arXiv:submit/1433827 [cond-mat.dis-nn] 24 Dec 2015

December 24, 2015

Neurons as an Information-theoretic Engine

Hideaki Shimazaki *
RIKEN Brain Science Institute
Wako-shi, Satama, Japan
shimazaki@brain.riken.jp

Abstract

We show that dynamical gain modulation of neurons' stimulus response is described as an information-theoretic cycle that generates entropy associated with the stimulus-related activity from entropy produced by the modulation. To articulate this theory, we describe stimulus-evoked activity of a neural population based on the maximum entropy principle with constraints on two types of overlapping activities, one that is controlled by stimulus conditions and the other, termed internal activity, that is regulated internally in an organism. We demonstrate that modulation of the internal activity realises gain control of stimulus response, and controls stimulus information. A cycle of neural dynamics is then introduced to model information processing by the neurons during which the stimulus information is dynamically enhanced by the internal gain-modulation mechanism. Based on the conservation law for entropy production, we demonstrate that the cycle generates entropy ascribed to the stimulus-related activity using entropy supplied by the internal mechanism, analogously to a heat engine that produces work from heat. We provide an efficient cycle that achieves the highest entropic efficiency to retain the stimulus information. The theory allows us to quantify efficiency of the internal computation and its theoretical limit.

1 Introduction

Humans and animals change sensitivity to sensory stimulus either adaptively to the stimulus conditions or following a behavioural context even if the stimulus does not change. A potential neurophysiological basis underlying these observations is gain modulation that changes responsiveness of neurons to stimulus; an example is contrast gain-control found in retina [30] and primary visual cortex under anaesthesia [15, 25], or in higher visual area caused by attention [18, 28]. Theoretical considerations suggested the gain modulation as a nonlinear operation that integrates information from different origins, offering ubiquitous computation performed in neural systems (see [5, 32] for reviews). Regulation of the level of background synaptic inputs [4, 7], shunting inhibition [8, 20, 26], and synaptic depression [1, 29] among others have been suggested as potential biophysical mechanisms of the gain modulation (see [40] for a review). While such modulation of the informative neural activity is a hallmark of computation performed internally in an organism, a principled view to quantify the internal computation has not been proposed yet.

Neurons convey information about the stimulus in their activity patterns. To describe probabilities of a combinatorially large number of activity patterns of the neurons with a smaller number of activity features, the maximum entropy principle has been successfully used [33, 39]. This principle constructs the least structured probability distribution given the small set of specified constraints on the distribution, known as a maximum entropy model. It explains probabilities of activity patterns as a

*Webpage http://2000.jukuin.keio.ac.jp/shimazaki, an alternative email address shimazaki@jhu.edu

result of nonlinear operation on the specified features using a softmax function. Moreover, the model belongs to an exponential family distribution, or a Gibbs distribution. Equivalence of inference under the maximum entropy principle with aspects of the statistical mechanics and thermodynamics was explicated through the work by E. Jaynes [11]. Recently thermodynamic quantities were used to assess criticality of neural activity [44, 45]. However, analysis of neural populations under this framework only recently started to include 'dynamics' of a neural population [10, 12, 13, 23, 35–37], and has not yet reached maturity to include computation performed internally in an organism.

Based on a neural population model obtained under the maximum entropy principle, this study investigates neural dynamics during which gain of neural response to a stimulus is modulated with a delay by an internal mechanism to enhance the stimulus information. This process is expected for dynamics of neurons subject to a feedback gain-modulation mechanism, e.g., via recurrent networks [31, 41, 42]. Regardless of the mechanisms, the delay is observed in the gain modulation at different stages of visual pathways [16, 19, 28]. For example, effect of contrast gain-control by attention on response of V4 neurons to high contrast stimulus appears 200-300 ms after the stimulus presentation, but is absent during 100-200 ms time period during which the neural response is returning to a spontaneous rate [28]. We demonstrate that our hypothetical dynamics of delayed gain-modulation forms an information-theoretic cycle that generates entropy ascribed to the stimulus-related activity using entropy supplied by the internal gain-modulation mechanism. The process works analogously to a heat engine that produces work from heat supplied by reservoirs. We define entropic efficiency of gain-modulation performed to retain the stimulus information, and provide a cycle that achieves the highest entropic efficiency.

This paper is organised as follows. In Section 2 we construct a maximum entropy model of a neural population by constraining two types of activities, one that is directly regulated by stimulus and the other that represents background activity of neurons, termed 'internal activity'. We point out that modulation of the internal activity realises gain-modulation of stimulus response. In Section 3, we explain the conservation of entropy, equation of state for the neural population, and information on stimulus. In Section 4, we construct cycles of neural dynamics that model stimulus-evoked activity during which the stimulus information is enhanced by the internal mechanism. We show that an ideal cycle introduced in this section achieves the highest efficiency in retaining the stimulus information. Derivations of free energies of the neural population are summarised in Appendix.

2 A simple model of gain modulation by a maximum entropy model

Maximum entropy model of spontaneous neural activity. We start by modelling spontaneous activity of N spiking neurons. We represent a state of the ith neuron by a binary variable $x_i = (0, 1)$ ($i = 1 \cdots N$). Here silence of the neuron is represented by '0' whereas activity, or a spike, of the neuron is denoted by '1'. The simultaneous activity of the N neurons is represented by a vector of the binary variables, $\mathbf{x} = (x_1, \ldots, x_N)$. The joint probability mass function, $p(\mathbf{x})$, describes the probability of generating the pattern $\mathbf{x}$. There are 2^N different patterns. We characterise the combinatorial neural activity with a smaller number of characteristic features $F_i(\mathbf{x})$ ($i = 1, \ldots, d$, where $d < 2^N$), based on the maximum entropy principle. Here $F_i(\mathbf{x})$ is the ith feature that combines the activity of individual neurons. For example, these features can be the first and second order interactions, $F_i(\mathbf{x}) = x_i$ for $i = 1, \ldots, N$, and $F_{(N-i/2)(i-1)+j-i}(\mathbf{x}) = x_i x_j$ for $i < j$. The maximum entropy principle constructs the least structured probability distribution while expected values of these features are specified [11]. By representing expectation by $p(\mathbf{x})$ using a bracket $\langle \cdot \rangle$, these constraints are written as $\langle F_i(\mathbf{x}) \rangle = c_i$ ($i = 1, \ldots, d$), where c_i is the specified constant.

Maximisation of a function subject to the equality constraints is formulated by the method of Lagrange multipliers that alternatively maximises the following Lagrange function,

$$\mathcal{L}[p] = -\sum_{\mathbf{x}} p(\mathbf{x}) \log p(\mathbf{x}) - a \sum_{\mathbf{x}} p(\mathbf{x}) - \sum_i b_i \left\{ \sum_{\mathbf{x}} p(\mathbf{x}) F_i(\mathbf{x}) - c_i \right\}, \tag{1}$$

where a and b_i ($i = 1, \ldots, d$) are the Lagrange multipliers. The Lagrange function is a functional of the probability mass function. By finding a zero point of its variational derivative, we obtain

$$p(\mathbf{x}) \sim \exp\left(-\sum_i b_i F_i(\mathbf{x}) \right). \tag{2}$$

The Lagrange parameters b_i are obtained by simultaneously solving $\frac{\partial \mathcal{L}}{\partial b_i} = \langle F_i(\mathbf{x}) \rangle - c_i = 0$ for $i = 1, \ldots, d$. Many gradient algorithms and approximation methods have been developed to search the parameters. Activities of retinal ganglion cells [33, 39, 44, 45], hippocampal [38], and cortical neurons [37, 43, 46] were successfully characterised using Eq. 2. In the following, we use a vector notation $\mathbf{b}_0 = (b_1, ..., b_d)^T$ and $\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), \ldots, F_d(\mathbf{x}))^T$. Here $\mathcal{H}_0 \equiv \mathbf{b}_0^T \mathbf{F}(\mathbf{x})$ is a Hamiltonian of the spontaneously active neurons. In statistical mechanics, Eq. 2 is identified as the Boltzmann distribution with an unit thermodynamic *beta*.

Maximum entropy model of evoked neural activity. In this subsection, we model evoked activity of neurons caused by changes in extrinsic stimulus conditions. We define a feature of stimulus-related activity as $X(\mathbf{x}) = \mathbf{b}_1^T \mathbf{F}(\mathbf{x})$, where elements of $\mathbf{b}_1$ dictate response properties of each feature in $\mathbf{F}(\mathbf{x})$ to a stimulus. For simplicity, we represent the stimulus-related activity by this single feature, and consider that the evoked activity is characterised by the two features, $\mathcal{H}_0(\mathbf{x})$ and $X(\mathbf{x})$. To model it, we constrain expectation of the internal and stimulus features using U and X, respectively. Here we assume that $\mathbf{F}(\mathbf{x})$, $\mathbf{b}_0$, and $\mathbf{b}_1$ are known and fixed. For example, this would model responses of visual neurons when we change contrast of a stimulus while fixing the rest of the stimulus properties. The maximum entropy distribution subject to these constraints is again given by the method of Lagrange multipliers. The Lagrange function is given as

$$\begin{aligned}\mathcal{L}[p] = &- \sum_{\mathbf{x}} p(\mathbf{x}) \log p(\mathbf{x}) \\ &- a \sum_{\mathbf{x}} p(\mathbf{x}) - \beta \left\{ \sum_{\mathbf{x}} p(\mathbf{x}) \mathcal{H}_0(\mathbf{x}) - U \right\} + \alpha \left\{ \sum_{\mathbf{x}} p(\mathbf{x}) X(\mathbf{x}) - X \right\}. \end{aligned} \tag{3}$$

Here a, β, and α are the Lagrange parameters. By maximising the functional $\mathcal{L}$ with respect to p, we obtain the following maximum entropy model,

$$p(\mathbf{x}) = \exp[-\beta \mathcal{H}_0(\mathbf{x}) + \alpha X(\mathbf{x}) - \psi(\beta, \alpha)], \tag{4}$$

where $\psi(\beta, \alpha)(= 1 + a)$ is a logarithm of a normalisation term. It is computed as

$$\psi(\beta, \alpha) = \log \sum_{\mathbf{x}} e^{-\beta \mathcal{H}_0(\mathbf{x}) + \alpha X(\mathbf{x})}. \tag{5}$$

We call $\psi(\beta, \alpha)$ a log-partition function. The Lagrange multipliers, β and α, are adjusted such that $\langle \mathcal{H}_0(\mathbf{x}) \rangle = U$ and $\langle X(\mathbf{x}) \rangle = X$. Eq. 4 is a softmax function (generalisation of a logistic function to multinomial outputs) that returns the population output from a linear sum of the features weighted by $-\beta$ and α. With this view, we may alternatively regard β or α as an input parameter that controls U and X. Hereafter we simply call U internal activity, and X stimulus-related activity. Similarly, we call β an internal component, and α a stimulus component. We consider that the stimulus component α can be controlled by changing extrinsic stimulus conditions that an experimenter can manipulate. The stimulus component is written as $\alpha(s)$ if it is a function of a scalar stimulus condition s, such as stimulus contrast. In contrast, the internal activity β is not directly controllable by the stimulus conditions. The spontaneous activity is modelled at $\beta = 1$ and $\alpha = 0$.

Gain modulation by internal activity. We give a simple example of the maximum entropy model to show how the internal activity modulates the stimulus-related activity. Figure 1a illustrates an exemplary model composed of 5 neurons. With these particular model parameters (see figure caption), the stimulus component α controls activity rates of the first three neurons and their correlations. The internal component β controls background activity rates of all neurons. In our settings, decreasing β increases the baseline activity level of all neurons. Figure 1b displays activity rates of the individual neurons ($\langle x_i \rangle$ for $i = 1, \ldots, 5$) as a function of the stimulus component α with a fixed internal component β. Increasing α under these conditions activates the first three neurons without changing the activity rates of Neuron 4 and 5[1]. Furthermore, the response functions of the three neurons shift toward left when the background activity rates of all neurons is increased by *decreasing* the internal component β (Fig. 1b dashed lines). Thus Neuron 1-3 increase sensitivity to stimulus component α. This type of modulation is called input-gain control. For example, if α

[1]The activity rates of Neuron 4, 5 do not depend on α because $\mathbf{b}_0$ does not contain interactions that relate Neuron 1-3 with Neuron 4, 5. If there are non-zero interactions between any pair from Neuron 1-3 and Neuron 4, 5 in $\mathbf{b}_0$, the activity rates of Neuron 4, 5 increase with the increased rates of Neuron 1-3.

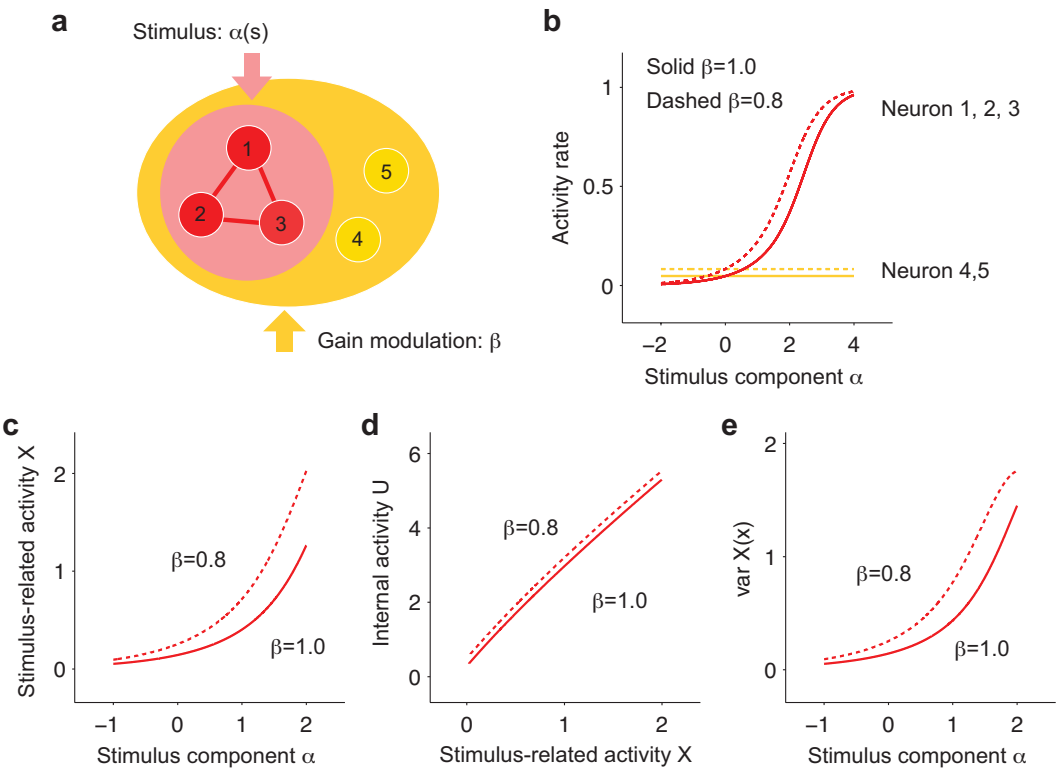

Figure 1: A simple model of gain modulation by a maximum entropy model of 5 neurons. (a) An illustration of neurons that are activated by a stimulus (neurons in a pink area) and controlled by an internal mechanism (neurons in a yellow area). The model is constrained by features containing up to the second order statistics: $\mathbf{F}(\mathbf{x}) = (x_1, \ldots, x_5,\ x_1x_2, x_1x_3, x_2x_3, \ldots, x_4x_5)^T$, where the first 5 elements are parameters for the individual activities x_i ($i = 1, \ldots, 5$) and the rest of the elements is the joint activities of two neurons x_ix_j ($i < j$). We assume that the stimulus-related activity is characterised by $\mathbf{b}_1 = (1, 1, 1, 0, 0,\ 0.3, 0.3, 0.3, 0, \ldots, 0)$. The first 3 elements are parameters for individual activity of the first three neurons x_i ($i = 1, 2, 3$). The value 0.3 is assigned to the joint activities of the first three neurons, namely the features specified by x_1x_2, x_1x_3, and x_2x_3. The internal activity is characterised by $\mathbf{b}_0 = (2, 2, 2, 2, 2, 0, \ldots, 0)$, which regulates activity rates of individual neurons but does not change their interactions. (b) The activity rates of neurons as a function of the stimulus component α at fixed internal components, $\beta = 1.0$ (solid line) and $\beta = 0.8$ (dashed line). (c) The stimulus component X as a function of α at different internal components. (d) The relation between the stimulus-related activity X and internal activity U. (e) The Fisher information about the stimulus component α.

is a logarithmic function of contrast s of visual stimulation presented to an animal while recording visual neurons ($\alpha(s) = \log s$), increasing the modulation (decreasing β) makes neurons respond to multiplicatively smaller stimulus contrast. This models the contrast gain-control observed in visual pathways [18, 25, 28, 30]. Other types of nonlinearity in the input-output relation can be constructed, depending on the nonlinearity in $\alpha(s)$.

Figure 1c displays a relation of the stimulus component α with the stimulus-related activity X at different internal component β. Similarly to the activity rates (Fig. 1b), the stimulus-related activity X is augmented if the internal component β is decreased. This nonlinear interaction between α and β is caused by the neurons that belong to both stimulus-related and internal activities. In this example, the stimulus component α also increases the internal activity U (Fig. 1d) because of increased activity rates of the shared neurons 1, 2, 3. Finally, Figure 1e displays the variance of stimulus feature $X(\mathbf{x})$ as a function of α. It quantifies the information about the stimulus component α, which we will discuss in the next section.

3 The conservation of entropy, equation of state, and stimulus information

Conservation of entropy for neural dynamics. The probability mass function, Eq. 4, belongs to the exponential family distribution. The Lagrange parameters are called natural or canonical parameters. The activity patterns of neurons are modelled as a linear combination of the two features $\mathcal{H}_0(\mathbf{x})$ and $X(\mathbf{x})$ using the canonical parameters $(-\beta, \alpha)$ in the exponent. Expectation of the features are called the expectation parameters U and X. Either natural or expectation parameters are sufficient to specify the probability distribution. We review dual structure of the two representations [2], and show that the relation provides the conservation law of entropy.

Negative entropy of the neural population is computed as

$$\begin{aligned} -S &= \langle \log p(\mathbf{x}) \rangle \\ &= -\beta \langle \mathcal{H}_0(\mathbf{x}) \rangle + \alpha \langle X(\mathbf{x}) \rangle - \psi(\beta, \alpha) \\ &= -U\beta + X\alpha - \psi(\beta, \alpha). \end{aligned} \tag{6}$$

Since the log-partition function of Eq. 4 is a cumulant generating function, U and X are related to the derivatives of $\psi(\beta, \alpha)$ as

$$\frac{\partial \psi(\beta, \alpha)}{\partial \beta} = -\langle \mathcal{H}_0(\mathbf{x}) \rangle = -U, \tag{7}$$

$$\frac{\partial \psi(\beta, \alpha)}{\partial \alpha} = \langle X(\mathbf{x}) \rangle = X. \tag{8}$$

Eqs. 6, 7 and 8 form a Legendre transformation from $\psi(\beta, \alpha)$ to $-S(U, X)$. The inverse Legendre transformation is constructed using Eq. 6 as well: $\psi(\beta, \alpha) = -\beta U + \alpha X - (-S(U, X))$. Thus dually to Eqs. 7 and 8, the natural parameters are obtained as derivatives of the entropy with respect to the expectation parameters,

$$\left(\frac{\partial S}{\partial U} \right)_X = \beta, \tag{9}$$

$$\left(\frac{\partial S}{\partial X} \right)_U = -\alpha. \tag{10}$$

The natural parameters represent sensitivities of the entropy to the independent variables U and X. From these results, the total derivative of $S(U, X)$ is written as

$$\begin{aligned} dS &= \left(\frac{\partial S}{\partial U} \right)_X dU + \left(\frac{\partial S}{\partial X} \right)_U dX \\ &= \beta dU - \alpha dX. \end{aligned} \tag{11}$$

This explains a change of neurons' entropy by changes in the internal and stimulus-related activities. We denote an entropy change caused by the internal activity as $dS^{\mathrm{int}} \equiv \beta dU$, and an entropy change caused by the extrinsic stimulus as $dS^{\mathrm{ext}} \equiv \alpha dX$, respectively. Then Eq. 11 is written as

$$dS = dS^{\mathrm{int}} - dS^{\mathrm{ext}} \tag{12}$$

We remark that dS is an infinitesimal difference of entropies at two close states, and its integral does not depend on a specific transition between the two states. In contrast, dS^{int} and dS^{ext} represent production of entropy separately by the internal and stimulus-related activities, and their integrals depend on the specific paths. Eq. 12 constitutes the conservation of entropy for neural dynamics. We stress that although it is the first law of thermodynamics, the neurons considered here interact with an environment differently from conventional thermodynamic systems[2]. While internal energy of the conventional systems is indirectly controlled via work and heat, we consider that the internal activity of neurons is controlled directly by the organism's internal mechanism. Thus we use dS^{int} and dS^{ext}, rather than the work and heat, as quantities that neurons exchange with an environment.

[2]We obtain $dU = TdS - fdX$, using $\beta \equiv 1/T$ and $\alpha \equiv \beta f$ in Eq. 11. In this form, the expectation parameter U is a function of (S, X). According to the conventions of thermodynamics, we may call U internal energy, T temperature of the system, and f force applied to neurons by a stimulus. It is possible to describe the evoked activity of a neural population using these standard terms of thermodynamics. However, this introduces the concepts of work and heat.

Equation of state for a neural population. Eq. 8 is an equation of the state for a neural population, which we rewrite here as

$$X(\beta, \alpha) = \frac{\partial \psi(\beta, \alpha)}{\partial \alpha}. \tag{13}$$

Through the log-partition function ψ, this equation relates state variables, β, α, and X, similarly to e.g., the classical ideal gas law that relates temperature, pressure, and volume. Figure 1c displayed the equation of state. We note that ψ is related to the Gibbs free energy (see Appendix). Furthermore, without loss of generality, we can assume that the ground state of the features is zero: $\mathcal{H}_0(\mathbf{0}) = X(\mathbf{0}) = 0$, where $\mathbf{x} = \mathbf{0}$ denotes the simultaneous silence of all neurons. We then obtain $p(\mathbf{0}) = e^{-\psi}$, namely

$$-\psi(\beta, \alpha) = \log p(\mathbf{0}). \tag{14}$$

Thus $-\psi(\beta, \alpha)$ is a logarithm of the simultaneous silence probability[3]. Since $d(\log p(\mathbf{0})) = dp(\mathbf{0})/p(\mathbf{0})$, $-d\psi$ gives a fractional increase of the simultaneous silence probability of the neurons. Accordingly Eq. 13 states that the stimulus-related activity X equals to the fractional decrease of the simultaneous silence probability by a small change of α, given β.

Information about stimulus. The Fisher information $J(\alpha)$ provides the accuracy of estimating a small change in the stimulus component α by an optimal decoder. More specifically, the inverse of the Fisher information provides a lower bound of variance of an unbiased estimator for α from a sample. For the exponential family distribution, it is given as the second order derivative of the log-partition function with respect to α, which is also the variance of stimulus feature $X(\mathbf{x})$:

$$\begin{aligned} J(\alpha) &\equiv \left\langle \left(\frac{\partial \log p(\mathbf{x})}{\partial \alpha} \right)^2 \right\rangle = \frac{\partial^2 \psi(\beta, \alpha)}{\partial \alpha^2} \\ &= \frac{\partial X}{\partial \alpha} = \langle X(\mathbf{x})^2 \rangle - \langle X(\mathbf{x}) \rangle^2. \end{aligned} \tag{15}$$

The first equality in the second line of Eq. 15 is obtained using the first order derivative of ψ, namely the equation of state (Eq. 13). The second equality in Eq. 15 represents the fluctuation-dissipation relation of the stimulus feature. The equalities show that the Fisher information can be computed in three different manners given that the internal component β is fixed: (i) the second derivative of ψ with respect to α using the simultaneous silence probability, (ii) the derivative of X with respect to α using the equation of state, or (iii) the variance of the stimulus feature.

The Fisher information computed at two fixed internal components was shown in Fig. 1d. The stimulus component α becomes relatively dominant in characterising the neural activity if the internal component β decreases. This results in the larger Fisher information $J(\alpha)$ for the smaller internal component β at given α. If the stimulus condition s controlls the stimulus component as $\alpha(s)$, and it is not related to β, the information about s is given as $\frac{\partial \alpha(s)}{\partial s} J(\alpha) \frac{\partial \alpha(s)}{\partial s}$.

4 Information-theoretic cycles by a neural population

We now introduce neural dynamics that models dynamical gain-modulation performed by an internal mechanism while neurons are processing stimulus. Since there are neurons that belong to both stimulus-related and internal activities, the internal mechanism not only changes the internal activity but also the stimulus-related activity, which realises the modulation. From an information-theoretic point of view, this process converts entropy generated by the internal mechanism to entropy associated with stimulus-related activity after one cycle of the neural response is completed. To explain this in detail, we first provide an intuitive example of delayed gain-modulation using a dynamical model, and then provide an ideal cycle that efficiently enhance stimulus information. Using the latter model, we explain why the process works similarly to a heat engine, and show how to quantify efficiency of the gain-modulation performed by the internal mechanism.

[3]Importantly, $-\psi$ is a logarithm of the simultaneous silence probability predicted by the model, Eq. 4. The observed probability of the simultaneous silence could be different from the prediction if the model is inaccurate. For example, an Ising model can be inaccurate, and it was shown that neural higher-order interactions significantly contribute to increasing the silence probability [24, 38].

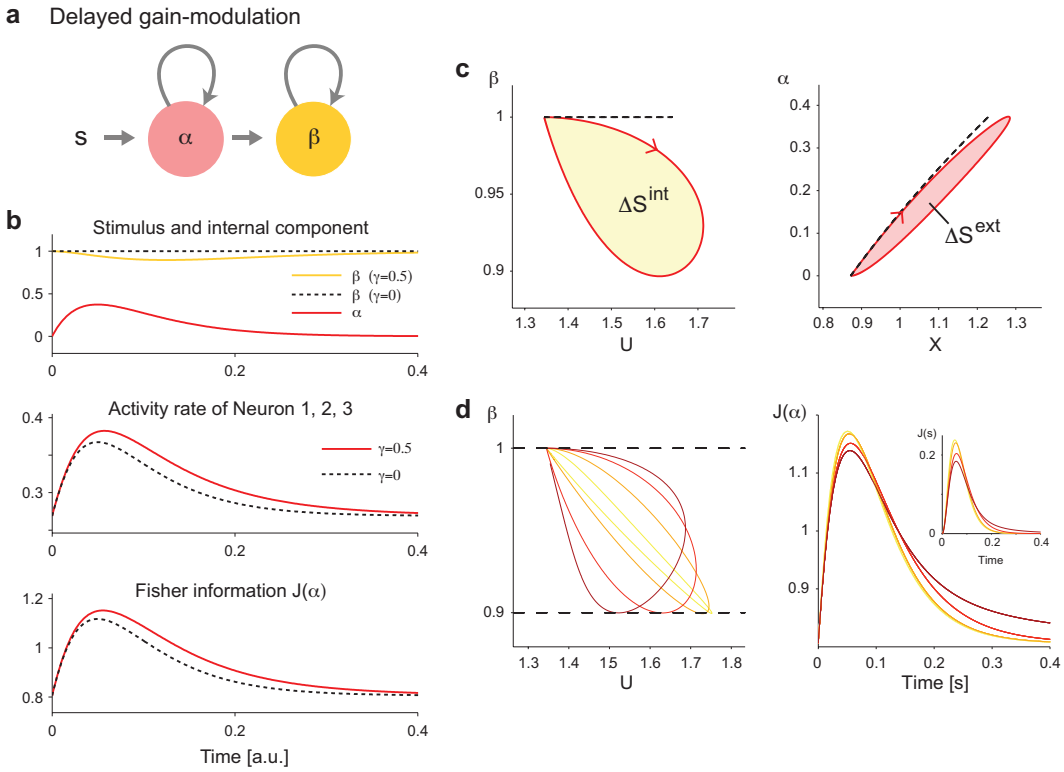

Figure 2: The delayed gain-modulation by internal activity. The parameters of the maximum entropy model ($N = 5$) follow those in Fig. 1. (a) An illustration of delayed gain-modulation described in Eqs. 16 and 17. The stimulus increases the stimulus component α that activates Neuron 1, 2, and 3. Subsequently, the internal component β is increased, which increases the background activity of all 5 neurons. We assume a slower time constant for the gain-modulation than the stimulus activation ($\tau_\beta = 0.1$ and $\tau_\alpha = 0.05$). (b) Top: Dynamics of the stimulus and internal components (solid lines, $\gamma = 0.5$). The internal component β without the delayed gain-modulation ($\gamma = 0$) is shown by a dashed black line. Middle: Activity rates [a.u.] of Neuron 1-3 with (solid red) and without (dashed black) the delayed gain-modulation. Bottom: The Fisher information about stimulus component α (Eq. 15). (c) The X-α (Left) and U-β (Right) phase diagrams. A red solid cycle represents dynamics when the delayed gain-modulation is applied ($\gamma = 0.5$). The dashed line is a trajectory when the delayed gain-modulation is not applied to the population ($\gamma = 0$). (d) Left: The U-β phase diagrams of neural dynamics with different combinations of τ_β and γ that achieve the same level of the maximum modulation (the minimum value of $\beta = 0.9$). Right: The Fisher information about the stimulus component α for different cycles. The colour code is the same as in the Left panel. The inset shows the Fisher information about the stimulus intensity s (Eq. 18).

An example of delayed gain-modulation. We first consider a simple dynamical model of delayed gain-modulation. We use the feature vector, $\mathbf{b}_0$ and $\mathbf{b}_1$ based on those described in Fig. 1. In this model, neurons are activated by a stimulus input, which subsequently increases modulation by an internal mechanism (Fig. 2a). Such a process can be modelled through dynamics of the controlling parameters given by,

$$\tau_\alpha^2 \dot{\alpha}(t) = -\tau_\alpha \alpha(t) + s\, e^{-t/\tau_\alpha} \tag{16}$$

$$\tau_\beta \dot{\beta}(t) = -\beta(t) + \beta_0 - \gamma \alpha(t) \tag{17}$$

for $t \geq 0$. Here s is intensity of an input stimulus. Neurons are initially at a spontaneous state: $\alpha(0) = 0$ and $\beta(0) = \beta_0 = 1$. The top panel of Figure 2b displays the dynamics of $\alpha(t)$ and $\beta(t)$. The population activity is sampled from the maximum entropy model with these dynamical

parameters. Here we consider a continuous-time representation of the maximum entropy model[4] [12, 13]. The activity rates of neurons are increased by the delayed gain-modulation (solid lines in Fig. 2b Middle) from those obtained without the modulation ($\gamma = 0$; dashed lines). Accordingly, the information about the stimulus component α contained in the population activity as quantified by the Fisher information (Eq. 15) increases and lasts longer by the delayed gain-modulation (Fig. 2b Bottom). Note that in this example, the information about the stimulus strength s is carried in both $\beta(t)$ and $\alpha(t)$ as time passes. The result obtained from the Fisher information about s using both $\beta(t)$ and $\alpha(t)$ is qualitatively the same as the result of the Fisher information about α (not shown) [5].

The U-β phase diagram (Fig. 2c Left) shows that dynamics without the gain-modulation is represented as a line because β is constant. In contrast, dynamics with the gain-modulation forms a cycle because weaker and then stronger modulation (larger and then smaller β) is applied to neurons when the internal activity U increases and then decreases, respectively. Similarly, the dynamics forms a cycle in the X-α plane (Fig. 2c Right) if the stimulus activity X is augmented by the delayed gain-modulation. By applying the conservation law for entropy (Eq. 12) to the cycle, we obtain

$$0 = \oint \beta dU - \oint \alpha dX. \tag{19}$$

Here $\oint \beta dU \equiv \Delta S^{\mathrm{int}}$ is entropy produced by the internal activity during the cycle due to the delayed gain-modulation, and $\oint \beta dU \equiv \Delta S^{\mathrm{ext}}$ is entropy produced by the activity related to extrinsic stimulus condtions. These are the areas within the circles in the phase diagrams. Eq. 19 states that the two cycles have the same area ($\Delta S^{\mathrm{int}} = \Delta S^{\mathrm{ext}}$).

The left panel in Figure 2d displays the U-β phase diagram for dynamics with given maximum strength of modulation (the minimum value of β). Among these cycles, larger cycles retain the information about the stimulus component α for a longer time period (Fig. 2d Right). The same conclusion is made from the Fisher information about s (Fig. 2d an inset in Right panel). The larger cycles were made because the modulation was only weakly applied to neurons when the internal activity U increased, then the strong modulation was applied when U decreased. Such modulation is considered to be efficient because it allows neurons to retain the stimulus information for a longer time period by using the slow time-scale of β without excessively increasing activity rates of neurons at its initial rise. In the next section, we introduce the largest cycle that maximises the entropy produced by the gain modulation when the maximum strength of the modulation is given. Using this cycle, we explain how the cycle works analogously to a heat engine, and define efficiency of the cycle to retain the stimulus information.

The efficient cycle by a neural population. The largest cycle is made if the modulation is not applied when the internal activity U increases, then applied when U decreases. Figure 3 displays a cycle of hypothetical neural dynamics that maximises the entropy production when the ranges of the internal component and activity are given. The model parameters follow those in Fig. 1. This cycle is composed of four steps. The process starts at the state A at which neurons exhibit spontaneous activity ($\beta = \beta_H = 1$, $\alpha = 0$). Figure 3a displays a sample response of the neural population to a stimulus change. Figure 3b and c display the X-α and U-β phase diagrams of the cycle. Heat capacity of the neural population and the Fisher information about α are shown in Fig. 3d. Details of the cycle steps are now described as follows.

[4]Under the assumption that rates of synchronous spike events scale with $\mathcal{O}(\Delta^k)$, where Δ is a bin size of discretisation and k is the number of synchronous neurons, Kass *et al.* [12] proved that it is possible to construct a continuous-time limit ($\Delta \to 0$) of the maximum entropy model that takes the synchronous events into account. Here we follow their result to consider the continuous-time representation.

[5]When α and β are both dependent on the stimulus, the Fisher information about s is given as

$$J(s) = \frac{\partial \boldsymbol{\theta}(s)^T}{\partial s} \mathbf{J} \frac{\partial \boldsymbol{\theta}(s)}{\partial s}, \tag{18}$$

where $\boldsymbol{\theta}(s) \equiv (-\beta, \alpha)^T$ and $\mathbf{J}$ is a Fisher information matrix given by Eq. 23, which will be discussed in the later section. We computed Eq. 18 using analytical solutions of the dynamical equations given as $\alpha(t) = \frac{st}{\tau_\alpha} e^{-t/\tau_\alpha}$ and $\beta(t) = 1 - \frac{sg}{\tau_\beta - \tau_\alpha} \left\{ \frac{\tau_\alpha \tau_\beta}{\tau_\beta - \tau_\alpha} (e^{-t/\tau_\beta} - e^{-t/\tau_\alpha}) - t e^{-t/\tau_\alpha} \right\}$.

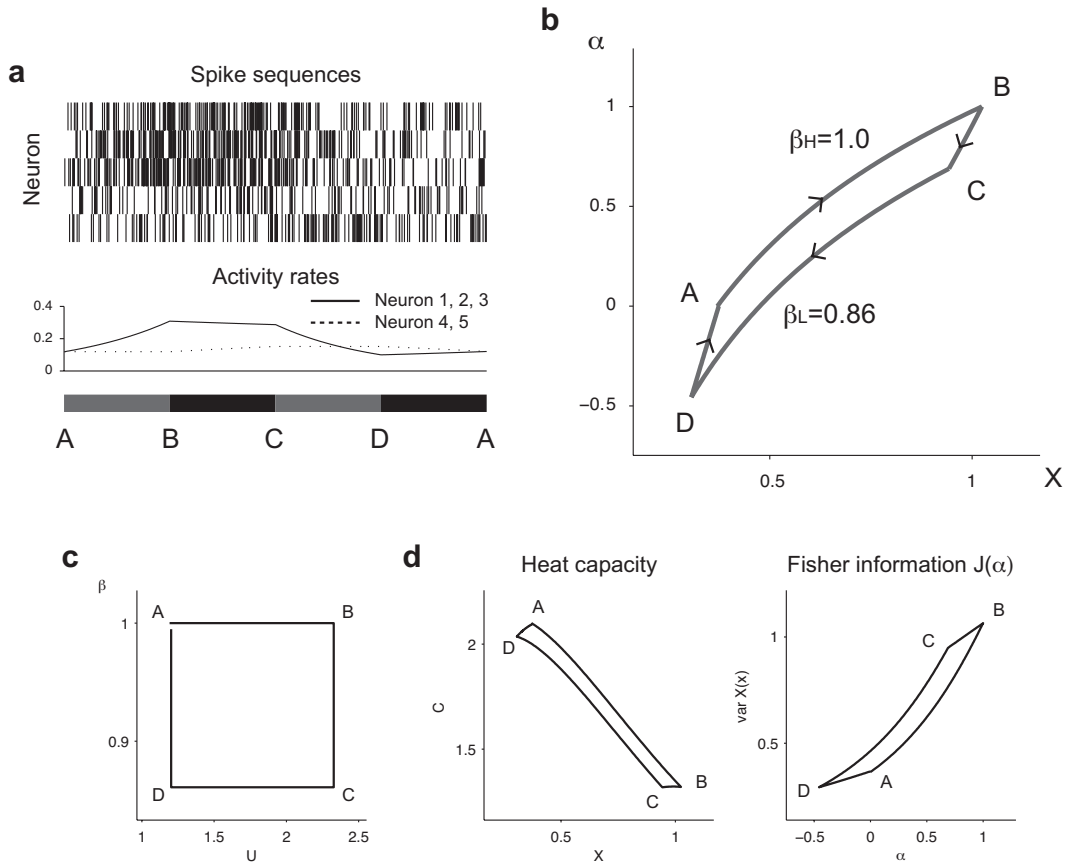

Figure 3: The efficient circle by a neural population ($N = 5$). The parameters of the maximum entropy model follow those in Fig. 1. The cycle starts from the state A at which $\beta = \beta_H = 1$ and $\alpha = 0$. See the main text for details of the steps. The efficiency of this cycle is 0.14. (a) Top: Spike raster plots during the cycle. Middle: Activity rates of neurons. Bottom: The cycle steps. (b) The X-α phase diagram. (c) The U-β phase diagram. (d) Left: X v.s. heat capacity. The heat capacity is defined as $C = \langle h^2 \rangle - \langle h \rangle^2$, where $h = -\log p(\mathbf{x})$ is information content. Right: Fisher information about the stimulus component α.

A→B **Increased stimulus response** The stimulus-related activity X is increased by increasing the stimulus component α while the internal component is fixed at $\beta = \beta_H$. In this process the internal activity U also increases.

B→C **Internal computation** An internal mechanism decreases the internal component β while keeping the internal activity ($dU = 0$). In this process the stimulus-related activity X decreases. The process ends at $\beta = \beta_L$.

C→D **Decreased stimulus response** The stimulus-related activity X is decreased by decreasing the stimulus component α while the internal component is fixed at $\beta = \beta_L$. In this process the internal activity U also decreases.

D→A **Internal computation** An internal mechanism increases the internal component β while keeping the internal activity ($dU = 0$). In this process the stimulus-related activity X increases. The process ends at $\beta \equiv \beta_H$.

The processes B→C and D→A represent additional computation performed by an internal neural mechanism on the neurons' stimulus information processing. It is applied after the initial increase of stimulus-related activity during A→B, therefore manifests delayed modulation. Without these processes, the neural dynamics is represented as a line in the phase diagrams. The Fisher information about α also increases during the process between C and D (Fig. 3d right panel). We reiterate that the

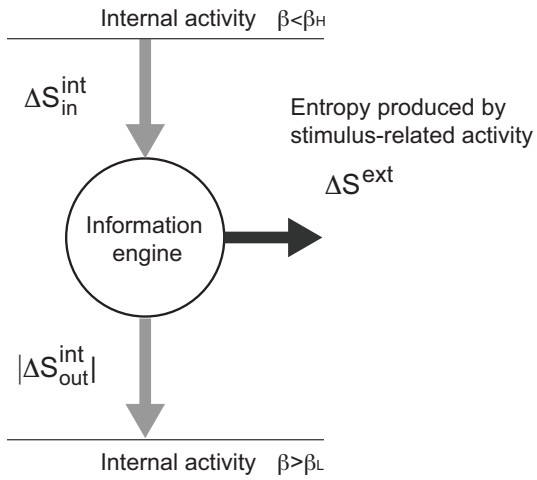

Figure 4: An information-theoretic cycle by a neural population.

Fisher information quantifies the accuracy of estimating a small change in α by an optimal decoder. Thus operating along the path between C and D is more advantageous than the path between A and B for downstream neurons if their goal is to detect a change in the stimulus-related activity of the upstream neurons that is not explained by the internal activity.

Interpretation as an information-theoretic cycle. We start our analysis on the cycle by examining how much entropy is generated by the internal and stimulus-related activities at each step. First, we denote by $\Delta S^{\rm int}_{\rm AB}$ and $\Delta S^{\rm int}_{\rm CD}$ the entropy changes caused by the internal activity during the process A→B and C→D, respectively. Since the internal component β is fixed at β_H during the process A→B, we obtain $\Delta S^{\rm int}_{\rm AB} = \beta_H \Delta U$, where ΔU is a change of the internal activity (see Fig. 3c). This change in the internal activity is positive ($\Delta U > 0$). Since the internal activity does not change during B→C and D→A, a change of the internal activity during C→D is given by $-\Delta U$ (Note that the internal activity is a state variable). We obtain $\Delta S^{\rm int}_{\rm CD} = -\beta_L \Delta U$ for the process during C→D. The total entropy change caused by the internal activity during the cycle is given as $\Delta S^{\rm int}_{\rm AB} + \Delta S^{\rm int}_{\rm CD} = (\beta_H - \beta_L)\Delta U$, which is positive because $\beta_H > \beta_L$ and $\Delta U > 0$. Thus the internal activity increases the entropy of neurons during the cycle. Second, we denote by $\Delta S^{\rm ext}$ the total entropy change caused by the stimulus-related activity during the cycle. According to the conservation law (Eq. 12) applied to this cycle, we obtain

$$0 = \Delta S^{\rm int}_{\rm AB} + \Delta S^{\rm int}_{\rm CD} - \Delta S^{\rm ext}. \tag{20}$$

Note that the sign of $\Delta S^{\rm ext} = \Delta S^{\rm int}_{\rm AB} + \Delta S^{\rm int}_{\rm CD}$ is positive. Hence the stimulus-related activity decreases the entropy of neurons during the cycle.

This cycle belongs to the following cycle that is analogous to a heat engine (Fig. 4). In this paragraph, we temporarily use *receive entropy* and *emit entropy* to express the positive and negative path-dependent entropy changes caused by the internal or stimulus-related activity in order to facilitate comparison with a heat engine[6]. In this cycle, neurons receive *entropy* as internal activity from an environment ($\Delta S^{\rm int}_{\rm in} > 0$) and emit *entropy* to the environment ($\Delta S^{\rm int}_{\rm out} < 0$). The received *entropy* as the internal activity is larger than the emitted *entropy* ($\Delta S^{\rm int}_{\rm in} + \Delta S^{\rm int}_{\rm out} > 0$). The surplus *entropy* is emitted to the environment in the form of the stimulus-related activity ($-\Delta S^{\rm ext} < 0$). Thus we may regard the cycle as the process that produces stimulus-related entropy using entropy supplied by the internal mechanism. We hereafter denote this cycle as an information-theoretic cycle, or engine. The cycle in Fig. 2 is also regarded as an information-theoretic cycle by separating the process at which the internal activity is maximised. The conservation law prohibits a perpet-

[6]Here we use *entropy* synonymously with heat in thermodynamics to facilitate the comparison with a heat engine. However this is not an accurate description because the entropy is a state variable.

ual information-theoretic cycle that can indefinitely produce the stimulus-related entropy without entropy production by the internal mechanism[7].

Efficiency of a cycle. As we discussed for the example dynamics in Fig. 2, we may consider that the modulation is efficient if it helps neurons to retain stimulus information without excessively increasing the internal and stimulus-related activities during the initial response. Such a process was achieved when gain modulation was only weakly applied to neurons when the internal activity U increased, then strong gain modulation was applied when U decreased. We can formally assess this type of efficiency by defining entropic efficiency, similarly to thermal efficiency of a heat engine. It is given by a ratio of the entropy change caused by the stimulus-related activity as opposed to the entropy change gained by the internal activity as:

$$\eta \equiv \frac{\Delta S^{\mathrm{ext}}}{\Delta S^{\mathrm{int}}_{\mathrm{in}}} = 1 - \frac{|\Delta S^{\mathrm{int}}_{\mathrm{out}}|}{\Delta S^{\mathrm{int}}_{\mathrm{in}}}. \tag{21}$$

For the proposed information-theoretic cycle in Fig. 3, it is computed as

$$\eta_e = 1 - \frac{|\Delta S^{\mathrm{int}}_{\mathrm{CD}}|}{\Delta S^{\mathrm{int}}_{\mathrm{AB}}} = 1 - \frac{\beta_L}{\beta_H}, \tag{22}$$

which is a function of the internal components, β_H and β_L. This cycle is the most efficient in terms of the entropic efficiency defined by Eq. 21 when the highest and lowest internal components and activities are given. The square cycle in the U-β phase diagram (Fig. 3c) already suggests this claim, and we can formally prove this by comparing the information-theoretic cycle with an arbitrary cycle $\mathcal{C}$ whose internal component β satisfies $\beta_L \leq \beta \leq \beta_H$[8]. Thus the proposed cycle bounds efficiency of the additional computation made by the delayed gain-modulation mechanism. We call the proposed cycle in Fig. 3, the ideal information-theoretic cycle. Note that this cycle is similar to, but different from the Carnot cycle [6] that can be realised by replacing the processes B→C and D→A with adiabatic processes. The Carnot cycle achieves the highest *thermal* efficiency.

Geometric interpretation. Finally, to consider conditions for the information-theoretic cycle, we introduce geometric interpretation of the cycle. Let us denote the internal and stimulus components as $\boldsymbol{\theta} = (-\beta, \alpha)^T$. In addition, we represent the expected internal and stimulus features by $\boldsymbol{\eta} = (U, X)^T$. The parameters $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ form dually flat affine coordinates, and are called θ and η-coordinates in information geometry [2]. For the ideal information-theoretic cycle, we indicate the parameters at A, B, C, and D using a subscript of $\boldsymbol{\theta}$ or $\boldsymbol{\eta}$. For example the parameters at A are $\boldsymbol{\theta}_A$ and $\boldsymbol{\eta}_A$.

The first process A→B of the ideal information-theoretic cycle is a straight line (geodesic) between $\boldsymbol{\theta}_{\mathrm{A}}$ and $\boldsymbol{\theta}_{\mathrm{B}}$ in the curved space of θ-coordinates. It is called e-geodesic. In addition, the internal component β is fixed while the stimulus component decreases, therefore the e-geodesic is a vertical line in the θ-coordinates. The second process B→C is the shortest line between $\boldsymbol{\eta}_{\mathrm{B}}$ and $\boldsymbol{\eta}_{\mathrm{C}}$ in the curved space of η-coordinates. The path is called an m-geodesic. In addition, the internal activity U is fixed while the stimulus-related activity decreases, therefore the m-geodesic is a vertical line in the η-coordinates. Similarly, the process C→D is an e-geodesic, and the process D→A is an m-geodesic.

The change in the internal component β during the processes along m-geodesic manifested the internal computation in the ideal information-theoretic cycle. A small change in $\boldsymbol{\theta}$ is related to a change in $\boldsymbol{\eta}$ as $d\boldsymbol{\eta} = \mathbf{J} d\boldsymbol{\theta}$. Here $\mathbf{J}$ is the Fisher information matrix with respect to $\boldsymbol{\theta}$. It is given as

$$\mathbf{J} = \begin{bmatrix} \langle \mathbf{b}_0, \mathbf{b}_0 \rangle & \langle \mathbf{b}_0, \mathbf{b}_1 \rangle \\ \langle \mathbf{b}_1, \mathbf{b}_0 \rangle & \langle \mathbf{b}_1, \mathbf{b}_1 \rangle \end{bmatrix}, \tag{23}$$

[7]This is synonymous with the statement that the first law prohibits a perpetual motion machine of the first kind, a machine that can work indefinitely without receiving heat.

[8]Let us consider the efficiency η achieved by an arbitrary cycle $\mathcal{C}$ during which the internal component β satisfies $\beta_L \leq \beta \leq \beta_H$. Let the minimum and maximum internal activity in the cycle be $U_{\min}$ and $U_{\max}$. We decompose $\mathcal{C}$ into the path $\mathcal{C}_1$ from $U_{\min}$ to $U_{\max}$ and the path $\mathcal{C}_2$ from $U_{\max}$ to $U_{\min}$ during which the internal component is given as $\beta_1(U)$ and $\beta_2(U)$, respectively. Because the cycle acts as an engine, we expect $\beta_1(U) > \beta_2(U)$. The entropy changes produced by the internal activity during the path C_i ($i = 1, 2$) is computed as $\Delta S^{\mathrm{int}}_{\mathcal{C}_1} = \int_{U_{\min}}^{U_{\max}} \beta_1(U)\, dU \leq \beta_H \int_{U_{\min}}^{U_{\max}} dU = \beta_H (U_{\max} - U_{\min})$ and $|\Delta S^{\mathrm{int}}_{\mathcal{C}_2}| = |\int_{U_{\max}}^{U_{\min}} \beta_2(U)\, dU| \geq |\beta_L \int_{U_{\max}}^{U_{\min}} dU| = \beta_L (U_{\max} - U_{\min})$. Hence we obtain $|\Delta S^{\mathrm{int}}_{\mathcal{C}_2}| / \Delta S^{\mathrm{int}}_{\mathcal{C}_1} \geq \beta_L / \beta_H$, or $\eta \leq \eta_e$.

where $\langle \mathbf{b}_i, \mathbf{b}_j \rangle \equiv \mathbf{b}_i^T \mathbf{G} \mathbf{b}_j$ $(i, j = 0, 1)$ is an inner product of the vectors $\mathbf{b}_i$ and $\mathbf{b}_j$ with a metric given by $\mathbf{G} = \langle \mathbf{F}(\mathbf{x})\mathbf{F}(\mathbf{x})^T \rangle - \langle \mathbf{F}(\mathbf{x}) \rangle \langle \mathbf{F}(\mathbf{x}) \rangle^T$. Note that $\langle \mathbf{b}_0, \mathbf{b}_0 \rangle$ is equivalent to Eq. 15. Likewise, the small change in $\boldsymbol{\eta}$ is related to the change in $\boldsymbol{\theta}$ by $d\boldsymbol{\theta} = \mathbf{J}^{-1} d\boldsymbol{\eta}$. Since the m-geodesic processes B→C and D→A are characterised by $d\boldsymbol{\eta} = (0, dX)^T$, the small change in θ-coordinates is given as

$$d\boldsymbol{\theta} = \begin{bmatrix} -\langle \mathbf{b}_0, \mathbf{b}_1 \rangle \\ \langle \mathbf{b}_0, \mathbf{b}_0 \rangle \end{bmatrix} |\mathbf{J}|^{-1} dX, \tag{24}$$

Conversely, the internal mechanism needs to change the internal and stimulus component according to the above gradient in order to accomplish the most efficient cycle. Thus if the internal mechanism can not access the stimulus component α, the ideal information-theoretic cycle is not realised. Further, if $\langle \mathbf{b}_0, \mathbf{b}_1 \rangle = 0$, the internal component β is not allowed to change, which however means that the entire process does not form a cycle. Therefore we impose $\langle \mathbf{b}_0, \mathbf{b}_1 \rangle \neq 0$. This equation indicates that the modulation by an internal mechanism is achieved through the activity features shared by the two components. Accordingly, this condition is violated if neurons participate in the stimulus-related activity and neurons subject to the internal modulation do not overlap (namely if neurons that appear in the features corresponding to non-zero elements of $\mathbf{b}_0$ are separable from those of $\mathbf{b}_1$). In general, in order to make a change of the internal component β influence the stimulus-related activity X, therefore controls stimulus information, one requires $\langle \mathbf{b}_0, \mathbf{b}_1 \rangle \neq 0$ because $dX = -\langle \mathbf{b}_1, \mathbf{b}_0 \rangle d\beta + \langle \mathbf{b}_1, \mathbf{b}_1 \rangle d\alpha$ from $d\boldsymbol{\eta} = \mathbf{J} d\boldsymbol{\theta}$.

5 Discussion

In this study, we provided hypothetical neural dynamics that efficiently encodes stimulus information with the aid of delayed gain-modulation by an internal mechanism, and demonstrated that the dynamics forms an information-theoretic cycle that acts similarly to a heat engine. This view provided us to quantify the efficiency of the gain-modulation in retaining the stimulus information. The ideal information-theoretic cycle introduced here bounded the entropic efficiency.

As an extension of a logistic activation function of a single neuron to multinomial outputs, the maximum entropy model explains probabilities of activity patterns by a softmax function of the features, therefore allows nonlinear interaction of the inputs (here β and α) in producing the stimulus-related activity X (Fig. 1). This interaction was caused by shared activity features in $\mathbf{b}_1$ and $\mathbf{b}_0$. The gain modulation more effectively changes the stimulus-related activity if the features of the stimulus-related and internal activities resemble (i.e., $\langle \mathbf{b}_1, \mathbf{b}_0 \rangle$ is close to 1), which may have implications in similarity between evoked and spontaneous activities [14] that can be acquired during development [3].

The model's statistical structure common to thermodynamics (the Legendre transformation; see Appendix) allowed us to construct the first law for neural dynamics (Eq. 12), the equation of state (Eq. 13), fluctuation-dissipation relation (Eq. 15), and neural dynamics similar to a thermodynamic cycle (Figs. 2 and 3) although we emphasised the differences from conventional thermodynamics in terms of the controllable quantities. The dynamics forms a cycle if the gain modulation is applied after the initial increase of the stimulus-related activity. This scenario is expected when the stimulus response is modulated by a feedback mechanism of recurrent networks [31, 41, 42], and is associated with short-term memory of the stimulus [31, 32]. Consistently with the idea of efficient stimulus-encoding by a cycle, effect of attentional modulation on neural response typically appears several hundred milliseconds after stimulus onset (later than the onset of the stimulus response) [9, 17, 19, 22, 28, 34] although the temporal profile can be altered by task design [9, 17].

To apply the theory to empirical data, the internal and stimulus feature need to be specified. Since even spontaneous neural activity is known to exhibit ongoing dynamics [14], estimation of these features is nontrivial. The optimal sequential Bayesian algorithms have been proposed to smoothly estimate the parameters of the neural population model when they vary in time [35–37]. These approaches can be used to select dominant features of spontaneous and evoked activities, and then to estimate the time-varying internal and stimulus-related components. By including multiple stimulus features in the model, the theory is expected to make quantitative predictions on competitive mechanisms of selective attention [17, 21, 22, 27]. The conservation law of entropy imposes competition among the stimuli given a limited entropic resource generated by the internal mechanism.

In sum, a neural population that works as an information-theoretic engine produces entropy ascribed to stimulus-related activity out of entropy supplied by an internal mechanism. This process is expected to appear during stimulus response of neurons subject to feedback gain-modulation. It is thus hoped that quantitative assessment of the neural dynamics as an information-thertoeic cycle contributes to understanding neural computation performed internally in an organism.

Acknowledgments

The author thanks C Donner, D Hirashima, S Koyama, and S Amari for critically reading the manuscript.

Appendix: Free energies of neurons

In this appendix, we introduce free energies of a neural population. Let us first discuss the relation of state variables and free energies that appear in our analysis of the neural population with those found in conventional thermodynamics. Assume that the small change in internal activity of neurons has the following linear relations to entropy S, expected feature X, and the number of neurons N:

$$dU = TdS + fdX + \mu dN. \tag{25}$$

Eq. 25 is the first law of thermodynamics, and the parameters are temperature T, force f, and chemical potential μ. The first law describes the internal activity as a function of (S, X, N). In thermodynamics, the Helmholtz free energy $F = U - TS$, Gibbs free energy $G = F - fX$, or enthalpy $H = U - fX$ are introduced to change the independent variables to (T, X, N), (T, f, N), and (S, f, N), respectively. These free energies are useful to analyse isothermal or other processes in which only one of the independent variables is changed. For example, the Helmholtz free energy can be used to compute the work done by force f under the isothermal condition. However, the concepts of the force and work may not be directly relevant to information-theoretic analysis of a neural population. Here we introduce the free energies that are more consistent with the framework based on entropy changes.

The first law is alternatively written as

$$dS = \beta dU - \alpha dX - \gamma dN, \tag{26}$$

Here we used $\beta = 1/T$, $\alpha = f/T$, and $\gamma = \mu/T$. This first law describes a small entropy change as a function of (U, X, N). The parameters are defined as

$$\beta(U, X, N) = \left(\frac{\partial S}{\partial U}\right)_{X,N}, \tag{27}$$

$$\alpha(U, X, N) = -\left(\frac{\partial S}{\partial X}\right)_{N,U}, \tag{28}$$

$$\gamma(U, X, N) = -\left(\frac{\partial S}{\partial N}\right)_{U,X}. \tag{29}$$

We change the independent variable U to β. For this goal, here we define the *scaled* Helmholtz free energy $\mathcal{F}$ as

$$\mathcal{F} = S - \beta U. \tag{30}$$

Note that $\mathcal{F} = -\beta F$. It is a function that changes the independent variables from (S, X, N) to (β, X, N). This can be confirmed from the total derivative of $\mathcal{F}$: $d\mathcal{F} = dS - d(\beta U) = -Ud\beta - \alpha dX - \gamma dN$. From this equation, we have

$$U(\beta, X, N) = -\left(\frac{\partial \mathcal{F}}{\partial \beta}\right)_{X,N}, \tag{31}$$

$$\alpha(\beta, X, N) = -\left(\frac{\partial \mathcal{F}}{\partial X}\right)_{N,\beta}, \tag{32}$$

$$\gamma(\beta, X, N) = -\left(\frac{\partial \mathcal{F}}{\partial N}\right)_{\beta,X}. \tag{33}$$

The entropy change caused by the stimulus-related activity when X changes from X_1 to X_2 is given by the area under the curve of $\alpha(\beta, X, N)$ in the X-α phase plane. From Eq. 32, if the process satisfies $d\beta = dN = 0$, the entropy change is computed as reduction of the scaled Helmholtz free energy as

$$\Delta S^{\mathrm{ext}} = \int_{X_1}^{X_2} \alpha(\beta, X, N)\, dX = \mathcal{F}(\beta, X_2, N) - \mathcal{F}(\beta, X_1, N). \tag{34}$$

Further change of the independent variables from (β, X, N) to (β, α, N) is done by introducing the *scaled* Gibbs free energy:

$$\mathcal{G} = \mathcal{F} + \alpha X = S - \beta U + \alpha X. \tag{35}$$

Note that $\mathcal{G} = -\beta G$. The independent variables of the Gibbs free energy are (β, α, N) since $d\mathcal{G} = d\mathcal{F} + (d\alpha X + X d\alpha) = -U d\beta + X d\alpha - \gamma dN$. From this equation, we find

$$\left(\frac{\partial \mathcal{G}}{\partial \beta}\right)_{\alpha, N} = -U(\beta, \alpha, N), \tag{36}$$

$$\left(\frac{\partial \mathcal{G}}{\partial \alpha}\right)_{\beta, N} = X(\beta, \alpha, N). \tag{37}$$

Note that the definition of the Gibbs free energy by Eq. 35 is obtained from Eq. 6 if we identify $\mathcal{G} = \psi$. Accordingly, Eqs. 36 and 37 coincide with Eqs. 7 and 8.

References

[1] L. F. Abbott, J. A. Varela, K. Sen, and S. B. Nelson. Synaptic depression and cortical gain control. *Science*, 275(5297):220–224, jan 1997.

[2] S.-i. Amari and H. Nagaoka. *Methods of Information Geometry*. The American Mathematical Society, the United States of America, 2000.

[3] P. Berkes, G. Orbn, M. Lengyel, and J. Fiser. Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, 331(6013):83–87, jan 2011.

[4] A. N. Burkitt, H. Meffin, and D. B. Grayden. Study of neuronal gain in a conductance-based leaky integrate-and-fire neuron model with balanced excitatory and inhibitory synaptic input. *Biol Cybern*, 89(2):119–125, aug 2003.

[5] M. Carandini and D. J. Heeger. Normalization as a canonical neural computation. *Nat Rev Neurosci*, 13(1):51–62, jan 2012.

[6] S. Carnot. *Rflexions sur la puissance motrice du feu et sur les machines propres dvelopper cette puissance*. Bachelier, Paris, 1824.

[7] F. S. Chance, L. F. Abbott, and A. D. Reyes. Gain modulation from background synaptic input. *Neuron*, 35(4):773–782, aug 2002.

[8] B. Doiron, A. Longtin, N. Berman, and L. Maler. Subtractive and divisive inhibition: effect of voltage-dependent inhibitory conductances and noise. *Neural Comput*, 13(1):227–248, jan 2001.

[9] G. M. Ghose and J. H. R. Maunsell. Attentional modulation in visual cortex depends on task timing. *Nature*, 419(6907):616–620, oct 2002.

[10] E. Granot-Atedgi, G. Tkaik, R. Segev, and E. Schneidman. Stimulus-dependent maximum entropy models of neural population codes. *PLoS Comput Biol*, 9(3):e1002922, mar 2013.

[11] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106(4):620–630, may 1957.

[12] R. E. Kass, R. C. Kelly, and W.-L. Loh. Assessment of synchrony in multiple neural spike trains using loglinear point process models. *Ann Appl Stat*, 5(2B):1262–1292, jun 2011.

[13] R. C. Kelly and R. E. Kass. A framework for evaluating pairwise and multiway synchrony among stimulus-driven neurons. *Neural Comput*, 24(8):2007–2032, aug 2012.

[14] T. Kenet, D. Bibitchkov, M. Tsodyks, A. Grinvald, and A. Arieli. Spontaneously emerging cortical representations of visual attributes. *Nature*, 425(6961):954–956, oct 2003.

[15] S. B. Laughlin. The role of sensory adaptation in the retina. *J Exp Biol*, 146:39–62, sep 1989.

[16] B. B. Lee, D. M. Dacey, V. C. Smith, and J. Pokorny. Dynamics of sensitivity regulation in primate outer retina: the horizontal cell network. *J Vis*, 3(7):513–526, sep 2003.

[17] S. J. Luck, L. Chelazzi, S. A. Hillyard, and R. Desimone. Neural mechanisms of spatial selective attention in areas v1, v2, and v4 of macaque visual cortex. *J Neurophysiol*, 77(1):24–42, jan 1997.

[18] J. Martnez-Trujillo and S. Treue. Attentional modulation strength in cortical area mt depends on stimulus contrast. *Neuron*, 35(2):365–370, jul 2002.

[19] C. J. McAdams and J. H. Maunsell. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4. *J Neurosci*, 19(1):431–441, 1999.

[20] S. J. Mitchell and R. A. Silver. Shunting inhibition modulates neuronal gain during synaptic excitation. *Neuron*, 38(3):433–445, may 2003.

[21] J. Moran and R. Desimone. Selective attention gates visual processing in the extrastriate cortex. *Science*, 229(4715):782–784, aug 1985.

[22] B. C. Motter. Focal attention produces spatially selective processing in visual cortical areas v1, v2, and v4 in the presence of competing stimuli. *J Neurophysiol*, 70(3):909–919, sep 1993.

[23] H. Nasser, O. Marre, and B. Cessac. Spatio-temporal spike train analysis for large scale networks using the maximum entropy principle and monte carlo method. *J. Stat. Mech.*, 2013(03):P03006, mar 2013.

[24] I. E. Ohiorhenuan, F. Mechler, K. P. Purpura, A. M. Schmid, Q. Hu, and J. D. Victor. Sparse coding and high-order correlations in fine-scale cortical networks. *Nature*, 466(7306):617–621, jul 2010.

[25] I. Ohzawa, G. Sclar, and R. D. Freeman. Contrast gain control in the cat's visual system. *J Neurophysiol*, 54(3):651–667, sep 1985.

[26] S. A. Prescott and Y. De Koninck. Gain control of firing rate by shunting inhibition: roles of synaptic noise and dendritic saturation. *Proc Natl Acad Sci U S A*, 100(4):2076–2081, feb 2003.

[27] J. H. Reynolds, L. Chelazzi, and R. Desimone. Competitive mechanisms subserve attention in macaque areas v2 and v4. *J Neurosci*, 19(5):1736–1753, mar 1999.

[28] J. H. Reynolds, T. Pasternak, and R. Desimone. Attention increases sensitivity of v4 neurons. *Neuron*, 26(3):703–714, jun 2000.

[29] J. S. Rothman, L. Cathala, V. Steuber, and R. A. Silver. Synaptic depression enables neuronal gain control. *Nature*, 457(7232):1015–1018, feb 2009.

[30] B. Sakmann and O. D. Creutzfeldt. Scotopic and mesopic light adaptation in the cat's retina. *Pflugers Arch*, 313(2):168–185, jan 1969.

[31] E. Salinas and L. F. Abbott. A model of multiplicative neural responses in parietal cortex. *Proc Natl Acad Sci U S A*, 93(21):11956–11961, oct 1996.

[32] E. Salinas and T. J. Sejnowski. Gain modulation in the central nervous system: where behavior, neurophysiology, and computation meet. *Neuroscientist*, 7(5):430–440, oct 2001.

[33] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, apr 2006.

[34] E. Seidemann and W. T. Newsome. Effect of spatial attention on the responses of area mt neurons. *J Neurophysiol*, 81(4):1783–1794, apr 1999.

[35] H. Shimazaki. Single-trial estimation of stimulus and spike-history effects on time-varying ensemble spiking activity of multiple neurons: a simulation study. *J. Phys.: Conf. Ser.*, 473:012009, dec 2013.

[36] H. Shimazaki, S.-i. Amari, E. N. Brown, and S. Grün. State-space analysis on time-varying correlations in parallel spike sequences. *ICASSP 2009. IEEE International Conference on*, apr 2009.

[37] H. Shimazaki, S.-i. Amari, E. N. Brown, and S. Grün. State-space analysis of time-varying higher-order spike correlation for multiple neural spike train data. *PLoS Comput Biol*, 8(3):e1002385, mar 2012.

[38] H. Shimazaki, K. Sadeghi, T. Ishikawa, Y. Ikegaya, and T. Toyoizumi. Simultaneous silence organizes structured higher-order interactions in neural populations. *Sci Rep*, 5:9821, apr 2015.

[39] J. Shlens, G. D. Field, J. L. Gauthier, M. I. Grivich, D. Petrusca, A. Sher, A. M. Litke, and E. J. Chichilnisky. The structure of multi-neuron firing patterns in primate retina. *J Neurosci*, 26(32):8254–8266, aug 2006.

[40] R. A. Silver. Neuronal arithmetic. *Nat Rev Neurosci*, 11(7):474–489, jul 2010.

[41] M. W. Spratling and M. H. Johnson. A feedback model of visual attention. *J Cogn Neurosci*, 16(2):219–237, mar 2004.

[42] C. Sutherland, B. Doiron, and A. Longtin. Feedback-induced gain control in stochastic spiking networks. *Biol Cybern*, 100(6):475–489, jun 2009.

[43] A. Tang, D. Jackson, J. Hobbs, W. Chen, J. L. Smith, H. Patel, A. Prieto, D. Petrusca, M. I. Grivich, A. Sher, P. Hottowy, W. Dabrowski, A. M. Litke, and J. M. Beggs. A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *J Neurosci*, 28(2):505–518, jan 2008.

[44] G. Tkačik, O. Marre, D. Amodei, E. Schneidman, W. Bialek, and M. J. Berry. Searching for collective behavior in a large network of sensory neurons. *PLoS Comput Biol*, 10(1):e1003408, jan 2014.

[45] G. Tkačik, T. Mora, O. Marre, D. Amodei, S. E. Palmer, M. J. Berry, and W. Bialek. Thermodynamics and signatures of criticality in a network of neurons. *Proc Natl Acad Sci U S A*, 112(37):11508–11513, sep 2015.

[46] S. Yu, D. Huang, W. Singer, and D. Nikolic. A small world of neuronal synchrony. *Cereb Cortex*, 18(12):2891–2901, dec 2008.