Preprint Accepted for publication in *Phys. Rev. E Rapid Commun.* (2019) **100**: 050401(R)

Critical size of neural population for reliable information transmission

Lubomir Kostal[1,*] and Ryota Kobayashi[2,†]

[1] *Institute of Physiology of the Czech Academy of Sciences, Videnska 1083, 14220 Prague 4, Czech Republic*
[2] *Principles of Informatics Research Division, National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, Japan*

It is known that the probability of decoding error has a phase transition at information rate equal to the channel capacity. The corresponding thermodynamic limit requires infinite coding dimension, hence making the actual decoding practically impossible. In this letter we analyze finite-size effects that occur in limited neural populations. We report that the achievable rate approaches the asymptote in a remarkably non-linear manner with the population size. Qualitatively, our findings do not seem to depend on the details of the model.

The ultimate rate of information (i.e., the channel capacity) that can be communicated through a noisy channel is an asymptotic quantity, assuming arbitrarily reliable communication, infinite decoder complexity and delays [1]. Stepping back from the asymptotic limit thus brings new parameters into play. In fact, in all practical applications of coding theory, the actual value of mutual information is of lesser interest than the probability that the input is incorrectly estimated at the destination. Although information transfer in unlimited neural populations may approach capacity [2], realistic neuronal populations are limited, in both their size and signal-to-noise ratio. Hence, the 'intermediate' regime of optimal decoding performance is of actual interest [3].

In this letter we investigate the information-optimality conditions for a size-restricted neuronal population [4, 5]. We examine the approach of the achievable rate to the asymptotic limit in dependence on the number of active neurons. Both classic [1] and recently improved bounds [6, 7] on the information rates with finitely-dimensional codes are employed. We find that there exists a critical (or threshold) population size, above which the information transfer scales almost asymptotically, while below the transmission deteriorates rapidly (Fig. 1A). Qualitatively similar phenomenon is known in the parameter estimation theory, where the estimator mean-squared error sometimes deviates from the Cramer-Rao bound abruptly as the sample size or signal-to-noise ratio decreases below the critical region [8].

Information capacity of a single neuron. Following the standard approach [5, 9–11], let x be a scalar input parameter (e.g., stimulus intensity, or feature), taking value in a possibly continuous set. Stimulation is formally described by the random variable X, with probability distribution or density function $p(x)$, which accounts for different frequencies of individual inputs in the stimulus ensemble. Denote y as the output value (neuronal response), e.g., the firing frequency, spike timing precision or some other characteristics [9]. For each input x the response is typically described stochastically. Although neuronal systems operate in continuous time, we implicitly assume that all input-output signals can be sufficiently well represented in 'discrete time' with step $\varDelta$, also due to the finite speed of the biological processes and the consequent bandwidth limitations.

Assuming that neuron acts as a memoryless information channel without feedback, the stimulus-response relationship is completely described by the conditional distribution function $f(y|X = x)$ of the response random variable Y. The unconditional output probability distribution function is then $p(y) = \langle f(y|x)\rangle_x$. The mutual information, $I(X;Y)$, in nats per time unit, is

$$I(X;Y) = \frac{1}{\varDelta}\left\langle \log \frac{f(y|x)}{p(y)} \right\rangle_{x,y}, \tag{1}$$

where $\varDelta$ is the duration of each input and the expectation is over the joint probability distribution of X and Y. The value of $I(X;Y)$ has become indispensable in the analysis of information flow in various stochastic physical [5, 12–16] or biological systems [9, 17–20].

The channel capacity, C, characterizes the ultimate information transmission reliability of any given neuronal model by optimizing $I(X;Y)$ over the set of all input distributions,

$$C = \max_{p(x)} I(X;Y). \tag{2}$$

More precisely, in order to facilitate the transmission reliability, information theory enforces processing of input sequences, $\{x_1, x_2, \dots, x_n\}$, instead of per-symbol decoding (Fig. 1B). Assuming that there are m distinct input n-sequences, the information rate, R, is defined as [1]

$$R = \frac{\log m}{n\varDelta}. \tag{3}$$

Shannon's channel coding theorem [1, Thm. 5.6.4] states that for all rates below channel capacity, $R < C$, the probability of decoding error, P_e, i.e., the probability that the decoder fails to identify the correct input sequence, can be made arbitrarily small with increasing n. In other words, up to $m \approx e^{\varDelta nC}$ patterns can be decoded reliably for n sufficiently large. The probability of decoding error in fact undergoes a phase transition at $R = C$ in the 'thermodynamic' limit $n \to \infty$. Shannon's theorem guarantees that $P_e = 0$ for $R < C$ while Wolfowitz's converse theorem [1, Thm. 5.8.5] implies $P_e = 1$ for $R > C$. The key question therefore is what is the relationship between achievable R and P_e for restricted n.

---

[*] E-mail: kostal@biomed.cas.cz
[†] E-mail: r-koba@nii.ac.jp

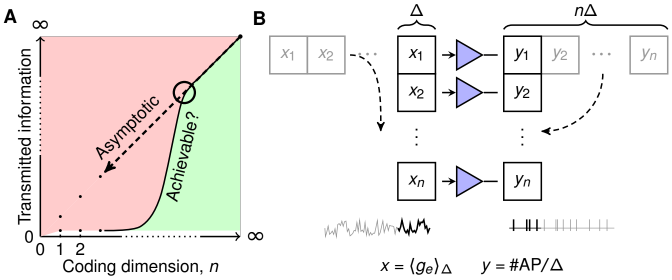

Figure 1. Framework and neural population set-up. (**A**) Typical results of information theory are asymptotic in the coding dimension n. As $n \to \infty$ it is possible to transmit information reliably up to rates equal to channel capacity per dimension (dashed line). However, the asymptotic limit is unlikely to be achieved with restricted n (schematic illustration). We report that there exists a critical n below which the performance deviates rapidly from the bound (circle). (**B**) Neuronal population model employed in this letter. The input (x_i) and output (y_i) are discretized with time-step Δ. For single neurons, Shannon's coding theorem assumes processing sequences of length n (coding dimension), resulting in delays ($n\Delta$) during the sequence formation and read-out. Instead we equivalently employ n identical neurons in parallel, using the Hodgkin-Huxley type model with point conductance input. Stimulus is the n-vector of average excitatory conductance components, response is the vector of individual firing frequencies.

Population model. We consider standard single-compartment neurons whose ionic currents are described by Hodgkin-Huxley type kinetics [4] (see Supplementary Material (SM, p. 10) for details and an additional neuronal model [21]). The synaptic input I_{syn} is given by the point-conductance model [22], where the excitatory (inhibitory) synaptic conductances follow the Ornstein-Uhlenbeck process.

The stimulus, x, is given by the mean excitatory synaptic conductance $\langle g_e \rangle$ during the time window of length Δ, which determines the time scale [23]. The inhibitory component is proportional to the excitatory one [24, 25], (SM, p. 12). For the purpose of this letter we confine x to the interval $[0, 0.4]$ mS/cm^2 and set $\Delta = 50$ ms (SM, p. 13) so that the effect of past stimulation can be approximately neglected. The response of the neuron, y, is the number of spikes observed in Δ (Fig. 1B). The conditional probability distribution $f(y|x)$ cannot be obtained in a closed form, however, we utilized intensive simulations to obtain a reliable approximation.

In order to avoid the delay $n\Delta$ implied by the channel coding theorem for the single-neuron case (Fig. 1B) we instead consider the population of n independent neurons. The probability distribution of a population response $\mathbf{y} = \{y_1, \dots, y_n\}$ to the input vector $\mathbf{x} = \{x_1, \dots, x_n\}$ is

$$f(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} f(y_i|x_i). \tag{4}$$

Due to the convenience of Eq. (4), populations of uncoupled neurons are used frequently, e.g., in [3, 26–29], see also [30] and references therein.

Since in each Δ we use n neurons in parallel, rather than a single neuron n-times, the information capacity C_n of the population grows as (Fig. 2A, dash-dotted line)

$$C_n = nC, \tag{5}$$

where $C \doteq 17.3$ bit/s is the single-neuron capacity from Eq. (2). (For convenience, results in Fig. 2A are expressed in bits rather than nats, i.e., 1 nat equals $1/\log(2)$ bit.) The capacity-achieving input distribution, $p(x)$, obtained by maximizing Eq. (2), is shown in Fig. 2B. The corresponding n-dimensional population input is then described by the ensemble Eq. (8).

The linear scaling in Eq. (5) does not equally hold for the population rate, $R_n = nR = (\log m)/\Delta$, as m is only asymptotically exponential in n, while the operational interpretation of C implies $n \to \infty$ (Fig. 1A, SM, p. 4). Unfortunately, no accurate expression for the best achievable rate, not exceeding the given decoding error tolerance, for arbitrary n is known in general [6]. We employ the recently improved upper bound [7] together with the optimized Gallager's achievability bound [1] under the maximum likelihood decoding rule to obtain an approximate description of the optimal rate R_n as a function of n.

Lower achievable bound. Given the set of available input vectors $\mathbf{x}^{(k)}, k = 1, \dots, m$, the current input $\mathbf{x}^{(i)}$ and the observed response vector $\mathbf{y}'$, the maximum-likelihood decoder chooses

$$\mathbf{x}^{(D)}: \quad D = \arg\max_k f(\mathbf{y}'|\mathbf{x}^{(k)}), \tag{6}$$

where $f(\mathbf{y}|\mathbf{x})$ follows Eq. (4). Decoding error occurs if $\mathbf{x}^{(D)} \neq \mathbf{x}^{(i)}$. Denote the probability of the k-th input vector as $\mathrm{P}_{\mathbf{x}}(k)$. The average probability of decoding error, P_e, is then

$$P_e = \sum_{k=1}^{m} \mathrm{P}_{\mathbf{x}}(k) \int_{\mathcal{E}(k)} f(\mathbf{y}|\mathbf{x}^{(k)})\,\mathrm{d}\mathbf{y}, \tag{7}$$

where $\mathcal{E}(k)$ is the set of all response vectors $\mathbf{y}$ such that if $\mathbf{x}^{(k)}$ is on the input the decoder in Eq. (6) fails. Without loss of generality we may assume a uniform distribution of input patterns, $\mathrm{P}_{\mathbf{x}}(k) = 1/m$, since the bounds employed below are valid for maximal error also [1, 7], and the maximum-likelihood decoder becomes the minimum-error decoder in this case.

Gallager's random coding bound on the achievable information rate relies on the average properties of the statistical ensemble of all possible m-sets of input n-vectors generated by the input probability distribution $p(x)$. The probability of a particular set of m inputs is

$$\prod_{k=1}^{m} p(\mathbf{x}^{(k)}) = \prod_{k=1}^{m} p(x_1^{(k)}) \cdots p(x_n^{(k)}), \tag{8}$$

due to the channel being memoryless [1]. For rates $R < C$ the optimal input distribution $p(x)$ generally differs from the maximizer in Eq. (2) and has to be be optimized in a different way. For the purposes of the population analysis in this letter, Gallager's random coding bound can be manipulated into the following form [1, Ch. 5.6]: for a given value of average decoding error P_e, maximum-likelihood decoding in Eq. (6) and

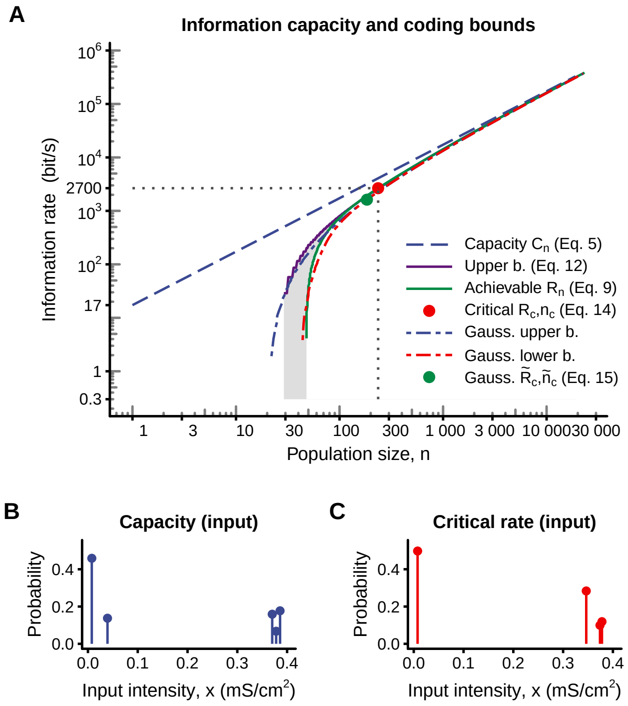

Figure 2. Optimal information transmission through a homogeneous neural population. (**A**) For finite-sized populations (following the scheme in Fig. 1B) the achievable information rate differs from the asymptotic capacity limit (divided per neuron, dashed). The area between the lower achievable bound and the upper bound (solid) indicates the performance of optimal decoders with probability of decoding error not exceeding $P_e = 10^{-10}$. There is a notable decrease in the information rate with respect to the asymptote (*cf.* Fig. 1A) below the critical population size ($n_c = 236$). The closed-form Gaussian approximation (dash-dotted) describes the upper/lower bounds (SM, p. 9) and the critical parameters well. (**B**) Asymptotically optimal (capacity-achieving) probability distribution over the input intensities maximizing Eq. (2). The distribution is discrete with five points of support. (**C**) Input probability distribution for the critical rate (Eq. 14). Note the simpler structure with respect to the capacity-achieving distribution (B).

population size n, it is guaranteed that the achievable population information rate R_n satisfies

$$R_n \geq \frac{n}{\Delta} E_r^{-1}\left(- \frac{\log P_e}{n} \right). \tag{9}$$

That is, there exists a set of $m \approx e^{\Delta R_n}$ input patterns such that the actual average probability of decoding error does not exceed P_e in Eq. (7). The function $E_r^{-1}(\cdot)$ is the inverse to the random coding error exponent,

$$E_r(\Delta R) = \max_{0 \leq \rho \leq 1} \left[\max_{p(x)} E_0\big(\rho, p(x)\big) - \rho \Delta R \right], \tag{10}$$

$$E_0\big(\rho, p(x)\big) = -\log \int \left(\int f(y|x)^{1/(1+\rho)} p(x)\, \mathrm{d}x \right)^{1+\rho} \mathrm{d}y. \tag{11}$$

The optimization problems in Eqs. (2) and (10) are convex in $p(x)$. We adopt the support-line method proposed by Gallager [1, Ch. 5.6] for the solution of Eq. (10) together with the cutting-plane algorithm [31, 32] that converts the originally nonlinear convex problem into a sequence of linear programs (SM, p. 7). Achievable rates from Eq. (9) are shown in Fig. 2A for $P_e = 10^{-10}$.

Upper bound. The converse theorems to Shannon's channel coding theorem state that reliable information transmission above certain rate is impossible since the prescribed error tolerance would have to be exceeded [1]. The upper bound on the information rate, as given by the converse, is not necessarily achievable.

We use the recently improved bound based on Strassen's normal approximation [7, 33],

$$R_n \leq nC - \frac{1}{\Delta}\left[\sqrt{nV(P_e)}\, Q^{-1}(P_e) + \frac{\log n}{2} \right] + O(1). \tag{12}$$

Here $Q^{-1}(\cdot)$ is the inverse to the Q-function, $Q(z) = \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, \mathrm{d}t$, and $V(P_e)$ is the channel dispersion, defined for $P_e \leq 1/2$ as

$$V(P_e) = \min_{p(x) \in \mathcal{C}} \left[\left\langle \log^2 \frac{f(y|x)}{p(y)} \right\rangle_{x,y} - \Delta^2 C^2 \right], \tag{13}$$

where $\mathcal{C}$ is the set of all input capacity-achieving distributions on the given channel. In other words, given the population size n and the prescribed probability of decoding error P_e, the right-hand side of Eq. (12) gives the information rate above which the probability of decoding error of any code exceeds P_e.

The rates upper-bounded by Eq. (12) are shown in Fig. 2A (gray region below the solid line) for $P_e = 10^{-10}$. The remaining terms in the bound that do not grow with n must be ignored for the numerical evaluation (they do not affect the asymptotic tightness), hence for very small n the upper bound is approximate [6].

Critical population size. Both the lower achievability bound in Eq. (9) and the upper bound in Eq. (12) approach the asymptote in a non-linear way with increasing n (Fig. 2A). Heuristically, it is possible to distinguish three performance regions, or regimes, in the diagram. To some extent, the described situation is analogous to the 'threshold effect' known from the parameter estimation theory [8, 34].

First, the initial 'no information' regime for very small n, where it is impossible to transmit any information while meeting the required error tolerance. The existence of this regime depends on the structure of the channel and on the value of P_e.

Second, the 'threshold' regime characterized by supra-linear growth of information rate with n, i.e., by a relatively rapid approach towards the capacity bound.

Third, for sufficiently high n, the 'asymptotic' regime where the amount of transmitted information grows almost linearly with the population size. The constant of proportionality is close to the capacity, the theoretical optimum. The logarithmic scale of the diagram allows its interpretation in terms of relative increments of the population size. If the number of neurons is increased by, e.g., 10 % while in the threshold regime, the benefit in terms of the added information rate is disproportionally larger than what would be expected in the asymptotic regime.

Next, we define the critical population size, n_c, to mark the transition towards the asymptotic regime more quantitatively. The critical rate R_c can be defined as the smallest rate at which the lower achievable bound in Eq. (9) and the upper bound agree in their error exponents [1, Ch. 5.8], and therefore indicates the presence of the 'almost'-asymptotic regime [35]. The value of R_c is best expressed as normalized per neuron [1],

$$R_c = \frac{1}{\Delta}\frac{\mathrm{d}}{\mathrm{d}\rho}\max_{p(x)} E_0\big(\rho, p(x)\big)\Big|_{\rho=1}. \tag{14}$$

The corresponding critical population size, n_c, is then found from Eq. (9) as $n_c = \lceil -(\log P_e)/E_r(\Delta R_c)\rceil$. The Eq. (14) gives consistent results across different neuronal models (SM, p. 12–14), The upper-lower bound equality is a desirable property, in addition, note that R_c does not depend on n for the population of independent and identical neurons.

The critical point $(n_c, n_c R_c)$ is marked in Fig. 2A. In this case $R_c \doteq 11.3$ bit/s, $n_c = 236$ and $n_c R_c \doteq 2568$ bit/s. The value of n_c is not affected by omitting the $O(1)$ term in Eq. (12), which essentially affect only the 'spread' of the gray region below the critical rate in Fig. 2A. The corresponding critical-rate input distribution, $p_c(x)$, is obtained by maximizing Eq. (10) for $R = R_c$ or using Eq. (14) (Fig. 2C). The distribution has a simpler structure (less points of support) than the capacity-achieving input distribution (Fig. 2B). The operational interpretation of the critical-rate input distribution follows from the concept of the input ensembles. If the ensemble is generated in accord with Eq. (8) using $p_c(x)$ and $n = n_c$ then at least one member of the ensemble (set of $m = \lceil \exp(\Delta n_c R_c)\rceil$ input vectors) can be used to communicate information at the critical rate, using the decoder in Eq. (6), and not exceeding the prescribed P_e.

The Eqs. (9) and (14) rarely lead to closed-form expressions. Nonetheless, for many cases of interest where $f(y|x)$ is well-behaved and varies continuously with x, the Gaussian approximation is practical,

$$\tilde{R}_c = \frac{1}{2\Delta}\log\left(\frac{1}{2}+\frac{S}{4}+\frac{1}{2}\sqrt{1+\frac{S^2}{4}}\right) \tag{15}$$

and $\tilde{n}_c = \lceil -\xi^{-1}\log P_e\rceil$, where $\xi = \big[2 + S - \sqrt{4+S^2} - 4\log 2 + 2\log\big(2 - S + \sqrt{4+S^2}\big)\big]/4$ and $S = (e^{2\Delta C} - 1)$ (SM, p. 9). For the model studied: $\tilde{R}_c \doteq 8.9$ bit/s, $\tilde{n}_c = 184$ and $\tilde{n}_c \tilde{R}_c \doteq 1622$ bit/s (Fig. 2).

Discussion. The concept of random input ensembles is employed universally in the information-theoretic literature, and apparently cannot be avoided [6]. One potential problem with this approach is that the average properties of the ensemble tend to be dominated by poorly performing m-sets for small n. We employ a relatively 'noisy' neuronal model and require the probability of decoding error to be small, $P_e \leq 10^{-10}$, naturally increasing the minimal value of n for which at least two input patterns can be formed and transmitted reliably. Similarly we mitigate the potential impact of the neglected $O(1)$ term in Eq. (12) that affects the spread between the lower and upper bounds on R_n below the critical rate. Independently, there is a supporting evidence that the performance of the actual state-of-art Turbo and LDPC codes over some classical information channel models (BSC, AWGN) falls within the mentioned bounds [6, 36]. We therefore consider our estimates of the optimal decoding performance in Fig. 2A to be adequate. Additional simulations also indicate that a change in model parameters preserves the qualitative aspects of Fig. 2A, including the critical behavior (SM, p. 12–14). In future research, we plan to determine the dependence of n_c, R_c on model parameters, number of patterns m required and to balance the unconstrained optimality conditions with the energetic (metabolic) expenses.

The maximum-likelihood decoding rule in Eq. (6) is sometimes proposed as biologically plausible [37]. The expected candidate, however, for a hypothetical decoder inside a living organism, is a neural network. Until comparatively recently, the available learning rules deemed neural networks as unsuitable for channel decoding [38]. The contemporary progress seems to have changed this, as it has been demonstrated that certain convolutional neural networks have a memory capacity that is exponential in the size of the network, provided that the input pattern components are correlated [39]. The network correctly retrieves a set of previously memorized patterns from their noisy versions, and the exponential scaling is theoretically (asymptotically) optimal. Although the input ensemble methods we employ here cannot be used to identify the detailed structure of the optimal sets of input vectors in the ensemble, it is known from coding theory that a statistical dependence between elements of input vectors is generally necessary to achieve reliable transmission [1].

Finally, available information-theoretic methods do not allow to include many of the detailed biological phenomena easily. In particular, different time scales of neural dynamics and realistic connectivity patterns dictate that both dependence on the history and feedback are frequently present in neural systems [40]. Only the model factorizing as Eq. (4) is amenable to the treatment presented in this letter, though. The channel capacity in the presence of memory can be only roughly estimated by its upper and lower bounds [41], and similarly, sufficiently precise evaluation of bounds analogous to Eqs. (9) and (12) is possible only in special cases [1, Ch. 5.9.], [42]. One possibility is to neglect the presumably optimal decoding in Eq. (6) and employ a robust and sufficiently universal mismatched decoding. Nonetheless, even simplified models may provide some insight and a useful description of biological reality [9]. With that said, the main goal of this letter is to bring forward the difference between the asymptotic and actually achievable decoding performance, which, as we believe, universally constraints the optimal information processing in neural systems.

This work was supported by the Czech Science Foundation project 17-06943S, JSPS KAKENHI grant nos. JP17H03279, JP18K11560 and JP19H01133, by JST ACT-I grant no. JPMJPR16UC, and by the Okawa Foundation for Information and Telecommunications. We are also grateful for the open collaborative research and MOU grant from the National Institute of Informatics.

[1] R. G. Gallager, *Information Theory and Reliable Communication* (John Wiley and Sons, Inc., New York, USA, 1968).

[2] M. D. McDonnell and N. G. Stocks, Phys. Rev. Lett. **101**, 058103 (2008).

[3] K. R. Rad and L. Paninski, in *Advances in Neural Information Processing Systems (NIPS)*, Vol. 24, edited by J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger (MIT Press, Cambridge, Massachusetts, 2011) pp. 846–854.

[4] J. Benda and A. V. M. Herz, Neural Comput. **15**, 2523 (2003).

[5] T. Toyoizumi, K. Aihara, and S. Amari, Phys. Rev. Lett. **97**, 98102 (2006).

[6] Y. Polyanskiy, V. H. Poor, and S. Verdu, IEEE Trans. Inf. Theory **56**, 2307 (2010).

[7] M. Tomamichel and V. Y. F. Tan, in *IEEE International Symposium on Information Theory (ISIT), 2013.* (2013) pp. 1536–1540.

[8] H. L. van Trees and K. L. Bell, *Detection, Estimation and Modulation Theory, Part I* (John Wiley and Sons, New York, 2013).

[9] P. Dayan and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (MIT Press, Cambridge, 2001).

[10] M. J. Chacron, B. Lindner, and A. Longtin, Phys. Rev. Lett. **92**, 080601 (2004).

[11] P. E. Greenwood, L. M. Ward, D. F. Russell, A. Neiman, and F. Moss, Phys. Rev. Lett. **84**, 4773 (2000).

[12] L. Barnett, J. T. Lizier, M. Harré, A. K. Seth, and B. T., Phys. Rev. Lett. **111**, 177203 (2013).

[13] N. G. Stocks, Phys. Rev. Lett. **84**, 2310 (2000).

[14] K. S. Turitsyn, S. A. Derevyanko, I. V. Yurkevich, and S. K. Turitsyn, Phys. Rev. Lett. **91**, 203901 (2003).

[15] M. E. Inchiosa, J. W. C. Robinson, and A. R. Bulsara, Phys. Rev. Lett. **85**, 3369 (2000).

[16] L. Kostal, Phys. Rev. E **82**, 026115 (2010).

[17] S. P. Strong, R. Koberle, R. R. de Ruyter van Steveninck, and W. Bialek, Phys. Rev. Lett. **80**, 197 (1998).

[18] L. Kostal, P. Lansky, and M. D. McDonnell, Biol. Cybern. **107**, 355 (2013).

[19] M. Wibral, J. T. Lizier, and V. Priesemann, Front. Robot. AI **2**, 5 (2015).

[20] T. Toyoizumi, J.-P. Pfister, K. Aihara, and W. Gerstner, Proc. Natl. Acad. Sci. USA **102**, 5239 (2005).

[21] J.-P. Pfister, T. Toyoizumi, D. Barber, and W. Gerstner, Neural Comput. **18**, 1318 (2006).

[22] A. Destexhe, M. Rudolph, J. M. Fellous, and T. J. Sejnowski, Neuroscience **107**, 13 (2001).

[23] M. Mattia and P. Del Giudice, Phys. Rev. E **66**, 051917 (2002).

[24] M. Wehr and A. M. Zador, Nature **426**, 442 (2003).

[25] R. W. Berg, A. Alaburda, and J. Hounsgaard, Science **315**, 390 (2007).

[26] M. A. Paradiso, Biol. Cybern. **58**, 35 (1988).

[27] D. Ganguli and E. P. Simoncelli, in *Advances in Neural Information Processing Systems (NIPS)*, Vol. 23, edited by J. Lafferty, C. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta (MIT Press, Cambridge, Massachusetts, 2010) pp. 658–666.

[28] N. Brunel, J. Comput. Neurosci. **8**, 183 (2000).

[29] M. Beiran, A. Kruscha, J. Benda, and B. Lindner, J. Comput. Neurosci. **44**, 189 (2018).

[30] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski, *Neuronal dynamics: From single neurons to networks and models of cognition* (Cambridge University Press, Cambridge, 2014).

[31] J. E. Kelley, J. Soc. Industrial App. Math. **8**, 703 (1960).

[32] J. Huang and S. P. Meyn, IEEE Trans. Inf. Theory **51**, 2336 (2005).

[33] V. Y. F. Tan and M. Tomamichel, IEEE Trans. Inf. Theory **61**, 2430 (2015).

[34] L. Kostal, P. Lansky, and S. Pilarski, J. Neural Eng. **12**, 036012 (2015).

[35] Y. Polyanskiy, *Channel coding: non-asymptotic fundamental limits* (PhD thesis, Princeton University, New Jersey, 2010).

[36] S. Dolinar, D. Divsalar, and F. Pollara, "Code performance as a function of block size," The Telecommunications and Mission Operations Progress Report, TMO PR 42-133 (Jet Propulsion Laboratory, Pasadena, CA, USA) (1998).

[37] S. Deneve, P. E. Latham, and A. Pouget, Nat. Neurosci. **2**, 740 (1999).

[38] R. J. McEliece, E. C. Posner, E. R. Rodemich, and S. S. Venkatesh, IEEE Trans. Inf. Theory **33**, 461 (1987).

[39] A. Karbasi, A. H. Salavati, and A. Shokrollahi, (2014), arXiv:cs.NE/1407.6513.

[40] W. Bialek, *Biophysics: Searching for Principles* (Princeton University Press, Princeton, 2012).

[41] P. Sadeghi, P. O. Vontobel, and R. Shams, IEEE Trans. Inf. Theory **55**, 663 (2009).

[42] A. Trofimov, in *Proceedings of the 14 th International Symposium On Problems of Redundancy in Information and Control Systems, Redundancy 2014, Saint-Petersburg, Russia* (2015) pp. 89–94.