs into two channels, are also possible. Partial lumping is a common strategy in sensory systems with different levels of convergence [57]. Furthermore, we assumed no weighting of inputs during the lumping process. This is an oversimplification since in neural circuits spikes from different presynaptic neurons could have a different impact on the membrane potential of the postsynaptic neuron depending on the synaptic connection strengths. These individual weights could also be optimized [30], which is beyond the scope of our paper.

Optimal number of distinct thresholds as a function of noise

The number of distinct optimal thresholds decreases with increasing noise of either kind at critical noise levels by successive bifurcations of the optimal thresholds (Fig 4). We mapped these characteristic bifurcations of the optimal thresholds at critical noise levels to phase transitions of different orders with order parameters being the threshold differences. At finite noise levels, the lumped-coding channel undergoes discontinuous threshold bifurcations which correspond to a first-order phase transition with respect to noise where the threshold differences are the order parameters. In contrast, for the independent-coding channel, the threshold differences change continuously and the phase transitions are of second-order.

Interestingly, for a range of noise parameters, we found a non-monotonic change in the number of distinct optimal thresholds with noise levels (S1 Fig). A similar non-monotonicity has also been reported under maximization of the Fisher information for neurons encoding sound direction [28]. This happens because of how the different neurons tile their thresholds to optimally encode the one-dimensional stimulus in the presence of multiple noise sources which interact non-trivially. The biological implications of such a non-monotonic change in the number of optimal thresholds as a function of noise are unclear. A related phenomenon in physics is that of *retrograde phenomena* [58]. For example, in a mixture of liquids, a phase

transition from liquid to gas, followed by another transition from gas to liquid, and then liquid to gas again can be observed while increasing temperature [58].

Analogies and differences to phase transitions in statistical physics

Our results suggest that input and output noise influence the mutual information in a very similar way how temperature affects free energy in statistical models of physical systems [49]; in that sense, both noise sources act as external parameters with respect to which the phase transition occurs. As in classical physical systems, the order of our phase transitions can be consistently linked to the continuity of the threshold differences: a continuous (discontinuous) order parameter corresponds to a second (first) order phase transition. In statistical models, the transition of an order parameter from zero to non-zero is accompanied by a spontaneous symmetry breaking of the system. Similarly, there is a symmetry breaking in our system as optimal thresholds become unequal at critical noise levels and thus the statistical equivalence of neurons breaks. For the case of continuous order parameters, i.e. optimal threshold differences of the independent-coding channel, we found the critical exponents of the order parameter to be 0.5—irrespective of the noise source. This value corresponds to the mean-field theory of continuous phase transitions [51], which underscores the similarity of our phase transitions to those of physical systems. Furthermore, we found the critical exponents for eigenvalues of the Hessian matrix of the information landscape to be 1—for which we have not established a direct correspondence in physical systems. As before, the source of the noise has no impact on the critical exponents, which again highlights that additive input noise and Poisson output noise have a similar influence. Previous work has also made the connection between phase transitions and information theory, showing that the maximization of the Fisher information is related to divergences in specific elements of the Fisher information matrix observed in experimental networks of finite size [59]. Similarly to our work, while these divergences occur at a critical point when the corresponding order parameter changes continuously, they disappear at the critical point when the first derivative of the order parameter diverges.

As in this previous work [59], we have more than one order parameter, specifically the number of subsequent threshold differences which correspond to the number of neurons minus one. Our scenario with three neurons shows similarities with a system with three mixed liquids where the miscibility depends on the liquids' relative concentration differences [60]. As the temperature varies, the system undergoes phase transitions where the miscibility changes, from having one phase in which all three liquids are mixable (similar to our scenario with three identical thresholds), to two phases where in one phase two liquids are mixable but which is separated from a second phase containing the third liquid (corresponding to two distinct thresholds in our neuronal population), to three phases where none of the liquids are mixable with each other (corresponding to the case of all distinct thresholds).

Even though our phase transitions have similar properties to the ones from physical systems, there are some noteworthy differences. In statistical physics, phase transitions are characterized by a non-analytic behavior of the moment-generating function, which is directly related to the free energy [61, 62]. The moment-generating function is a sum of exponentials (see S1 Text) and should thus be non-analytic only when the size of the system is infinitely large, $N \to \infty$. In our work, we characterize phase transitions by non-analytic behavior of the maximized mutual information and find phase transitions for finite N, as small as two. Interestingly, the moment-generating function in our case is a smooth function of the thresholds and also—when the thresholds are not optimized but fixed—of both noises (see S8 Fig). However, in our case the moment-generating function becomes a non-analytic function of the

noises when the optimized thresholds are used. Furthermore, in standard phase transitions the order parameters are statistical quantities since they are the moments of a function, for example, magnetization in the Ising model is the mean over spin directions. In contrast, our order parameters are not statistical variables but are obtained by optimizing the mutual information. They might be related to the statistical moments of some function of neural activity or to a function of the statistical moments of neural activity, however, we have not found such a relationship.

Information loss at non-optimal thresholds

An important, but often neglected, question for optimal coding theories is how much worse are suboptimal solutions in comparison to optimal ones in terms of information transmission. In the independent-coding channel, near critical noise levels, the information landscape becomes flat in the directions of principal curvature. This suggests that multiple threshold combinations yield nearly identical information, a property of the neural population that is closely related to the concept of "stiff vs. sloppy modeling", whereby a system's output is insensitive to changes in "sloppy" directions of the parameter space, but very sensitive to changes in "stiff" directions [63–66]. Hence, even population codes that utilize suboptimal thresholds often achieve information very close to the maximal, and it is unclear whether such small information differences could be measured experimentally. This also raises the question whether a few percent more information about a stimulus realized by optimal codes could be sufficiently beneficial for the performance of a sensory system to become a driving force during evolution. It has been shown that mutations which have very small effects on evolutionary fitness are fixated in a population with a probability almost irrespective of the mutation being advantageous or deleterious [67, 68]. On the other hand, in certain sensory systems like the retina, entire populations of retinal ganglion cells perform multiple functions [69, 70] or fulfill different computations under different light conditions [71]. For such systems, there must be a fundamental trade-off in performance, since such a system cannot be optimal at all functions [72, 73]. The sloppiness of nearly-equivalent optimal thresholds that we observe near critical noise levels should resolve when considering that neurons have multiple constraints and often perform more than just one function or encode different stimulus features.

Assumptions in our model and comparison to other theoretical frameworks

There are several modeling assumptions in our theoretical framework that make mathematical treatment possible. First, we considered the encoding of a static stimulus, even though natural stimuli have correlations in space and time. Previous studies have exploited their correlation structure to explain various aspects of sensory coding, for example, the size and shape of receptive fields of retinal ganglion cells [12, 13, 21, 33, 35, 38]. Since correlations in the stimulus are thought to reduce effective noise values [38], by considering stimuli independent in time, we likely underestimated effective noise levels.

Moreover, our coding framework assumed a one-dimensional stimulus; thus, it is appropriate for explaining the number of the population's distinct thresholds which encode a *single* stimulus feature—this could be the contrast at a single spatial position on the retina (as found to be coded by two different types of OFF retinal ganglion cells that encode the same linearly filtered stimulus [5]), or sound intensity at a single frequency (as found to be coded by ANFs, which get input from the same inner hair cell [8, 74]). Throughout this study we investigated the encoding of a one-dimensional stimulus drawn from a Gaussian distribution; however, natural stimulus distributions have a higher level of sparseness than the Gaussian distribution

[75, 76]. Therefore, we also explored information maximization using the generalized normal distribution allowing us to continuously vary the kurtosis—how heavy the tails are—of both the stimulus and the input noise distributions. Our results remain qualitatively the same as for the Gaussian distributions (S9 Fig).

Second, we modeled each neuron in the population solely with a binary nonlinearity. This nonlinearity describes the tuning curve of the neuron as a function of a given stimulus feature. In general, a tuning curve with respect to a stimulus feature is measured by reverse correlating the stimulus variable with the output variable and fitting a linear-nonlinear model [77]. The linear part of the model denotes the stimulus feature to which the neuron responds and the nonlinear part represents the tuning curve. We did not incorporate the linear part in our model but rather assumed that the input to the nonlinearity is already linearly preprocessed because simultaneous optimization under different noise sources and stimulus convergence would be mathematically intractable. We chose binary nonlinearities as they are theoretically optimal under certain conditions of high (and biologically plausible) Poisson noise [25, 30, 78]. Importantly, however, under conditions of non-negligible input noise the optimal nonlinearity could be interpreted to acquire a finite slope thus making our analysis relevant also for continuous nonlinearities with sigmoidal shape. This is consistent with neuronal recordings; for example the steepness of the tuning curve of the H1 blowfly neuron increases with contrast, and for high contrast—which corresponds to low noise—the tuning curve is almost binary [16].

Third, we considered a constraint on the maximum expected spike count since the total encoded information cannot be infinite. Such a constraint is motivated by a biophysical limit of a neuron's firing rate and the biological reality of a short reaction time. Instead, one could constrain the *mean* spike count [5, 14, 26], which would be interpreted as a metabolic constraint. Maximum and mean rate constraints lead to qualitatively similar conclusions regarding the optimal number thresholds, as shown in small populations of two neurons [5, 14].

Many previous studies make very similar assumptions but consider certain limiting scenarios, for instance considering only one noise source [5, 29, 30, 39], studying a population with only two neurons [5, 25, 39], or introducing an additional source of additive output noise [25]. Table 2 summarizes these studies with regards to the different optimization measures, constraints, information convergence strategies, sources of noise and neural population size. While our results are in agreement with these previous studies in the specific limiting conditions, we extend the optimal coding framework by mapping the full space of noise and stimulus convergence thus linking and extending previous findings.

Table 2. Comparison of different studies with regards to the different optimization measures, constraints, information convergence strategies, sources of noise and neuronal population size. MI stands for Mutual Information and MSE for Mean Square Error.

Study	Optimality measure	Constraint	Lumped or indepedent	Input or output noise	# Neurons
Our study	MI	Maximum rate	Both	Both	≤ 6
Brinkman et al., [25]	MI and MSE	Maximum rate	Independent	Both	2
Gjorgjieva et al., [14]	MI and MSE	Maximum and mean rate	Independent	Both	2
Kastner et al., [5]	MI	Mean rate	Independent	Input	2
Gjorgjieva et al., [39]	MI and MSE	Maximum rate	Independent	Output	any
Nikitin et al., [30]	MI	Maximum rate	Lumped	Output	4
McDonnell et al., [29]	MI	Maximum rate	Lumped	Input	15
Bethge et al., [27]	MSE	Maximum rate	Independent	Output	≤ 4
Harper and McAlpine [28]	Fisher Info.	Bell-shaped tuning curves	Independent	Output	200

Implications of our model

With these considerations, our coding model can be applied to a population of neurons coding for a one-dimensional stimulus that could apply to any sensory system, including the coding of sound intensity in auditory nerve fibers [8, 74], the coding of temperature in thermosensation by heat- and cold-activated ion channels [79, 80], the coding of vibration frequency by mechanosensory neurons [81, 82] and the coding of contrast by retinal ganglion cells coding for the same spatial location and visual feature with different thresholds [5]. Given the generality of our theoretical framework, we studied two coding strategies commonly used in previous studies and the contribution of two sources of noise without going into detail of the origins of this noise. We found that low noise favors parallel encoding with different thresholds while high noise favors equal thresholds. We applied and tested our framework on data from the ANF, but our theory remains relevant for other sensory systems where the two sources of noise can be distinguished and measured. For instance, in the mammalian retina multiple sources of noise can be identified in the retinal circuits, including from the photoreceptors [83–86] or at the bipolar cell output synapses [38, 87–89]. In the case that our model was applied to coding by retinal ganglion cells at the same spatial locations and with the same visual feature, these sources would all count as input noise. Their relative contributions could change with ambient light level [90]. The output noise in this case would come after the thresholding nonlinearity, and would likely correspond to noise expected from stochastic vesicle release at synapses. This noise is often taken to follow Poisson statistics where the variance in output scales with the output strength [91]. Applying the theory to a different experimental system would depend on the specific circuitry of that system and the identification of noise sources that enter the circuit at different points. Our results could then be used to make predictions of the coding thresholds of a population of neurons as a function of the strength of each noise source. The success of such applications would depend on the ability to extract the relevant components of the neural circuit in question and to develop a mathematically tractable description of its computations.

Conclusion

In sum, we considered an optimal coding framework with contributions from two sources of noise and investigated information transmission under two differen scenarios of stimulus convergence. Since we did not model a specific sensory system, but rather aimed to uncover general principles of optimal coding solutions under the two sets of independent scenarios above (noise and stimulus convergence), the sources of noise in our model do not directly correspond to circuit elements, making direct comparison to experimental data difficult. However, by applying our framework to coding by two types of ANFs in the mouse with a higher and lower spontaneous rate, we found that their thresholds are close to the optimal ones when maximizing information per spike. More importantly, we extended previous theoretical results that considered specific limiting scenarios, in the process providing a unifying framework for how different noise sources and the strategy of stimulus convergence influence information transmission and number of distinct thresholds in populations of nonlinear neurons.

Methods

We assume that the stimulus s follows a Gaussian distribution with mean zero and variance σ_s^2: $P(s) \sim \mathcal{N}(0, \sigma_s^2)$. It is encoded by the spike counts $\{k_i\}$ of N binary neurons $i = \{1, .., N\}$ in a given coding time window ΔT. Input noise z, $P(z) \sim \mathcal{N}(0, \sigma_z^2)$, is added to the stimulus before the nonlinear processing, $\sigma := \sigma_z/\sigma_s$ denotes the effective amount of input noise. For both the

stimulus and the noise distribution, we also considered other distributions with different kurtosis. However, we did not find significant differences to the Gaussian distributions (S9 Fig). We assume N binary nonlinearities $v_i(x) = v_{max}\Theta(x - \theta_i)$ with the two firing rate levels $v_i = \{0, v_{max}\}$ and a respective threshold θ_i. The input to each nonlinearity is the sum of stimulus and input noise: $x = s + z$. Poisson output noise is implemented by assuming that the spike count k_i in the coding window ΔT follows the Poisson distribution, $P(k_i|v_i) = (v_i\Delta T)^{k_i}/k_i! \; e^{-v_i\Delta T}$.

We denote the expected spike count of a neuron firing with its maximum firing rate v_{max} by $R := v_{max}\Delta T$. If R is small it means more output noise since even in the presence of maximum firing rate there are more occurrences of zero spikes and thus there is a higher ambiguity about the real firing rate. The above implementation of output noise can be understood as a constraint on the maximum firing rate level v_{max} while having a fixed coding window length ΔT.

For given noise levels σ and R the goal is to find nonlinearities which optimally encode the stimulus s with a vector of spike counts $\vec{k} \equiv \{k_i\}$ (independent-coding channel) or the lumped spike count $k = \Sigma k_i$ (lumped-coding channel). Since we assume binary nonlinearities and keep the two firing rates fixed, the only variables to optimize are the components of the threshold vector $\vec{\theta} := \{\theta_1, \ldots, \theta_N\}$. As a measure for optimality for the independent- and lumped-coding channels we choose the mutual information between stimulus s (input) and observed spike count $\vec{k}$ or k, respectively (output). The mutual information gives an upper bound on how much information can on average be obtained about the input by observing the output. It is given as the difference between output entropy $H(\vec{k})$ and noise entropy $H(\vec{k}|s)$ [92]:

$$I_m(\vec{k};s) = H(\vec{k}) - H(\vec{k}|s) \tag{2}$$

$$= -\sum_{k_1=0}^{\infty} \ldots \sum_{k_N=0}^{\infty} P(\vec{k}) \log(P(\vec{k})) + \sum_{k_1=0}^{\infty} \ldots \sum_{k_N=0}^{\infty} \int_s ds\, P(s)\, P(\vec{k}|s) \log(P(\vec{k}|s)) \tag{3}$$

$$= \sum_{k_1=0}^{\infty} \ldots \sum_{k_N=0}^{\infty} \int_s ds\, P(s)\, P(\vec{k}|s) \log\left(\frac{P(\vec{k}|s)}{\int_{s'} ds' P(s') P(\vec{k}|s')}\right) \tag{4}$$

where the input-output kernel $P(\vec{k}|s)$ is the probability of obtaining a certain vector of output spikes for a given stimulus value. In the case of the lumped-coding channel the calculations are the same, except that the spike count is now one-dimensional, i.e. we have $I_m(k; s)$ as the mutual information and $P(k|s)$ as the input-output kernel.

Independent-coding channel

In the case of the independent-coding channel, $P(\vec{k}|s) \equiv P(k_1, \ldots, k_N|s)$ can be expressed by multiplying $P(k_1, \ldots, k_N|v_1, \ldots, v_N)$ and $P(v_1, \ldots, v_N|s)$ and summing over all possible firing rate states $\{0, v_{max}\}$:

$$P(\vec{k}|s) = \sum_{v_1 \in \{0, v_{max}\}} \ldots \sum_{v_N \in \{0, v_{max}\}} P(k_1, \ldots, k_N|v_1, \ldots, v_N) P(v_1, \ldots, v_N|s) \tag{5}$$

We assume no noise correlations and thus v_i conditional on s are independent of each other:

$$P(v_1, \ldots, v_N|s) = \prod_i P(v_i|s) \tag{6}$$

Furthermore, all k_i are independent of each other conditional on a set of firing rates $\{v_1, \ldots,$

v_N}, and every k_i only depends on $v_{j=i}$:

$$P(k_1, \ldots, k_N | v_1, \ldots, v_N) = \prod_i P(k_i | v_1, \ldots, v_N) = \prod_i P(k_i | v_i) \tag{7}$$

Taken together:

$$P(\vec{k}|s) = \sum_{\vec{v} \in \{0, v_{max}\}^N} \prod_i^N P(k_i|v_i)P(v_i|s) = \prod_i^N \sum_{v_i \in \{0, v_{max}\}} P(k_i|v_i)P(v_i|s) \tag{8}$$

$P(k_i|v_i)$ follows a Poisson distribution and $P(v_i|s)$ denotes the probability of having a firing rate of zero (or v_{max}) for a given stimulus s. Since the input noise fluctuations are on a much faster time scale than the length of the coding window (over which the stimulus is assumed to be constant), an averaging over z can be performed. Thus $P(v_i = 0|s)$ (or $P(v_i = v_{max}|s)$) is given as the probability that stimulus plus noise is smaller (or larger, respectively) than threshold θ_i, which is the area under the noise distribution for which $s + z < \theta_i$ (or $s + z \geq \theta_i$, respectively):

$$P(v_i = v_{max}|s) = \int_{\theta_i - s}^{\infty} dz P_z(z) =: H_i(s), \tag{9}$$

$$P(v_i = 0|s) = \int_{-\infty}^{\theta_i - s} dz P_z(z) = 1 - H_i(s). \tag{10}$$

$H_i(s)$ can be viewed as the "effective" tuning curve that one would measure electrophysiologically (see also Fig 1, top right). It is the cumulative of the dichotomized noise distribution. If the noise distribution is normally distributed with variance σ^2, the effective tuning curve is given by the complementary error function:

$$H_i(s) = \frac{1}{2} \operatorname{erfc}\left(\frac{\theta_i - s}{\sqrt{2}\sigma}\right). \tag{11}$$

Then one can calculate the mutual information by performing the summation over all output variables $k_1, \ldots, k_N$. The output noise is included since $P(k_i|v_i)$ is Poisson distributed. According to the Poisson distribution, $P(k_i = 0|v_i = v_{max}) = e^{-R}$. For each k_i, all spike counts greater than zero can be lumped into one state due to the fact that if there is one or more spikes emitted, the firing rate can not be zero, i.e. $P(k_i > 0|v_i > 0) = 0$ for $v_i = \{0, v_{max}\}$. This state is denoted as 1 and from now on we have $k_i \in \{0, 1\}$. Thus $P(k_i = 1|v_i = v_{max}) = 1 - P(k_i = 0|v_i = v_{max}) = 1 - e^{-R}$. The mutual information can then be calculated as

$$I_m(\vec{k}; s) = \sum_{k_1, \ldots, k_N \in \{0,1\}^N} \int_s P_s(s) \prod_{i=1}^{N} \sum_{v_i \in \{0, v_{max}\}} P(k_i|v_i)P(v_i|s) \log\left(\frac{\ldots}{\int_{s'} ds' P(s') \ldots}\right) ds \tag{12}$$

with

$$\sum_{v_i} P(k_i = 0|v_i)P(v_i|s) = (1 - H_i(s)) + e^{-R}H_i(s) =: Q_i(s), \tag{13}$$

$$\sum_{v_i} P(k_i = 1|v_i)P(v_i|s) = (1 - e^{-R})H_i(s) =: S_i(s) \tag{14}$$

where output noise is denoted as $R := v_{max}\Delta T$. Taken together, the mutual information for the

independent-coding channel is

$$I_m(\vec{k}; s) = \sum_{k_1=0}^{1} \cdots \sum_{k_N=0}^{1} \int_s P_s(s) \left(\prod_{i=1}^{N} P_{k_i}(s) \right) \log \left(\frac{\prod_i P_{k_i}(s)}{\int_{s'} \mathrm{d}s' P(s') \prod_i P_{k_i}(s')} \right) \mathrm{d}s \tag{15}$$

with

$$P_{k_i}(s) = \begin{cases} Q_i(s), & \text{for } k_i = 0 \\ S_i(s), & \text{for } k_i = 1 \end{cases} \tag{16}$$

Lumped-coding channel

Next we turn to the calculation for the input-output kernel $P(k|s)$ in the case of the lumped-coding channel. For the case of only input input noise where $Q_i(s) = 1 - H_i(s)$ and $S_i(s) = H_i(s)$, McDonnell et al. [29, 93] explained how $P(k|s)$ can be calculated using a recursive formula. We extended these calculations to additional Poisson output noise. We write $P(k|s)$ as $P(k|N, s)$ and use the notation by McDonnell et al. [29, 93], for which

$$T_{k,s}^N := P(k|N, s). \tag{17}$$

Furthermore, $P_{k_i|s,i}$ is defined as the probability of cell i firing k_i spikes in a coding window ΔT when the stimulus is s. With that, one can express the probability of having k spikes with N cells as the probability of having k_N spikes by the N-th neuron multiplied by the probability of having $k - k_N$ spikes by the other neurons and taking into account all possibilities of k_N by summing over k_N:

$$T_{k,s}^N = \sum_{k_N=0}^{k} P_{k_N|s,i=N} \cdot T_{k-k_N,s}^{N-1} \tag{18}$$

where

$$P_{k_i|s,i} = \sum_{v_i \in \{0, v_{\max}\}} P(k_i|v_i) P(v_i|s) \tag{19}$$

$$= P(k_i|v_i = 0) P(v_i = 0|s) + P(k_i|v_i = v_{\max}) P(v_i = v_{\max}|s) \tag{20}$$

$$= \begin{cases} (1 - H_i(s)) + e^{-R} H_i(s), & \text{for } k_i = 0 \\ \dfrac{(R)^{k_i}}{k_i! \, e^{-R}} H_i(s), & \text{for } k_i > 0 \end{cases} \tag{21}$$

is the probability of cell i emitting k_i spikes given stimulus s, and

$$T_{0,s}^N = \prod_{i=1}^{N} P_{0|s,i} = (1 - H_i(s)) + e^{-R} H_i(s) = \prod_{i=1}^{N} Q_i(s) \tag{22}$$

being the probability of having zero spikes with N cells, as well as

$$T_{k,s}^1 = P_{k_1|s,i=1} \tag{23}$$

being the probability of having k spikes with $N = 1$. Thus for every $k = 0, 1, 2, \ldots$ until an

upper bound which is determined by the precision one wants to reach, $T^N_{k,s}$ is calculated for every $k_N = 0, 1, \ldots k$ by using the recursive formula in Eq 18. This is computationally very expensive and thus we studied only populations with up to $N = 3$ neurons and expected maximum spike count of $R = 10$ (note that calculating just *one* $P(k|s)$ for $N = 3$ and $R = 10$ requires on the order of 50 000 evaluations of Eq 19). As with the independent-coding channel, input noise σ is included in $H_i(s)$ (see Eq 9) and the output noise level is denoted by R.

Our goal is to find the optimal thresholds $\vec{\theta}$ which maximize mutual information for given levels of input and output noise σ and R:

$$\vec{\theta}^* = \arg\max_{\vec{\theta}} I_m(\vec{\theta}|\sigma, R). \tag{24}$$

Optimization procedure

For the calculation of Eq 4 we performed the integration numerically for both the independent- and the lumped-coding channel. For the independent-coding channel, this numerical integration is the computationally most expensive part of calculating the mutual information. We tested several numerical integration algorithms (Riemann, trapezoid, Romberg, Simpson, and adaptive algorithms) which all lead to very similar results. We performed numerical optimizations using the Nelder-Mead simplex algorithm implemented in the Scipy package [94]. It is a local optimizer which does not rely on estimating the gradient. Gradient based optimizers like the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm rely on calculating or estimating the inverse of the Hessian matrix. For the independent-coding channel this becomes problematic around critical noise values where one eigenvalue of the Hessian approaches zero and thus leads to large numerical imprecisions when inverting the Hessian. For the lumped-coding channel this is unproblematic and for speed purposes we also used an adaptation of the BFGS algorithm [95] implemented in Scipy. In order to spot possible local maxima—which are especially prevalent for large N—we applied a grid of initial conditions. After some trials it was possible to estimate what form of initial conditions lead to local maxima in the $N = 3$ case. Additionally, potential local maxima could in general be easily spotted and checked by considering the plots of optimal thresholds vs. noise.

The heavy numerical procedure limited our analysis to small population sizes with a maximum of three neurons in the case of the lumped-coding channel, and six neurons in the case of the independent-coding channel.

Generalized normal distribution

The generalized normal distribution (GND) is given by [41]

$$P_{\mathrm{GND}}(s) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{\left(-\frac{|s|}{\alpha}\right)^\beta} \tag{25}$$

where $\Gamma(z)$ is the gamma function given by

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} \mathrm{d}x. \tag{26}$$

The parameter β determines the kurtosis, particular values being $\beta = 1$ (for the Laplace distribution), $\beta = 2$ (the standard normal distribution) and $\beta \to \infty$ (the uniform distribution). The

variance of the GND is

$$\sigma^2 = \frac{\alpha^2 \Gamma(3/\beta)}{\Gamma(1/\beta)} \,.$$
(27)

The effective tuning curve of Eq 9 is in this case given as

$$H_i(s) = \frac{1}{2} - \text{sign}(\theta_i - s) \frac{\gamma\left(1/\beta, \left(\frac{|\theta_i - s|}{\alpha}\right)^\beta\right)}{2\Gamma(1/\beta)}$$
(28)

where $\gamma(x, y)$ is the lower incomplete gamma function defined as

$$\gamma(x, y) = \int_0^y t^{x-1} e^{-t} \mathrm{d}t \,.$$
(29)

Local curvature of information landscape

To investigate the curvature of the information landscape, we numerically calculated the Hessian matrix of the mutual information (using the Python package Numdifftools) at optimal thresholds and performed eigendecomposition. The Hessian matrix is defined as

$$H = \frac{\partial^2 I_m}{\partial \theta_i \partial \theta_j}.$$
(30)

Its eigenvectors give the directions of principal curvatures and the respective eigenvalues quantify the curvature in these directions.

Data fitting

We modeled ANF tuning curves as binary neurons, each neuron i with threshold θ_i so that if the stimulus (here, sound intensity at a given frequency) is higher (lower) than θ_i, the firing rate is $r_{m,i}$ (r_i). Here r_i denotes the spontaneous firing rate (SR) of the neuron, and $r_{m,i}$ denotes its maximal firing rate. The addition of Gaussian input noise with mean 0 and standard deviation σ_i transforms the effective tuning curve of the neuron into a sigmoid given by the equation (Figs 1A & 7A):

$$v_i(s) = r_i + (r_{m,i} - r_i)H_i(s),$$
(31)

where H_i denotes the complementary error function (Eq 11). To analyze the experimentally recorded ANF tuning curves from ref. [8], we first fit all the tuning curves with Eq 31. We used the approach from Balasooriya et al. [96] to identify a single outlier in the distribution of normalized SR, $r_i/r_{m,i}$. Upon removing the outlier, we pooled the measured tuning curves from the same ANF, and fitted each with our sigmoidal function (Fig 7A). This resulted in 148 tuning curves from 24 animals. To divide the tuning curves into two classes, since the distributions of normalized SR and the dynamic range are not center-symmetric, we calculated the cumulative distribution functions, $F(r_i/r_{m,i})$ and $F(\sigma_i)$ (S5 Fig).

$$F(r_i/r_{m,i}) = P(r/r_m \leq r_i/r_{m,i}),$$
(32)

$$F(\sigma_i) = P(\sigma \leq \sigma_i).$$
(33)

When $F(r_i/r_{m,i}) > F(\sigma_i)$, neuron i was classified as 'Type 1,' characterized by a higher SR, a smaller dynamic range and a lower firing threshold. Otherwise neuron i was classified as 'Type 2,' characterized by a lower SR, a broader dynamic range and a higher firing threshold (e.g. Fig 7B).

After fitting, we extracted the average values of all parameters for each type. The 'Type 1' neuron had a high normalized SR ($r_1/r_m = 0.159$) and a dynamic range of $\sigma_1 = 0.337$; while the 'Type 2' neuron had a very low normalized SR ($r_2/r_m = 0.036$) and a dynamic range of $\sigma_2 = 0.534$. These dynamic ranges were normalized to the stimulus distribution of natural sound frequencies used in the optimization, which is assumed to be a Gaussian distribution with mean 30 dB and standard deviation of 12.5 dB [97]. The maximum expected spike count R—the product of average maximal firing rate of the ANFs and the coding time window, $\Delta T = 50$ ms—was 13.8.

For the model with three neurons, we divided the data into three types (S7 Fig). The 'Type 1' neuron had a normalized SR of $r_1/r_m = 0.180$ and a dynamic range of $\sigma_1 = 0.328$, the 'Type 2' neuron had a normalized SR of $r_2/r_m = 0.087$ and a dynamic range of $\sigma_2 = 0.398$, while the 'Type 3' neuron had a normalized SR of $r_3/r_m = 0.019$ and a dynamic range of $\sigma_3 = 0.571$. These dynamic ranges were again normalized to a Gaussian stimulus distribution with mean 30 dB and standard deviation of 12.5 dB [97].

Finding local maxima of mutual information for the three-neuron model

Calculation of the mutual information landscape for the three-neuron model is much more computationally demanding compared to the two-neuron model. To find the information maximum, we started from a coarse grid of thresholds for the three neurons, and first calculated the mutual information for all the points in the threshold grid. After selecting the thresholds corresponding to the local maxima of the mutual information, we then zoomed in and used a finer grid of thresholds around each local maximum. Next, we found the local maxima in this new grid. We repeated this process until reaching the desired precision. Because of the centrosymmetry of the information landscape, we can do the calculation in only half of the space of thresholds.

Supporting information

S1 Fig. Non-monotonicity of the number of distinct optimal thresholds with output noise level R. A. For $N = 3$ neurons there is a non-monotonicity of the number of distinct optimal thresholds with output noise for a relatively small input noise parameter range ($0.54 < \sigma < 0.6$, see Fig 4A). For high output noise (low R), first the upper two thresholds merge, before they split again with decreasing output noise and for even lower output noise the middle threshold merges with the lower threshold. **B**. For $N = 6$ neurons a similar transformation of thresholds happens, where the two middle thresholds split with decreasing output noise, thus increasing number of distinct optimal thresholds.
(PDF)

S2 Fig. Optimal thresholds for higher number of neurons in the case of the independent-coding channel. A. Number of distinct optimal thresholds for $N = 4$ cells depending on input noise σ and output noise R. **B**. As in A but for $N = 6$.
(PDF)

S3 Fig. Threshold differences as phase transitions in the limit of just one noise source. A. Threshold bifurcations for the independent-coding channel with respect to input noise σ for vanishing output noise. The derivatives of mutual information with respect to input noise

indicate a second-order phase transition. **B**. As in A but for the lumped-coding channel. There is a first-order phase transition for low noise (left inset) and a second-order phase transition for high noise (right inset). **C**. The Independent-coding channel with vanishing input noise. No phase transition is visible since the "bifurcation" happens in the limit of infinite output noise. **D**. The Lumped-coding channel with vanishing input noise exhibits second-order phase transitions.
(PDF)

S4 Fig. Critical exponents obtained from fitting threshold differences and eigenvalues in proximity of critical noise values. A. Obtaining the critical exponent β_σ by fitting a monomial function to the threshold differences for input noise values slightly smaller than the critical input noise value σ_c. **B**. As in A, the critical exponent β_R is obtained by fitting output noise values slightly smaller than the critical output noise R_c. **C,D**. Similarly, one obtains the critical exponents of the eigenvalues of the Hessian matrix of the information landscape, ϕ_l and ϕ_r, by fitting the eigenvalues for both slightly smaller (ϕ_l) and slightly larger (ϕ_r) noise values than the critical noise value.
(PDF)

S5 Fig. Analysis of auditory nerve tuning curves of mice divided into two types. A. Distribution of normalized spontaneous firing rate (SR, r/r_m). An outlier is identified and marked in orange. **B**. Transforming normalized SR r/r_m and dynamic range σ into cumulative distribution functions. **C**. A diagram showing the method to classify neurons into two types. **D**. Scatter plot and linear fit between normalized SR r/r_m and dynamic range σ. Black dots denote the 'center of mass' within each 'type', and the green dots show the values of example neurons in Fig 7B. **E**. As in D but for the relationship between σ and threshold θ. Magenta dots and red dots denote where mutual information is maximized, with corresponding σ_1 and σ_2 as the black dots. Average firing rate corresponding to red dots are lower.
(PDF)

S6 Fig. Contour plots of mutual information and average firing rate, with different combination of σ_1, σ_2 and r_1/r_m (note, $r_2 = r_1$). In each panel, the left plot corresponds to mutual information and the right one shows average firing rate. Cyan dots show optimal thresholds (θ_1, θ_2) which maximize mutual information. Maximal spike count is set to $R = 10$ for every panel. **A**. Both neurons have zero input noise $\sigma_1 = \sigma_2 = 0$, and zero spontaneous rate $r_1/r_m = 0$. **B**. The two neurons have identical but nonzero input noise $\sigma_1 = \sigma_2 = 0.25$, and zero spontaneous rate $r_1/r_m = 0$. **C**. The two neurons have two different and nonzero input noise $\sigma_1 = 0.1$, $\sigma_2 = 0.5$, and zero spontaneous rate $r_1/r_m = 0$. **D**. Both neurons have zero input noise $\sigma_1 = \sigma_2 = 0$, and nonzero spontaneous rate $r_1/r_m = 0.2$. **E**. The two neurons have identical but nonzero input noise $\sigma_1 = \sigma_2 = 0.25$, and non-zero spontaneous rate $r_1/r_m = 0.2$. **F**. The two neurons have two different and nonzero input noise $\sigma_1 = 0.1$, $\sigma_2 = 0.5$, and nonzero spontaneous rate $r_1/r_m = 0.2$. **G**. $\sigma_1 = 0.5$, $\sigma_2 = 0.1$, $r_1/r_m = 0.2$. **H**. The mechanism behind symmetry breaking of the mutual information landscape. The case (left) where the neuron with the larger input noise has a larger threshold located in the region where the stimulus rarely occurs is more efficient than in the case (right) where the neuron with the larger input noise has a smaller threshold near the stimulus mean where its dynamic range covers a large range of possible stimuli.
(PDF)

S7 Fig. Analysis of auditory nerve tuning curves of mice divided into three types. A. A diagram showing the method to classify neurons into three types. **B**. Scatter plot and linear fit between normalized spontaneous firing rate (SR, r/r_m) and dynamic range σ. Black dots denote

the 'center of mass' within each 'type'. **C**. As in B but for the relationship between σ and threshold θ.
(PDF)

S8 Fig. The moment generating function of the spike output vector $\vec{k}$ is smooth for fixed thresholds but not for optimized thresholds of both noise sources. A. The two components ($N = 2$) of the moment-generating function $\vec{M}(\vec{t})$ for $\vec{t} = (1, 1)$ (see S1 Text) depending on input noise σ. Output noise value and threshold vector are fixed to $R = 1$ and $\vec{\theta} = (-0.5, 0.5)$, respectively. The first two derivatives show no discontinuities. **B**. As A but depending on R with $\sigma = 0.2$. **C**. As A,B but depending on first threshold vector θ_1. **D,E**. As A,B but with optimized threshold vector for each noise value. The components of the moment-generating function show a bifurcation and the first derivatives show discontinuities.
(PDF)

S9 Fig. Number of distinct optimal thresholds when using input and noise distributions different from Gaussian. Instead of a Gaussian stimulus and noise distribution we also used a generalized normal distribution and varied the kurtosis (small β means high kurtosis, see Eq 25). **A**. Laplacian (having high kurtosis) as input distribution. **B**. Input distribution with low kurtosis (similar to uniform). **C**. Laplace distribution as noise distribution. **D**. Noise distribution with low kurtosis.
(PDF)

S1 Text. The moment-generation function of the independent-coding channel for two neurons.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Kai Röth, Julijana Gjorgjieva.

Formal analysis: Kai Röth, Shuai Shao, Julijana Gjorgjieva.

Funding acquisition: Julijana Gjorgjieva.

Investigation: Kai Röth, Shuai Shao, Julijana Gjorgjieva.

Methodology: Kai Röth, Shuai Shao, Julijana Gjorgjieva.

Project administration: Julijana Gjorgjieva.

Software: Kai Röth, Shuai Shao, Julijana Gjorgjieva.

Supervision: Julijana Gjorgjieva.

Visualization: Kai Röth, Shuai Shao.

Writing – original draft: Kai Röth, Shuai Shao, Julijana Gjorgjieva.

Writing – review & editing: Kai Röth, Shuai Shao, Julijana Gjorgjieva.

References

1. Masland RH. The neuronal organization of the retina. Neuron. 2012; 76(2):266–280. http://www.sciencedirect.com/science/article/pii/S0896627312008835. PMID: 23083731

2. Sanes JR, Masland RH. The types of retinal ganglion cells: Current status and implications for neuronal classification. Annual Review of Neuroscience. 2015; 38(1):221–246. https://doi.org/10.1146/annurev-neuro-071714-034120 PMID: 25897874

3. Baden T, Berens P, Franke K, Rosón MR, Bethge M, Euler T. The functional diversity of retinal ganglion cells in the mouse. Nature. 2016; 529(7586):345–350. http://dx.doi.org/10.1038/nature16468. PMID: 26735013

4. Kastner DB, Baccus SA. Coordinated dynamic encoding in the retina using opposing forms of plasticity. Nature Neuroscience. 2011; 14(10):1317–1322. http://dx.doi.org/10.1038/nn.2906. PMID: 21909086

5. Kastner DB, Baccus SA, Sharpee TO. Critical and maximally informative encoding between neural populations in the retina. Proceedings of the National Academy of Sciences. 2015; 112(8):2533–2538. http://www.pnas.org/content/112/8/2533.abstract. PMID: 25675497

6. Segev R, Puchalla J, Berry MJ. Functional organization of ganglion cells in the salamander retina. Journal of Neurophysiology. 2006; 95(4):2277–2292. https://doi.org/10.1152/jn.00928.2005. PMID: 16306176

7. Nayagam BA, Muniak MA, Ryugo DK. The spiral ganglion: Connecting the peripheral and central auditory systems. Hearing Research. 2011; 278(1):2–20. http://www.sciencedirect.com/science/article/pii/S0378595511001043. PMID: 21530629

8. Taberner AM, Liberman MC. Response properties of single auditory nerve fibers in the mouse. Journal of Neurophysiology. 2005; 93(1):557–569. https://doi.org/10.1152/jn.00574.2004. PMID: 15456804

9. Si G, Kanwal JK, Hu Y, Tabone CJ, Baron J, Berck ME, et al. Structured odorant response patterns across a complete olfactory receptor neuron population. Neuron. 2019; 101(5):950–962.e7. https://doi.org/10.1016/j.neuron.2018.12.030 PMID: 30683545

10. Tsunozaki M, Bautista DM. Mammalian somatosensory mechanotransduction. Current Opinion in Neurobiology. 2009; 19(4):362–369. http://www.sciencedirect.com/science/article/pii/S0959438809000890. PMID: 19683913

11. Bell CC. Mormyromast electroreceptor organs and their afferent fibers in mormyrid fish. III. Physiological differences between two morphological types of fibers. Journal of Neurophysiology. 1990; 63(2):319–332. https://doi.org/10.1152/jn.1990.63.2.319. PMID: 2313348

12. Atick JJ, Redlich AN. Towards a theory of early visual processing. Neural Computation. 1990; 2(3):308–320. http://dx.doi.org/10.1162/neco.1990.2.3.308.

13. Ratliff CP, Borghuis BG, Kao YH, Sterling P, Balasubramanian V. Retina is structured to process an excess of darkness in natural scenes. Proceedings of the National Academy of Sciences. 2010; 107(40):17368–17373. http://www.pnas.org/content/107/40/17368.abstract. PMID: 20855627

14. Gjorgjieva J, Sompolinsky H, Meister M. Benefits of pathway splitting in sensory coding. The Journal of Neuroscience. 2014; 34(36):12127–12144. https://doi.org/10.1523/JNEUROSCI.1032-14.2014 PMID: 25186757

15. Laughlin S. A simple coding procedure enhances a neuron's information capacity. Zeitschrift für Naturforschung c. 1981; 36(9-10):910–912. https://doi.org/10.1515/znc-1981-9-1040 PMID: 7303823

16. Brenner N, Bialek W, van Steveninck RR. Adaptive rescaling maximizes information transmission. Neuron. 2000; 26(3):695–702. http://www.sciencedirect.com/science/article/pii/S0896627300812052. PMID: 10896164

17. Faisal AA, Selen LPJ, Wolpert DM. Noise in the nervous system. Nature Reviews Neuroscience. 2008; 9:292–303. http://dx.doi.org/10.1038/nrn2258. PMID: 18319728

18. Rieke F, Bodnar DA, Bialek W. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. Proceedings of the Royal Society of London B: Biological Sciences. 1995; 262(1365):259–265. http://rspb.royalsocietypublishing.org/content/262/1365/259. PMID: 8587884

19. Warland DK, Reinagel P, Meister M. Decoding visual information from a population of retinal ganglion cells. Journal of Neurophysiology. 1997; 78(5):2336–2350. https://doi.org/10.1152/jn.1997.78.5.2336. PMID: 9356386

20. Chichilnisky EJ, Rieke F. Detection sensitivity and temporal resolution of visual signals near absolute threshold in the salamander retina. Journal of Neuroscience. 2005; 25(2):318–330. http://www.jneurosci.org/content/25/2/318. PMID: 15647475

21. Haft M, van Hemmen JL. Theory and implementation of infomax filters for the retina. Network: Computation in Neural Systems. 1998; 9(1):39–71. https://doi.org/10.1088/0954-898X_9_1_003. PMID: 9861978

22. Tkačik G, Callan CG, Bialek W. Information flow and optimization in transcriptional regulation. Proceedings of the National Academy of Sciences. 2008; 105(34):12265–12270. https://doi.org/10.1073/pnas.0806077105 PMID: 18719112

23. Tkačik G, Walczak AM, Bialek W. Optimizing information flow in small genetic networks. Physical Review E. 2009; 80:031920. https://link.aps.org/doi/10.1103/PhysRevE.80.031920. PMID: 19905159

24. Berens P, Ecker AS, Gerwinn S, Tolias AS, Bethge M, Geisler WS. Reassessing optimal neural population codes with neurometric functions. Proceedings of the National Academy of Sciences. 2011; 108 (11):4423–4428. http://www.jstor.org/stable/41061130. PMID: 21368193

25. Brinkman BAW, Weber AI, Rieke F, Shea-Brown E. How do efficient coding strategies depend on origins of noise in neural circuits? PLOS Computational Biology. 2016; 12(10):1–34. http://dx.doi.org/10.1371%2Fjournal.pcbi.1005150. PMID: 27741248

26. Wang Z, Stocker AA, Lee DD. Efficient neural codes that minimize lp reconstruction error. Neural Computation. 2016; 28(12):2656–2686. http://dx.doi.org/10.1162/NECO_a_00900. PMID: 27764595

27. Bethge M, Rotermund D, Pawelzik K. Optimal neural rate coding leads to bimodal firing rate distributions. Network: Computation in Neural Systems. 2003; 14(2):303–319. http://www.tandfonline.com/doi/abs/10.1088/0954-898X_14_2_307. PMID: 12790186

28. Harper NS, McAlpine D. Optimal neural population coding of an auditory spatial cue. Nature. 2004; 430:682–686. http://dx.doi.org/10.1038/nature02768. PMID: 15295602

29. McDonnell MD, Stocks NG, Pearce CEM, Abbott D. Optimal information transmission in nonlinear arrays through suprathreshold stochastic resonance. Physics Letters A. 2006; 352(3):183–189. http://www.sciencedirect.com/science/article/pii/S0375960105018074.

30. Nikitin AP, Stocks NG, Morse RP, McDonnell MD. Neural population coding is optimized by discrete tuning curves. Physical Review Letters. 2009; 103:138101. http://link.aps.org/doi/10.1103/PhysRevLett.103.138101. PMID: 19905542

31. Barlow H. Redundancy reduction revisited. Network: Computation in Neural Systems. 2001; 12(3):241–253. http://dx.doi.org/10.1080/net.12.3.241.253. PMID: 11563528

32. van Hateren JH. A theory of maximizing sensory information. Biological Cybernetics. 1992; 68(1):23–29. https://doi.org/10.1007/BF00203134. PMID: 1486129

33. van Hateren JH. Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. Journal of Comparative Physiology A. 1992; 171(2):157–170. https://doi.org/10.1007/BF00188924.

34. Tkačik G, Prentice JS, Balasubramanian V, Schneidman E. Optimal population coding by noisy spiking neurons. Proceedings of the National Academy of Sciences. 2010; 107(32):14419–14424. http://www.pnas.org/content/107/32/14419.abstract. PMID: 20660781

35. Karklin Y, Simoncelli EP. Efficient coding of natural images with a population of noisy linear-nonlinear neurons. In: Advances in neural information processing systems. MIT Press; 2011. p. 999–1007.

36. Pitkow X, Meister M. Decorrelation and efficient coding by retinal ganglion cells. Nature Neuroscience. 2012; 15(4):628–635. http://dx.doi.org/10.1038/nn.3064. PMID: 22406548

37. Doi E, Gauthier JL, Field GD, Shlens J, Sher A, Greschner M, et al. Efficient coding of spatial information in the primate retina. Journal of Neuroscience. 2012; 32(46):16256–16264. http://www.jneurosci.org/content/32/46/16256. PMID: 23152609

38. Borghuis BG, Ratliff CP, Smith RG, Sterling P, Balasubramanian V. Design of a neuronal array. Journal of Neuroscience. 2008; 28(12):3178–3189. http://www.jneurosci.org/content/28/12/3178. PMID: 18354021

39. Gjorgjieva J, Meister M, Sompolinsky H. Functional diversity among sensory neurons from efficient coding principles. PLOS Computational Biology. 2019; 15(11):1–38. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007476. PMID: 31725714

40. Atick JJ. Could information theory provide an ecological theory of sensory processing? Network: Computation in Neural Systems. 1992; 3(2):213–251. https://doi.org/10.1088/0954-898X_3_2_009.

41. Nadarajah S. A generalized normal distribution. Journal of Applied Statistics. 2005; 32(7):685–694. https://doi.org/10.1080/02664760500079464.

42. Roska B, Meister M. The retina dissects the visual scene into distinct features. In: The new visual neurosciences. MIT Press; 2014. p. 163–182.

43. Hallem EA, Carlson JR. Coding of odors by a receptor repertoire. Cell. 2006; 125(1):143–160. http://www.sciencedirect.com/science/article/pii/S0092867406003631. PMID: 16615896

44. Tan J, Savigner A, Ma M, Luo M. Odor information processing by the olfactory bulb analyzed in gene-targeted mice. Neuron. 2010; 65(6):912–926. http://www.sciencedirect.com/science/article/pii/S0896627310001029. PMID: 20346765

45. Geisler CD. From sound to synapse: Physiology of the mammalian ear. Oxford University Press; 1998.

46. Sachs MB, Winslow RL, Sokolowski BH. A computational model for rate-level functions from cat auditory-nerve fibers. Hearing research. 1989; 41(1):61–69. https://books.google.de/books?id=baRutmLuDJsC. PMID: 2793615

47. Borst A, Theunissen FE. Information theory and neural coding. Nature Neuroscience. 1999; 2(11):947–957. http://dx.doi.org/10.1038/14731. PMID: 10526332

48. Uzzell VJ, Chichilnisky EJ. Precision of spike trains in primate retinal ganglion cells. Journal of Neurophysiology. 2004; 92(2):780–789. https://doi.org/10.1152/jn.01171.2003. PMID: 15277596

49. Stanley HE. Introduction to phase transitions and critical phenomena. Oxford University Press; 1971.

50. Ehrenfest P. Phasenumwandlungen im ueblichen und erweiterten Sinn, klassifiziert nach dem entsprechenden Singularitaeten des thermodynamischen Potentiales. Verhandlingen der Koninklijke Akademie van Wetenschappen (Amsterdam). 1933; 36:153–157.

51. Landau LD, Lifšic EM, Lifshitz EM, Pitaevskii L. Statistical physics: theory of the condensed state. vol. 9. Butterworth-Heinemann; 1980.

52. Jackson BS, Carney LH. The spontaneous-rate histogram of the auditory nerve can be explained by only two or three spontaneous rates and long-range dependence. Journal of the Association for Research in Otolaryngology. 2005; 6(2):148–159. https://doi.org/10.1007/s10162-005-5045-6. PMID: 15952051

53. Perge JA, Koch K, Miller R, Sterling P, Balasubramanian V. How the optic nerve allocates space, energy capacity, and information. Journal of Neuroscience. 2009; 29(24):7917–7928. http://www.jneurosci.org/content/29/24/7917. PMID: 19535603

54. Balasubramanian V, Sterling P. Receptive fields and functional architecture in the retina. The Journal of Physiology. 2009; 587(12):2753–2767. http://dx.doi.org/10.1113/jphysiol.2009.170704. PMID: 19525561

55. Kolb H. How the Retina Works: Much of the construction of an image takes place in the retina itself through the use of specialized neural circuits. American Scientist. 2003; 91(1):28–35. http://www.jstor.org/stable/27858152.

56. Gutierrez GJ, Rieke F, Shea-Brown E. Nonlinear convergence preserves information. bioRxiv. 2019. https://www.biorxiv.org/content/early/2019/12/13/811539.

57. Euler T, Haverkamp S, Schubert T, Baden T. Retinal bipolar cells: elementary building blocks of vision. Nature Reviews Neuroscience. 2014; 15(8):507–519. http://dx.doi.org/10.1038/nrn3783. PMID: 25158357

58. Danesh A. PVT and phase behaviour of petroleum reservoir fluids. vol. 47. Elsevier; 1998.

59. Prokopenko M, Lizier JT, Obst O, Wang XR. Relating Fisher information to order parameters. Physical Review E. 2011 Oct; 84:041116. https://link.aps.org/doi/10.1103/PhysRevE.84.041116. PMID: 22181096

60. Campbell FC. Phase diagrams: understanding the basics. ASM International; 2012.

61. Bulmer M. Statistical inference. In: Principles of statistics. Dover Publications; 1979. p. 165–187.

62. Martin OC, Monasson R, Zecchina R. Statistical mechanics methods and phase transitions in optimization problems. Theoretical Computer Science. 2001; 265(1):3–67. http://www.sciencedirect.com/science/article/pii/S0304397501001499.

63. Brown KS, Sethna JP. Statistical mechanical approaches to models with many poorly known parameters. Physical Review E. 2003; 68:021904. https://link.aps.org/doi/10.1103/PhysRevE.68.021904. PMID: 14525003

64. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally sloppy parameter sensitivities in systems biology models. PLOS Computational Biology. 2007; 3(10):1–8. https://doi.org/10.1371/journal.pcbi.0030189. PMID: 17922568

65. Machta BB, Chachra R, Transtrum MK, Sethna JP. Parameter space compression underlies emergent theories and predictive models. Science. 2013; 342(6158):604–607. https://science.sciencemag.org/content/342/6158/604. PMID: 24179222

66. Panas D, Amin H, Maccione A, Muthmann O, van Rossum M, Berdondini L, et al. Sloppiness in spontaneously active neuronal networks. Journal of Neuroscience. 2015; 35(22):8480–8492. http://www.jneurosci.org/content/35/22/8480. PMID: 26041916

67. Kimura M. Some Problems of Stochastic Processes in Genetics. The Annals of Mathematical Statistics. 1957; 28(4):882–901. http://www.jstor.org/stable/2237051.

68. Kimura M. Evolutionary rate at the molecular level. Nature. 1968; 217(5129):624–626. https://doi.org/10.1038/217624a0 PMID: 5637732

69. Münch TA, da Silveira RA, Siegert S, Viney TJ, Awatramani GB, Roska B. Approach sensitivity in the retina processed by a multifunctional neural circuit. Nature Neuroscience. 2009; 12. http://dx.doi.org/10.1038/nn.2389. PMID: 19734895

70. Deny S, Ferrari U, Macé E, Yger P, Caplette R, Picaud S, et al. Multiplexed computations in retinal ganglion cells of a single type. Nature Communications. 2017; 8(1):1964. https://doi.org/10.1038/s41467-017-02159-y. PMID: 29213097

71. Tikidji-Hamburyan A, Reinhard K, Seitter H, Hovhannisyan A, Procyk CA, Allen AE, et al. Retinal output changes qualitatively with every change in ambient illuminance. Nature Neuroscience. 2015; 18(1):66–74. http://dx.doi.org/10.1038/nn.3891. PMID: 25485757

72. Shoval O, Sheftel H, Shinar G, Hart Y, Ramote O, Mayo A, et al. Evolutionary trade-offs, pareto optimality, and the geometry of phenotype space. Science. 2012. http://science.sciencemag.org/content/early/2012/04/25/science.1217405. PMID: 22539553

73. Młynarski WF, Hermundstad AM. Adaptive coding for dynamic sensory inference. eLife. 2018; 7: e32055. https://doi.org/10.7554/eLife.32055. PMID: 29988020

74. Liberman MC. Single-neuron labeling in the cat auditory nerve. Science. 1982; 216(4551):1239–1241. http://science.sciencemag.org/content/216/4551/1239. PMID: 7079757

75. Bell AJ, Sejnowski TJ. The "independent components" of natural scenes are edge filters. Vision Research. 1997; 37(23):3327–3338. http://www.sciencedirect.com/science/article/pii/S0042698997001211. PMID: 9425547

76. Field DJ. What is the goal of sensory coding? Neural Computation. 1994; 6(4):559–601. https://doi.org/10.1162/neco.1994.6.4.559.

77. Dayan P, Abbott LF. Theoretical neuroscience. MIT Press; 2001.

78. Shamai S. Capacity of a pulse amplitude modulated direct detection photon channel. IEE Proceedings I (Communications, Speech and Vision). 1990; 137:424–430(6). https://doi.org/10.1049/ip-i-2.1990.0056

79. Lumpkin EA, Caterina MJ. Mechanisms of sensory transduction in the skin. Nature. 2007; 445 (7130):858–865. http://digital-library.theiet.org/content/journals/10.1049/ip-i-2.1990.0056. PMID: 17314972

80. Dhaka A, Viswanath V, Patapoutian A. TRP ion channels and temperature sensation. Annual Review of Neuroscience. 2006; 29:135–161. https://doi.org/10.1146/annurev.neuro.29.051605.112958

81. Salinas E, Hernandez A, Zainos A, Romo R. Periodicity and firing rate as candidate neural codes for the frequency of vibrotactile stimuli. Journal of Neuroscience. 2000; 20(14):5503–5515. https://doi.org/10.1523/JNEUROSCI.20-14-05503.2000

82. Romo R, Brody CD, Hernandez A, Lemus L. Neuronal correlates of parametric working memory in the prefrontal cortex. Nature. 1999; 399:470–473. https://doi.org/10.1038/20939

83. Schneeweis DM, Schnapf JL. The photovoltage of macaque cone photoreceptors: adaptation, noise, and kinetics. Journal of Neuroscience. 1999; 19(4):1203–1216. https://doi.org/10.1523/JNEUROSCI.19-04-01203.1999 PMID: 9952398

84. Ala-Laurila P, Greschner M, Chichilnisky EJ, Rieke F. Cone photoreceptor contributions to noise and correlations in the retinal output. Nature Neuroscience. 2011; 14(10):1309–1316. http://dx.doi.org/10.1038/nn.2927. PMID: 21926983

85. Angueyra JM, Rieke F. Origin and effect of phototransduction noise in primate cone photoreceptors. Nature Neuroscience. 2013; 16(11):1692–1700. https://doi.org/10.1038/nn.3534 PMID: 24097042

86. Hansen BC, Field DJ, Greene MR, Olson C, Miskovic V. Towards a state-space geometry of neural responses to natural scenes: A steady-state approach. NeuroImage. 2019; 201:116027. https://doi.org/10.1016/j.neuroimage.2019.116027 PMID: 31325643

87. Freed MA. Parallel cone bipolar pathways to a ganglion cell use different rates and amplitudes of quantal excitation. The Journal of Neuroscience. 2000; 20(11):3956–3963. https://doi.org/10.1523/JNEUROSCI.20-11-03956.2000 PMID: 10818130

88. Dunn FA, Rieke F. The impact of photoreceptor noise on retinal gain controls. Current Opinion in Neurobiology. 2006; 16(4):363–370. http://www.sciencedirect.com/science/article/pii/S0959438806000900. PMID: 16837189

89. Freed MA, Liang Z. Synaptic noise is an information bottleneck in the inner retina during dynamic visual stimulation. The Journal of Physiology. 2014; 592(4):635–651. https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.2013.265744. PMID: 24297850

90. Farrow K, Teixeira M, Szikra T, Viney TJ, Balint K, Yonehara K, et al. Ambient illumination toggles a neuronal circuit switch in the retina and visual perception at cone threshold. Neuron. 2013; 78(2):325–338. http://dx.doi.org/10.1016/j.neuron.2013.02.014. PMID: 23541902

91. Murphy RF Gabe J. Signals and noise in an inhibitory interneuron diverge to control activity in nearby retinal ganglion cells. Nature Neuroscience. 2008; 11:318–326. https://doi.org/10.1038/nn2045

92. Cover TM, Thomas JA. Elements of information theory. John Wiley & Sons; 2012.

93. McDonnell MD, Abbott D, Pearce CEM. An analysis of noise enhanced information transmission in an array of comparators. Microelectronics Journal. 2002; 33(12):1079–1089. http://www.sciencedirect.com/science/article/pii/S0026269202001131.

94. Gao F, Han L. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. Computational Optimization and Applications. 2012; 51(1):259–277. https://doi.org/10.1007/s10589-010-9329-3.

95. Zhu C, Byrd RH, Lu P, Nocedal J. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. ACM Trans Math Softw. 1997; 23(4):550–560. http://doi.acm.org/10.1145/279232.279236.

96. Balasooriya U. Detection of outliers in the exponential distribution based on prediction. Communications in Statistics—Theory and Methods. 1989; 18(2):711–720. https://doi.org/10.1080/03610928908829929.

97. Singh NC, Theunissen FE. Modulation spectra of natural sounds and ethological theories of auditory processing. The Journal of the Acoustical Society of America. 2003; 114(6):3394–3411. https://doi.org/10.1121/1.1624067. PMID: 14714819