How to find decision makers in neural circuits?

Alexei A. Koulakov[1], Dmitry A. Rinberg[2], and Dmitry N. Tsigankov[1]

[1] *Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 11724, USA*
[2] *Monell Chemical Senses Center, Philadelphia, PA, 19104, USA*

Neural circuits often face the problem of classifying stimuli into discrete groups and making decisions based on such classifications. Neurons of these circuits can be distinguished according to their correlations with different features of stimulus or response, which allows defining sensory or motor neuronal types. In this study we define the third class of neurons, which is responsible for making decisions. We suggest two descriptions for contribution of units to decision making: first, as a spatial derivative of correlations between neural activity and the decision; second, as an impact of variability in a given neuron on the response. These two definitions are shown to be equivalent, when they can be compared. We also suggest an experimental strategy for determining contributions to decision making, which uses electric stimulation with time- varying random current.

I. INTRODUCTION

Nervous system is continuously confronted by megabytes of information, representing light, sound, smell, etc. This information is compiled by the brain into a set of decisions, representing behaviors of living organisms. The mechanisms involved in this reduction have been under investigation for many years (Glimcher, 2003; Romo and Salinas, 2003). In this study we address a question complementary to the issue of decision making (DM) mechanisms. We define neuronal units involved in making perceptual decisions. For this purpose we determine DM activity in surrogate networks, defined mathematically, in which a complete control is present over stimuli, mechanisms, and responses. Such decision making analysis (DMA) has practical significance, since once units involved in making particular decision are located, further efforts could be concentrated on uncovering the underlying mechanisms.

In this study DM task is defined as evaluation of a function in the multidimensional stimulus space (Figure 1A). This function has a discrete set of values, representing the repertoire of responses available to the organism. The decisions may, of course, be stochastic, to reflect the uncertainty, pertinent to behavior. This definition is suitable for experiments where subjects perform poly-alternative forced-choice tasks, such as saccadic response to the direction of stimulus motion (Shadlen and Newsome, 2001).

Let us consider motion-discrimination task in more detail. Figure 1B lists some visual areas, which are involved in this task. The areas are arranged along a rough sensory-motor axis, so that the areas on the left are more "sensory", while those on the right are more "motor". This implies that the responses in these areas are more correlated with stimulus or response respectively. Where on this sensory-motor axis one should position the DM elements? One could argue that the elements most correlated with the decision itself are the decision makers, following the analogy with the definition of sensory and motor elements. It is, however, difficult, if not impossible, to distinguish such definition from the definition of purely motor units (Shadlen and Newsome, 2001). The latter relay the results of decision making process, without involvement in the formation of the decision. An alternative approach is therefore needed to define the DM units.

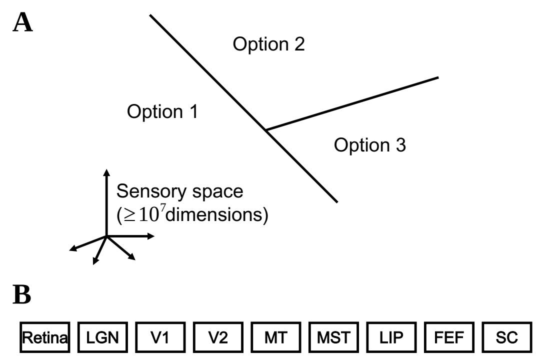

FIG. 1 **A,** Definition of decision making task. Nervous system evaluates a function, whose values represent discrete decisions, in the many-dimensional sensory space. **B,** Some of the visual areas involved in motion-discrimination task. The areas on the left are more sensory (response is correlated with the sensory input), while those of the right are more motor (correlated with the response).

The DM components may be surmised to be located on the interface between sensory and motor areas. More precisely, the *first* element in the sensory-motor chain, which carries significant correlation with the response, may be identified as the decision maker. In this study we develop this idea into rigorous mathematical formulation and find a special correlation function, which determines contributions of units to DM. This formalism allows us to answer two questions pertaining to the identities of DM units. First, we consider the case when not one but *several* elements are involved in the same decision simultaneously. Our approach allows us to evaluate relative importance of various units in such a distributed DM. Second, we consider the systems with loops in connectivity. For such systems the concept of 'the first element' becomes more arbitrary and one has to proceed more carefully in defining contributions to DM. We succeed in doing so for our surrogate networks and define DM units for recurrent networks in a way, which is consistent with the linear sensory-motor chains, thus satisfying the requirement of the correspondence principle.

This paper is organized as follows. We first analyze simple linear chain models, and networks, such as trees, which have similar properties. We then use this analysis to define decision makers in networks of arbitrary connectivity. Finally, we extend our study to the cases, when electric stimulation can be applied to units, and show that DM components can be identified in a way consistent with our preceeding analyses.

II. LINEAR CHAINS AND THEIR DERIVATIVES

The goal of this section is to formulate quantitative principles by which DM network elements can be identified. We approach this task by analyzing simple cases, which can be solved exactly without the use of computer, and in which the identities of DM elements are clear. These cases allow us to emphasize the properties of DM task we are attempting to describe. We proceed therefore to the analysis of the simplest network capable of making decisions.

A. The 'nematode' network

In this subsection we consider the network, which we call 'nematode', because of its resemblance to simple biological organisms, both in the layout and in the fundamental significance. We first define the model; then show that it can make simple decisions; and, finally, define the positions of decision makers in the network.

Consider a linear chain of units, whose response is characterized by a set of real numbers x_i, where $i = 1...N$ is the position of the element in the chain (Figure 2). Response of each element does not depend on time. This model is therefore static. This assumption is introduced here to simplify the analysis and can be relaxed as described below (section II.C). Each unit performs a simple linear transformation between the unit's input and the output. Thus, for element number i

$$x_i = x_{i-1} + \eta_i \tag{1}$$

Here η_i is noise associated with the element. In this work we assume that noise has zero mean, is individual to each unit, and, therefore, is uncorrelated between units, i.e.

$$\overline{\eta_i} = 0, \quad \overline{\eta_i \eta_j} = \begin{cases} \overline{\eta_i^2}, & i = j \\ 0, & i \neq j \end{cases} \tag{2}$$

We further assume that noise has a Gaussian distribution. The chain of linear elements is thus completely specified by a set of noise variances $\overline{\eta_i^2}$. The model described by (1) and (2) yields the following solution for the response of the last element in the chain

$$x_N = x_0 + \eta_1 + \eta_2 + ... + \eta_{N-1} + \eta_N. \tag{3}$$

Thus, the response of the last element is just a sum of the input into network x_0 and noise contributions from all units, independently on the order of unit in the chain.

The last element in the chain has non-linear response properties. Its response is defined by

$$d = H(x_N), \tag{4}$$

where $H(x)$ is the Heaviside step function, which is equal to one/zero if the argument is positive/negative. It follows then that our 'nematode' network is capable of making decisions based on the values of input variable x_0. This is if we interpret variable d, which is equal either 0 or 1, as the result of DM process, as defined in Figure 1A. The decisions are made stochastically and are dependent upon the instantiations of random variables η_i, which vary from trial to trial.

Our model is completely defined by the set of noise variances, pertinent to each unit $\overline{\eta_i^2}$. Although decisions made by this chain are quite simple, the identities of decision makers are not so easy to find. The distribution of impact to DM along the chain should depend upon the distribution of noise variables $\overline{\eta_i^2}$. Our next goal is to develop a sensible definition of contributions to DM based on the vector of variances $\overline{\eta_i^2}$. Before doing so we describe general input-output properties of the chain.

Since decision made by the network varies from trial to trial, one can define averaged over trials response of the system $\overline{d(x_0)}$. As shown in Figure 2B it has a sigmoid shape, smeared by the total amount of noise in the system. One can, therefore, consider two cases, depending

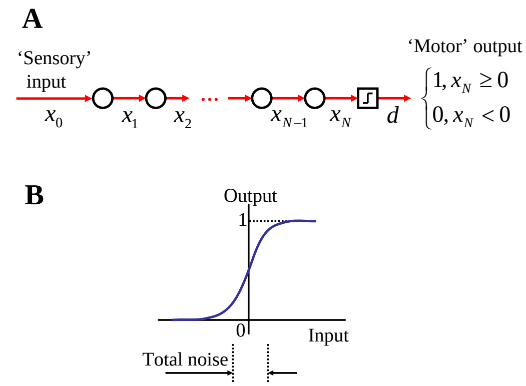

FIG. 2 **A,** a simple 'nematode' network consists of a linear chain of units. All units in the chain, but the last, are linear. The last unit, shown by the square, is non-linear and returns zero or one depending on the sign of the response of the preceding unit. **B,** The average input-output relationship for the 'nematode' is given by the sigmoid function (error function). The spread of the sigmoid is determined by the net noise in the chain.

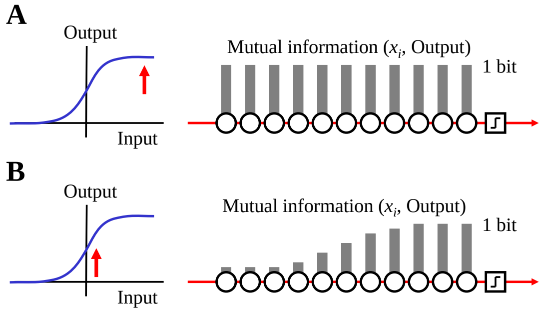

FIG. 3 **A,** If signal-to-noise ratio is high, responses of all units are well correlated with the output, as shown on the right by the mutual information between the response of given unit and the output. **B,** For the case of low signal-to-noise ratio, the output is more correlated with the motor units (right) than with the sensory ones (left).

on whether the signal-to-noise ratio for the chain is large or small. These two regimes are shown in Figure 3A and B respectively.

To analyze responses of units in these two cases we define their correlation with the decision. This correlation is defined for each element in the chain (Figure 3, right). As a measure of correlation we choose mutual information (MI) between response of the $i-$th unit, x_i, and the decision, d. MI has an advantage of being unitless (it is measured in bits) and having clear intuitive properties, as described below. We will also show below in this section that MI has limitations as a measure of DM.

MI describes the information transmission from the $i-$th unit to the output of the system. Since the output can only have values 0 or 1, MI cannot exceed the value of one bit. We now consider two cases, depending on the network's signal-to-noise ratio. If network input $|x_0|$ is large, as in Figure 3A, response of the system is well correlated with the input. Hence, activities of all units are well correlated with both input and output, and $MI(x_i, d) \approx 1$ for all of the units. In the opposite limit, when the signal-to-noise ratio is small, $|x_0|$ is smaller than noise, and the system's response is weakly correlated with the input (Figure 3B). In this case MI as a function of unit's position displays a structure, shown in Figure 3B (right). This structure, as shown below, has a key to the definition of DM components and is qualitatively discussed here. The units, which are close to the exit from the network, show strong correlation with the decision, similarly to the high signal-to-noise ratio case. Their MI is therefore close to 1 bit. On the other hand, more 'sensory' units, in the beginning of the chain are strongly correlated with the input. Since input-output correlation is weak in low signal-to-noise ratio case, the 'sensory' units display virtually *no* relation to the output and $MI(x_i, d) \approx 0$ for such units (Figure 3B, right). Thus, MI, as a function of i displays a transition from 0 to 1 in the low signal-to-noise ratio case.

How could one deduce identities of decision makers from these dependencies (Figure 3A and B)? One could suggest that the elements perfectly correlated with the output of the system, such as exit elements from the chain, are the ones that make the decision. However, such elements may be just the relay or 'motor' units, in which case their contribution to DM is small. Indeed, when we type, our decisions are perfectly correlated with activities of finger muscles; but one could hardly blame our fingers for the content of the typing. Thus, despite their high correlation with the output, exit elements could not be called decision makers. Input elements, having no correlation with the decision, are responsible for DM in even lesser degree. We thus need to analyze the dependence of MI on position in more detail and suggest another scheme for defining DM units.

Our discarding of motor units as decision makers can be further extended onto the entire high signal-to-noise ratio case (Figure 3A). We suggest that the deterministic regime is not descriptive from the point of view of DM analysis. First, in this regime all units become indistinguishable from motor. The latter are not decision makers, as suggested above. Second, the dependence shown in Figure 3A (right) does not reveal the contributions of individual units to the decision. Since all units have the same correlation, it is hard, if not impossible, to differentiate them and assign different contributions. Third, the responses of units in this case are deterministically related to the input. Hence, units act as relays, passively transmitting information along the chain. It can be argued that the external environment, providing the input variable x_0, acts as the decision maker. We conclude that to find decision making activity one has to concentrate on the low signal-to-noise ratio case.

We show below that the identities of decision making units can be deduced from the shape of transition in Figure 3B (right). To this end we analyze a set of examples of networks with various distributions of noise $\overline{\eta_i^2}$. We start from the simplest example of a single noisy unit.

1. Example 1: 'Noisy' neuron.

Consider a chain in which noise is absent from all units but one, whose order number in the chain is n (Figure 4). Since, according to our previous discussion, we need to consider the low signal-to-noise ratio case, we will assume that

$$x_0 = 0, \tag{5}$$

i.e. network receives no input. Making the decision in this case is still possible, based on the values of noise inside the network. Since noise is only present in one neuron, from (3) we conclude that

$$x_N = \eta_n. \tag{6}$$

The decision made by the network is

$$d = H(\eta_n). \tag{7}$$

Thus, decision is causally linked to the processes controlling unit number n, which leads us to conclusion that this neuron is the decision maker.

Paradoxically, the noisiest unit in this simple formulation makes the largest impact. All noiseless elements, even nonlinear, are deterministic, and work as simple relays which transmit information from the previous node to the next one. The output of the circuit is linked to the processes controlling noise in neuron number n, rather that in any other neuron in the network.

One would be tempted to conclude that the non-linear element is actually the decision maker in this case. We deduce that the non-linear element does not have a causal effect on output from the circuit; therefore its role is just to relay response from neuron n to the output. In this respect the non-linear element is not different from other noiseless elements.

To link this example to our previous discussion (Figure 3B) we plot MI as a function of position in the chain in Figure 4 (top). As we discussed, MI is high for exit ('motor') units and low for input ('sensory') elements. Figure 4 also shows the derivative of MI with respect to position in the chain. It is clear that this derivative represents the decision making element. Thus, we conclude that not correlation with the decision but the *rate of change* of the latter along the network is the indicator of DM.

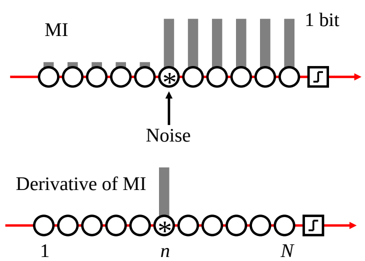

FIG. 4 The example of 'noisy' neuron (marked by asterisk). Top panel, mutual information between given unit and the decision. Bottom panel, derivative of mutual information. The derivative represents the decision making unit in this case.

2. Example 2: Uniformly distributed noise.

Our next example shows that the conclusion about derivative of MI is basically correct, but has to be slightly amended to be numerically precise. Consider the chain in which all elements are noisy and the variance of noise is the same for each element. In this case

$$x_N = \eta_1 + \eta_2 + ... + \eta_{N-1} + \eta_N \tag{8}$$

i.e. all units contribute to decision *equally*. This is because Eq. (8) does not distinguish the order in which contributions from the units are added, and all contributions are of equal strength on average. Can this conclusion be confirmed by the derivative of MI?

Figure 5A shows MI as a function of position in the chain for this case. This dependence is obtained in Appendix A. It increases smoothly from 0 to 1 resulting in a non-zero derivative at all units. This is consistent with (8) and the notion that all units participate in the decision. However, (8) suggests that all units participate in decision *equally*. The derivative of MI turns out to be slightly non-uniform, as seen in Figure 5A. This can be corrected if not MI itself but a non-linear function of MI, denoted $F(MI)$, is considered. This non-linear function is calculated in Appendix A and is shown in Figure 5B. The new correlator $F(MI)$ has the same basic properties as the MI. It rises from 0 to 1 monotonously when passing through the array (Figure 5C). But, in addition, its derivative turns out to be *uniform*, as shown in Figure 5C (bottom). This is consistent with equal participation of all units in DM in the uniformly distributed noise case and Eq. (8). Thus, we conclude that for this case the contributions to DM are given by the rate of increase of $F(MI)$ when moving through the array

$$DM_i = F(MI_i) - F(MI_{i-1}). \tag{9}$$

Here i is the index along the chain. Eq. (9) is the main result of this paper. It represents our definition of contributions to DM for networks of simple connectivity, such as chains.

Three points should be made about the definition (9). First, it reproduces the result obtained in the previous example of 'noisy' neuron. Indeed, the mutual information rises from 0 to 1 on the 'noisy' neuron in Figure 4. But $F(MI)$ coincides with MI at these values, as follows from its plot in Figure 5B. Thus, the derivative of $F(MI)$ is also given by a single spike at the position of 'noisy' neuron, as in Figure 4 (bottom). Second, Eq. (9) implies that, from point of view of DM, not mutual information, but another correlator, given by $F(MI)$, is more relevant. Function F deviates from linear function only slightly (Figure 5B), and for practical purposes the distinction between the MI and $F(MI)$ could be ignored. However, we retain it throughout the manuscript to ensure mathematical rigor. Third, when deducing (9) we did not postulate that contributions to DM are proportional to the variance of noise. Instead, we suggested that Eq. (8) implies that all units contribute equally, independently on the order in the chain. This simple qualitative statement is powerful enough to constrain our quantitative reasoning and lead to a measure of DM in form of function $F(MI)$ and definition (9). We do not know yet if the derivative of $F(MI)$ is proportional to the variance of noise, square root of this variance, or any other characteristic of noise in each element. All of these parameters give the same results in the uniform noise case. We need to have a difference between units to measure relative strength of their contributions. This is achieved by the next example.

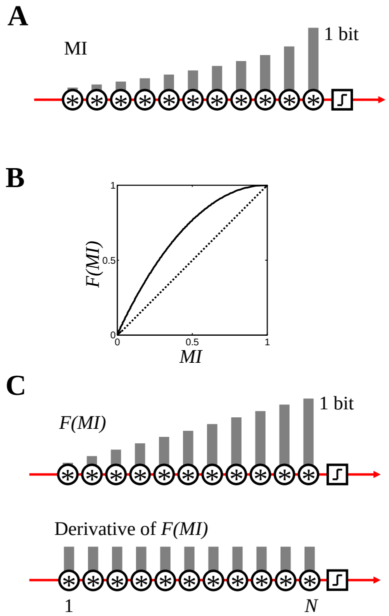

FIG. 5 The example with uniformly distributed noise. **A**, mutual information between response of given unit and the decision. The dependence has a non-uniform increase, suggesting that mutual information is not a good measure of decision making. **B**, if one applies a non-linear function (solid curve) to the mutual information in **A**, one obtains a uniformly increasing correlator in **C**. This non-linear function, called $F(MI)$, is close to linear, shown by the dotted line. **C**, the new correlator $F(MI)$ (top panel) has a uniform derivative (bottom panel). Thus, derivative of $F(MI)$ is a sensible measure of decision making in the case of uniform noise.

3. Example 3: 'Loud' neuron.

In this example the variances of noise on all neurons are the same, similarly to the previous case. However, here we amend the network definition given by (1). We do so for only one neuron. We assume that the link between units 5 and 6 is characterized by a very large strength $K >> 1$. Thus, for neuron number 6 (Figure 6) instead of (1) we have

$$x_6 = Kx_5 + \eta_6 \tag{10}$$

Therefore this example is the same as the previous, except that the single network connection is changed. What are the DM units in this case?

The network's output is given by

$$x_N = K(\eta_1 + \eta_2 + ... + \eta_5) + \eta_6 + ... + \eta_{11} \tag{11}$$

Thus, units 1 through 5 contribute equally to decision. In addition, their contributions are multiplied by a large factor K. Units 6 through 11 also contribute equally, but their contribution is much smaller than that of the former group. We conclude that units 1 through 5 are much stronger decision makers than units 6 through 11. This conclusion is supported by the derivative of $F(MI)$, as shown in Figure 6 (bottom).

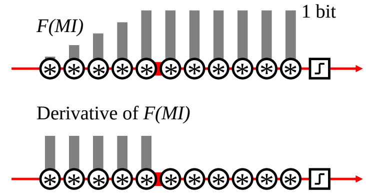

FIG. 6 The 'loud' neuron example. The link between units 5 an 6 is strengthened. Compare to Figure 5A.

Thus, changing one link in the chain produces large effect on the distribution of DM. The units downstream from the link contribute less to decisions, while the units upstream contribute a lot. What is the measure of decision making, which could differentiate these two types of units?

Calculations in Appendix A show that derivative of $F(MI)$ is proportional to K^2 for units 1 through 5. This is easy to understand qualitatively, since MI increases along the chain even for negative ($K < 0$) links. This is *not* possible if contribution from units 1 to 5 are multiplied by K for example. Thus, an even power of K is required, which is shown in Appendix A to be K^2.

4. Alternative definition of DM.

So far we have used definition (9), which is quite complex, since it involves calculation of a nonlinear function $F(MI)$. Is it possible to reproduce the results derived above in a simpler way? It turns out that the role of given unit in DM is proportional to its contribution to the variability of the output $\overline{x_N^2}$. This leads us to an alternative to (9) definition of DM.

Let us introduce the new definition using the examples, considered above. From (11) in the 'loud' neuron case we derive

$$\overline{x_N^2} = K^2(\overline{\eta_1^2} + ... + \overline{\eta_5^2}) + \overline{\eta_6^2} + ... + \overline{\eta_{11}^2}. \tag{12}$$

We could conjecture that the contributions to DM from different units are weighted proportionally to the corresponding summands in (12). Indeed, if we assume

$$DM_{1..5} = \overline{\eta_{1..5}^2}K^2$$

$$DM_{6..11} = \overline{\eta_{6..11}^2}, \tag{13}$$

by choosing appropriate values of variance of noise and gain, we can reproduce the results of all three of our previous examples. Thus, in the case of 'noisy' neuron the variance of noise is only present in one unit, rendering this unit decision maker, according to (13). In the case of uniform noise, when $K = 1$ and all $\overline{\eta_{1..11}^2}$ are the same, (13) gives uniform contributions to DM. In the case of 'loud' neuron, (13) gives the correct factor K^2 describing the advantage of upstream neurons. Thus, the contributions to DM are proportional to the variance of noise on given element, multiplied by the square of the gain from this element to the output. We can rewrite (13) in a more compact form to emphasize this latter statement

$$DM_i = \overline{\eta_i^2}\frac{d\overline{x_N^2}}{d\overline{\eta_i^2}}. \quad (14)$$

One could verify (14), by applying it to (12) and obtaining relationships (13). This justifies (14) in the three examples considered above.

Eq. (14) also applies to linear chains in general. In Appendix A we derive (14) from previous definition (9) for arbitrary distribution of connection strengths and noise. Thus, (14) can be considered an alternative definition to (9). The equivalence between (9) and (14) is demonstrated graphically in Figure 7.

Why should one consider an alternative definition? This is because (9) cannot be applied to networks of arbitrary connectivity, such as circuits containing loops. Definition (14) however applies to all topologies, including the linear chain examples, considered here.

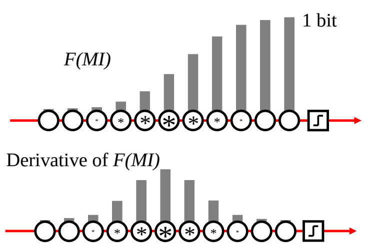

FIG. 7 Equivalence of two definitions. The top panel shows distribution of noise variance (asterisk diameter) and of $F(MI)$ (bars). The bottom panel displays the derivative of $F(MI)$, defined by (9). The derivative is numerically the same as the variance of noise. Both can be used as measures of decision making.

B. Conclusions from 'nematode' study

Let us review our findings. First, we arrived to the definition of DM activity using the information-theoretical approach (9). According to this definition, DM is the rate of change of correlation with decision along the chain. In other words, the *first* element or elements, which correlate with the decision, are the decision makers. This approach has its pros and contras. Indeed, the viewpoint expressed by (9) has a potential to be transferred to other systems, which contain non-linear elements. Eq. (9) has an information-theoretical origin; hence its applicability may be broader than our simple system. Another advantage of (9) is that it relies on the characteristics measurable in single-electrode recording experiments, such as response of single unit and its correlation with behavioral decision. Thus, (9) could be used experimentally. The disadvantage of the information-theoretical approach is that it is not clear how to apply it to the systems with loops, as we have mentioned above. Since biological networks almost always contain loops this significantly limits the applicability of information-theoretical formula (9).

Our second step was to derive an alternative definition (14). The latter is *equivalent* to the former definition (9) for linear-chain ('nematode') example, as we have demonstrated on simple examples and have shown more rigorously in Appendix A. The alternative definition (14) can be understood on the basis of the following two observations. First, the example of 'noisy' neuron shows that the variability is the source of decisions. Thus,

<u>Conclusion 1:</u> Under fixed other conditions, an increase in variability and noise in a single unit leads to a larger contribution to DM from this unit.

$$DM_i \sim \overline{\eta_i^2} \quad (15)$$

Second, the example of 'loud' neuron shows that not only variability and noise are important but also how much of this variability reaches the motor units. DM is hence a property of network connectivity too. Thus, we arrive to the next rule

<u>Conclusion 2:</u> The stronger is the pathway from given unit to the motor output, the larger is the contribution of this unit to DM.

$$DM_i \sim \frac{d\overline{x_N^2}}{d\overline{\eta_i^2}} \quad (16)$$

These two rules are combined into the definition (14). Although (14) and (16) assume that the output element is unique, this requirement will be removed below, when we consider arbitrary topology networks.

What are the features of (14)? It could be used for an arbitrary topology network, since it does not contain derivative along the chain, as (9) does. Definition (14) can also be used operationally to measure the contribution of each neuron to the decision experimentally. To do that one needs to vary noise at the given unit and measure the variability of the responses. The details are discussed in section IV below.

A special note should be made about normalization in (14). Throughout this work we adopt the convention that DM contributions are evaluated for all units and then normalized proportionally to (14), so that the total sum of all contributions is equal to one (or 100%). We will assume this to hold below without explicitly mentioning. Finally, we give another definition of DM contributions, which could be useful when noise in the system is the same for all units. In this case the only difference between units is due to difference in their position in the network. We therefore call such quantity *topological* DM.

$$TDM_i = \frac{\partial \sigma^2(x_N)}{\partial \overline{\eta_i^2}} \quad (17)$$

As seen from e.g. (12) it does not depend on the levels of noise, and can be obtained from (14) by assuming that $\overline{\eta_i^2} = 1$ for all units. It therefore describes how strongly each elements of the circuit affects the output. This quantity is sometimes helpful in describing the network's topology.

Lastly, we discuss the notion of noise and variability in our approach. Is this really noise, which leads networks to decisions? Not necessarily. Imagine that we have studied a chain-like network (Figure 8A) and performed the DM analysis, described above. We found that the network contains two decision makers, which are equally important. A more thorough investigation may suggest that these units are inputs from external network, which in effect is responsible for DM. For example, these hidden pathways may be inputs from other sensory modalities or regulatory inputs of other type. Thus, DMA may help identify entry points from other, less studied, parts of the network.

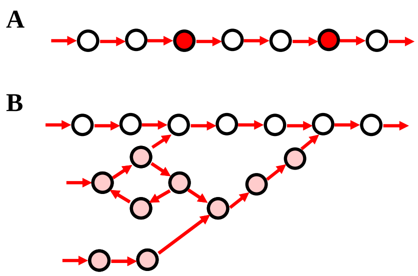

FIG. 8 Hidden pathway. The intensity of red shows contribution to decision making for each unit. **A,** analysis for an incomplete connectivity reveals two decision makers. **B,** a more thorough study may show that this results from other inputs to the network.

C. Trees

Our studies indicate that information-theoretical analysis [definition (9)] can be further extended to tree-like topologies (Figure 9). The latter are defined as connectivities with no loops in them. In addition, in our case, each network element is allowed to have only one outgoing link (Figure 9). We define a column-vector $\vec{f}$ such that $f_i = F(MI_i)$. Then (9) is equivalent to

$$\overrightarrow{DM} = (\hat{I} - \hat{S})\vec{f}. \tag{18}$$

Here $\hat{S}$ is the structure matrix defined as follows. An element S_{ij} of the structure matrix is equal to 1 if there is a connection from unit number i to j. Matrix $\hat{I}-\hat{S}$ thus implements evaluating differences between connected elements in (9). Structure matrix is related to connectivity matrix, containing network's weights through $S_{ij} = |sign(C_{ij})|$. Connectivity matrices for some networks are shown in Figures 9 and 10.

Information-theoretical approach can be even further extended on the cases, when signals propagate along the network in time, therefore resulting in delays between signal and response. In this case by $\vec{f}$ one should understand a sum of correlations over all times preceding the decision. This compensates for the presence of delays. So far there is no understanding if Eq. (18) [or (9)] can be used for topologies other than trees. Definition (14), however, can be used with networks of arbitrary connectivity. This is the topic of the next section.

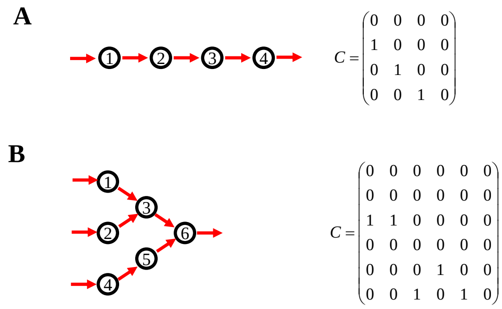

FIG. 9 Mutual information approach can be extended to connectivities other than linear chains (**A**). Thus, decision makers on trees (**B**) can also be found. Arbitrary network can be specified by connectivity matrices, which are provided for illustration purposes. The non-zero entries in a connectivity matrix indicate a connection between two elements numbered on the left. An entry value describes the strength of connection and does not have to be unitary or positive.

III. DYNAMIC MODELS

All previous examples, except the one mentioned at the end of the last session, were static, i.e. variables did not depend on time. The deficiency of this approach is that it is not clear how to treat networks with loops. To apply our analysis to the cases with loops, and, in general, to networks with *arbitrary* connectivity (Figure 10), we consider time-dependent models here. This allows us to observe propagation of noise around the loop explicitly and to make accurate conclusions about contributions to DM.

We limit ourselves to linear dynamical systems, where the single nonlinear element is the last one, transforming an analog system output to a binary response. As the first step we consider temporal dynamics in the discrete-time approximation, which contains all essential features of our approach. Later in the section we extend discrete model to the continuous-time case and show their equivalence.

A. Discrete-time model

In this section we consider a system of N elements, whose activity at each instant is described by an N-dimensional column-vector $\vec{x}(t)$. Time has discrete values separated by an interval τ. Therefore this model is called the discrete-time model. The values of activity at

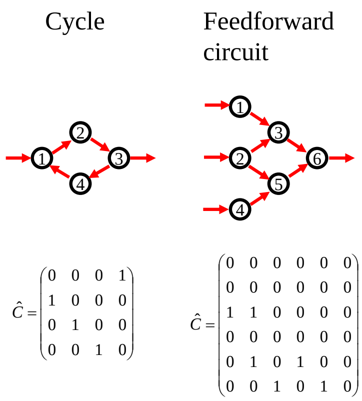

FIG. 10 The network topologies considered by the dynamic models. Arbitrary connectivities, such as cycles (left) or feed-forward networks (right) can be considered.

two neighboring time-slices are related by the connection matrix $\hat{C}$

$$\vec{x}(t+\tau) = \hat{C}\vec{x}(t) + \vec{\eta}(t) + \vec{s}(t) \tag{19}$$

Here $\vec{\eta}(t)$ is the vector describing noise added to activity vector on each time-slice. The variable $\vec{s}(t)$ describes sensory input into the system. The rules of temporal evolution of activities described by this equation are general enough to include almost all interesting phenomena and mimic modeling of real systems on digital computers. In appendix B we will prove that this model is equivalent to systems with continuously defined time.

Noise is specified by the parameter $\vec{\eta}(t)$, which has a zero mean and is defined by the correlation matrix

$$\overline{\eta_i(t_1)\eta_j(t_2)} = \mathcal{N}_{ij}\delta_{t_1,t_2} \tag{20}$$

We assume here that neighboring in time values on noise are not correlated, implying that we consider a system with white noise. This assumption can be easily relaxed and is used here to simplify the analysis. It becomes rigorously valid when time-interval τ is longer than the correlation time of noise. Further, if noise is specific to each neuron, the same-time correlation matrix $\hat{\mathcal{N}}$ is diagonal

$$\mathcal{N}_{ij} = \overline{\eta_i^2}\delta_{ij} \tag{21}$$

This takes place i.e. when stochasticity is induced by probabilistic nature of synaptic vesicle release, in which case every two neurons receive uncorrelated fluctuating inputs.

Some time after presentation of the stimulus $[\vec{s}(t) \neq 0]$ the system is forced to make a decision through the following process. First, a scalar quantity

$$y = \vec{v}^T \cdot \vec{x}(t) \tag{22}$$

is evaluated. Here time corresponds to the instant, when the choice is to be made. The output metrics vector $\vec{v}$ describes the way in which system's activity affects motor response. In the simplest case, which was considered in the previous section, when a single element number n evokes responses, $v_i = \delta_{in}$. In a more complex situation, when multiple areas/neurons have direct influence on decision, vector $\vec{v}$ has more than one non-zero element. On the second step, decision is made based on the sign of y

$$d = H(y) \tag{23}$$

Thus, this model describes a two-alternative forced-choice task.

Our system is completely defined by the following set of parameters: $\hat{C}$, $\vec{s}(t)$, $\hat{\mathcal{N}}$, and $\vec{v}$. As we have shown in the previous section, the presence of the stimulus is not required to define DM elements [Eq. (5)]. We therefore set $\vec{s}(t)$ to zero and are left with three parameters $\hat{C}$, $\hat{\mathcal{N}}$, and $\vec{v}$. We now are ready to determine DM elements in our simple model.

To find decision makers we will use Eq. (14). In this case it becomes

$$DM_i = \mathcal{N}_{ii}\frac{\partial\sigma^2(y)}{\partial\mathcal{N}_{ii}} \tag{24}$$

Therefore, we need to evaluate the variability on the output from the system $\sigma^2(y)$. This is accomplished if we notice that $y = \vec{v}^T \cdot \vec{x}(t) = \vec{x}^T(t) \cdot \vec{v}$ and

$$\sigma^2(y) = \vec{v}^T \cdot \overline{\vec{x}(t)\vec{x}^T(t)} \cdot \vec{v} = \vec{v}^T\hat{X}(t,t)\vec{v} \tag{25}$$

Here we introduced the cross-correlation matrix defined as follows

$$\hat{X}(n,k) \equiv \overline{\vec{x}(n)\vec{x}^T(k)} = \overline{\begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} \begin{pmatrix} x_1 & \cdots & x_N \end{pmatrix}} \tag{26}$$

We replace here the time variable by the integers, specifying the time-slice number. The averaging in (25) and (26) is assumed over different instantiations of noise (trials).

Due to the properties of noise in our model, this correlator does not depend on the absolute values of time (n and k), but only on the difference $(n-k)$. As follows from (25), of particular interest is the same-time correlator $\hat{X}_0 \equiv \hat{X}(n,n)$, which determines fluctuations in y. We now derive equation for same-time correlator $\hat{X}_0$.

Using (19) we obtain

$$\begin{aligned} \hat{X}_0 &= \overline{\vec{x}(n+1)\vec{x}^T(n+1)} = \\ &= \overline{\left[\hat{C}\vec{x}(n) + \vec{\eta}(n)\right]\left[\vec{x}^T(n)\hat{C}^T + \vec{\eta}^T(n)\right]} \end{aligned} \tag{27}$$

We then notice that the correlator $\overline{\vec{x}(n)\vec{\eta}^T(n)}$ is identically zero, since $\vec{x}(n)$ is a linear combination of values of noise at times $k < n$ [see Eq. (19)]. We thus deduce from Eq. (27) that

$$\hat{X}_0 = \overline{\hat{C}\vec{x}(n)\vec{x}^T(n)C^T} + \overline{\vec{\eta}(n)\vec{\eta}^T(n)},$$

which leads us, finally, to

$$\hat{X}_0 - \hat{C}\hat{X}_0\hat{C}^T = \hat{\mathcal{N}} \tag{28}$$

This equation allows us to determine the same-time correlator $\hat{X}_0$ from connectivity and noise cross-correlogram, defined in (20), which is a diagonal matrix.

We would like to pause here and describe the properties of this equation. First of all, in the most generic case (28) allows us to determine $\hat{X}_0$ from $\hat{C}$ and $\hat{\mathcal{N}}$ uniquely. Indeed, (28) is a system of N^2 linear equations for N^2 unknowns $\hat{X}_0$, arranged in the matrix form. Hence, this system, in most cases, can be solved uniquely. On the other hand, with one exception, $\hat{X}_0$ cannot be expressed explicitly in terms of matrices $\hat{C}$ and $\hat{\mathcal{N}}$. Thus, one has to either appeal to the representation of $\hat{X}_0$ in terms of eigenvectors and eigenvalues of $\hat{C}$, or use computer to arrange elements of matrix $\hat{X}_0$ in vector form and solve resulting linear system.

The contribution to DM from a given element can be determined from Eq. (25)

$$DM_i = \mathcal{N}_{ii}\frac{\partial\sigma^2(y)}{\partial\mathcal{N}_{ii}} = \vec{v}^T\frac{\partial\hat{X}_0}{\partial\ln\mathcal{N}_{ii}}\vec{v} \tag{29}$$

The topological DM contributions are

$$TDM_i = \vec{v}^T\frac{\partial\hat{X}_0}{\partial\mathcal{N}_{ii}}\vec{v} \tag{30}$$

Using Eqs. (28) and (29) one can analyze a variety of network connectivities. Some new effect emerging for non-tree systems are described next.

B. Case 1: fan-out hub effect

We now consider network shown in Figure 11A, in which all elements have the same variance of noise and all connections have unitary strength. Figure 11A shows two pathways from unit **2** to the exit unit, **6**. The resulting network gain from unit **2** to unit **6** is thus equal to two. All other units' gain at the exit is one. The contribution to DM from unit **2** is thus four times larger that from other units. This is because noise at this unit is multiplied by a factor of two, and the variance of noise, by a factor of four. We conclude that there may be some special elements in network, which occupy hub-like positions, gaining large influence due to abundance of their outputs. It should be noted that fan-in hubs are not special from the point of view of DM in any way.

C. Case 2: temporal integrator

Let us now examine the network with a loop. Figure 11B shows such an example with unitary link strength and uniform noise variance, as in previous case. The presence of loop affects DM drastically: our discrete-time model marks units belonging to the loop as decision makers[1]. This is easy to understand, since noise, generated by each unit on each time-step, cannot leave the loop and, therefore, builds up there without limits. Therefore the variance of noise in the output of element number three grows proportionally to time $\overline{[x_3(t) - \overline{x_3}(t)]^2} = \overline{\eta^2}t \to \infty$. Here averaging is assumed over instantiations of noise (trials). Thus, loop becomes the crucial decision maker. This case is somewhat analogous to our previous 'noisy' neuron example.

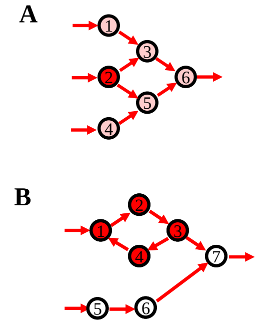

FIG. 11 Two cases, in which the identities of decision makers can be found using discrete-time approach. Variance of noise on all elements is the same; all network links have unitary strength. The degree of decision making is shown by the intensity of red. **A**, The fan-out effect. **B**, the temporal integrator.

What is the possible role of loops in biological networks? Why would one introduce such unreliable components? Loops, similar to shown in Figure 11B, have many useful properties. For instance, they can act as parametric memory systems. Indeed, imagine that responses of all units in the loop have the same values, equal to x. This could be accomplished by manipulating the sensory inputs. Assume that no more inputs are received from the outside of the system. It follows that, in the absence of noise on each element, this value of response will reverberate around the loop forever. This is because all links have unitary strength. Loops can thus memorize a graded value, such as x, functioning as parametric memory elements.

Suppose, in addition, that a non-zero input s is applied to element number **1** at all times. Since this element acts as a summator, its response on the next step is $x_1(1) = x + s$. The signal s propagates around the loop, and in four steps it reaches the first element again, at which time its response is $x_1(5) = x + 2s$. In four more steps $x_1(9) = x + 3s$. Thus, not only noise, but also signal can build up in the system. Therefore, a loop can operate as

[1] Rigorously speaking, the set of equations (28) and (29) does not have a valid solution for the loop with all connection equal to unity. One needs to set one of the connection as a parameter, $\alpha < 1$, solve the equations, and consider the limit $\alpha \to 1$.

a temporal integrator. The integration is not perfect if one of the links has a non-unitary strength, in which case integrator becomes leaky (Robinson, 1989).

Temporal integrators play special role in DM, since they act as accumulators of sensory information, which puts them into special position with respect to other areas (Gold and Shadlen, 2002). As an example, such is area LIP in primate visual cortex, which is involved in DM in direction-discrimination task (Shadlen and Newsome, 2001; Roitman and Shadlen, 2002; Mazurek et al., 2003).

D. Continuous-time model

We finally consider a model, in which time runs continuously. This model has potential relevance to real-life networks. The responses of units satisfy the following equation

$$\frac{d\vec{x}(t)}{dt} = -\hat{A}\vec{x}(t) + \vec{\eta}(t) + \vec{s}(t). \tag{31}$$

The network connectivity matrix $\hat{A}$ can be related to connection matrix from the discrete-time model in (19) through $\hat{C} = e^{-\hat{A}\tau}$ (see Appendix B). Noise is defined by its cross-correlation

$$\overline{\eta_i(t_1)\eta_j(t_2)} = \mathcal{N}_{ij}\delta(t_1 - t_2), \tag{32}$$

where

$$\hat{\mathcal{N}} = \begin{pmatrix} \overline{\eta_1^2} & & 0 \\ & \overline{\eta_2^2} & \\ & & \cdots & \\ 0 & & & \overline{\eta_N^2} \end{pmatrix} \tag{33}$$

is a diagonal cross-correlogram of noise. Eqs. (31)-(33) are analogous to the discrete-time case (19)-(21). Similarly, we define the output scalar and the decision variable

$$y = \vec{v}^T \cdot \vec{x}(t)$$

$$d = H(y). \tag{34}$$

Here t is time when the system makes the decision.

Our model is thus defined by Eqs. (31)-(34). We will now use definition (29) to find decision makers. As in discrete-time case we need to know the variance of the output variable, $\sigma^2(y)$, after which (29) leads to

$$DM_i = \overline{\eta_i^2}\frac{\partial \sigma^2(y)}{\partial \overline{\eta_i^2}} \tag{35}$$

Important for us is the time-dependent correlator

$$\hat{X}(t_1, t_2) = \overline{\vec{x}(t_1)\vec{x}^T(t_2)}, \tag{36}$$

which we now evaluate. Solution of (31) is obtained using matrix exponentials

$$\vec{x}(t) = \int_{-\infty}^{t} dt' e^{\hat{A}(t'-t)} \left[\vec{\eta}(t) + \vec{s}(t)\right] \tag{37}$$

If external stimulus is zero or a constant in time, due to (5), the correlator at $t_1 > t_2$

$$\hat{X}(t_1, t_2) = \int_{-\infty}^{t_2} dt' e^{\hat{A}(t'-t_1)}\hat{\mathcal{N}} e^{\hat{A}^T(t'-t_2)} \tag{38}$$

We seek $\hat{X}(t_1, t_2)$ in the form

$$\hat{X} = e^{\hat{A}(t_2-t_1)}\hat{X}_0, \tag{39}$$

where $\hat{X}_0$ is equal-time cross-correlation. To find equation for $\hat{X}_0$ we differentiate (38) as follows

$$\begin{aligned} \frac{\partial \hat{X}}{\partial t_2} &= \hat{A}e^{\hat{A}(t_2-t_1)}\hat{X}_0 = \\ &= e^{\hat{A}(t_2-t_1)}\hat{\mathcal{N}} - \hat{X}\hat{A}^T \end{aligned} \tag{40}$$

We arrive thus to the following equation for $\hat{X}_0$

$$\hat{A}\hat{X}_0 + \hat{X}_0\hat{A}^T = \hat{\mathcal{N}} \tag{41}$$

This equation is the central tool for the continuous-time theory. The contributions to DM from each unit are found by differentiating $\sigma^2(y) = \vec{v}^T\hat{X}_0\vec{v}$ with respect to noise, as in Eq. (29)

$$DM_i = \mathcal{N}_{ii}\vec{v}^T\frac{\partial \hat{X}_0}{\partial \mathcal{N}_{ii}}\vec{v} \tag{42}$$

Once the same-time correlation matrix $\hat{X}_0$ is found from Eq. (41), cross-correlation for arbitrary time is

$$\hat{X}(t_1, t_2) = \begin{cases} e^{\hat{A}(t_2-t_1)}\hat{X}_0, & t_1 \geq t_2 \\ \hat{X}_0 e^{\hat{A}^T(t_1-t_2)}, & t_1 < t_2 \end{cases} \tag{43}$$

This equation suggests a helpful strategy for determining noise matrix $\hat{\mathcal{N}}$. Indeed, (41) and (43) imply that

$$\hat{\mathcal{N}} = \left.\frac{\partial \hat{X}(t_1, t_2)}{\partial t_1}\right|_{t_1=t_2-\varepsilon} - \left.\frac{\partial \hat{X}(t_1, t_2)}{\partial t_1}\right|_{t_1=t_2+\varepsilon} \tag{44}$$

Here ε is infinitesimally small positive number. In other words, noise matrix is equal to discontinuity in time-derivative of cross-correlation at $t_1 = t_2$. Since noise correlation matrix is diagonal, the non-zero elements are

$$\overline{\eta_i^2} = \left.\frac{\partial X_{ii}(t_1, t_2)}{\partial t_1}\right|_{t_1=t_2-\varepsilon} - \left.\frac{\partial X_{ii}(t_1, t_2)}{\partial t_1}\right|_{t_1=t_2+\varepsilon} \tag{45}$$

Two comments are in order here. First, noise term $\vec{\eta}(t)$ plays the role of input noise in (31). It cannot be measured directly. Equation (44) provides a way to single it out. Second, (44) does not apply to the discrete-time model. Indeed, in the latter we either have $t_1 = t_2$, or $t_1 = t_2 \pm 1$, etc., i.e. the condition $t_1 = t_2 \pm \varepsilon$ with ε infinitesimally small is hard to enforce. It may happen that $\varepsilon \approx 1$ is acceptable due to presence of slow components in the circuit, such as temporal integrators. However, in general case (44) cannot be applied to the discrete-time case. For instance, it fails dramatically for the case of 'nematode' chain considered above.

Equations (41), (42), and (44) represent a useful set of tools to find DM components for various connectivities. We present here two possible cases, in which decision makers can be found. They differ in what is known about the system.

Scenario 1: Assume we know the network connectivity $\hat{A}$, output metrics vector $\vec{v}$, and autocorrelation for each unit $X_{ii}(t_1, t_2)$. The steps below allow finding the decision makers.

1. Since noise matrix is diagonal, as per (33), it can be found from autocorrelation using (44).
2. Solving (41) allows determining $\partial \hat{X}_0 / \partial \mathcal{N}_{ii}$, the derivative of equal-time crosscorrelation with respect to noise in each element.
3. Decision makers are found from (42).
4. Normalize contributions to DM so that $\sum_i DM_i = 1$.

Scenario 1 does not require simultaneous measurements from all units. It requires the knowledge of the network connectivity however. The next scenario is complementary in this respect.

Scenario 2: Suppose we have measured the full crosscorrelation matrix $\hat{X}(t_1, t_2)$ by simultaneous recordings from all units. Suppose also that we know how the output of the system is evaluated (vector $\vec{v}$). These are the steps to determine DM units.

1. Use (44) to find noise matrix $\hat{\mathcal{N}}$.
2. Use (41) to find the connection matrix $\hat{A}$.
3. Solve (41) to calculate $\partial \hat{X}_0 / \partial \mathcal{N}_{ii}$ for each element.
4. Use (42) to find decision makers.
5. Normalize contributions to DM so that $\sum_i DM_i = 1$.

Both scenarios use extensive knowledge about the system, which renders them useless in experimental conditions. In the next subsection we discuss a way to bypass these limitations.

Finally, we would like to provide solution to (41) using eigenbasis of matrix $\hat{A}$. Since $\hat{A}$ is not necessarily symmetric, a distinction should be made between right and left eigenvectors. The latter turn out to be useful for our purposes. They are defined by

$$\vec{\xi}_\alpha^{+} \hat{A} = \lambda_\alpha \vec{\xi}_\alpha^{+}. \tag{46}$$

Here and below Greek indexes denote numbers of eigenvalues, while Latin ones label spatial components of vectors and matrices. Solution of (41) is

$$X_{0ij} = \sum_{\alpha\beta\gamma\delta mn} \frac{\xi_{i\alpha}\xi_{j\beta}^* \xi_{m\gamma}^* \xi_{n\delta}}{\lambda_\gamma + \lambda_\delta^*} \left(G^{-1}\right)_{\alpha\gamma} \left(G^{-1}\right)_{\beta\delta}^* \mathcal{N}_{mn}, \tag{47}$$

where

$$G_{\alpha\beta} = \sum_i \xi_{i\alpha}^* \xi_{i\beta} \tag{48}$$

is the Gram matrix of eigenvectors. Eq. (47) is valid if the eigenvectors form a complete basis in the N-dimensional space. As follows from (47), eigenvalues of $\hat{A}$ with small real part contribute to DM in a large degree. This justifies the use of principal component analysis when such eigenvalues are present. An example of such principal component is the temporal integrator loop in Figure 11B, which has vanishing λ.

In case if matrix $\hat{A}$ is symmetric, its eigenvalues are real and eigenvectors are orthogonal. This leads to a unit Gram matrix. Then, Eq. (3.29) becomes more compact

$$X_{0ij} = \sum_{\alpha\beta mn} \frac{\xi_{i\alpha}\xi_{j\beta}^* \xi_{m\alpha}^* \xi_{n\beta}}{\lambda_\alpha + \lambda_\beta} \mathcal{N}_{mn} \tag{49}$$

Similar equations, called Kubo formulas, are obtained for various correlators in case of diffusion of particles in random media (Efetov, 1997). The distinguishing feature of (49) is that a product of four eigenvectors enters the expression. Thus, propagation of noise in this case can be accompanied by interference between different pathways. An example of destructive interference of this kind is given below, in section IV.

Eq. (49) can be further simplified. Indeed, our model uses diagonal noise matrices, i.e. $n = m$ in (49). Suppose also that the output from the network occurs through one exit element number i, which is specified by taking $\vec{v} = \hat{e}_i$. In this case the use of Eq. (42) gives

$$DM_i = \mathcal{N}_{nn} \sum_{\alpha\beta} \frac{\xi_{i\alpha}\xi_{i\beta}^* \xi_{n\alpha}^* \xi_{n\beta}}{\lambda_\alpha + \lambda_\beta}. \tag{50}$$

From this equation we conclude that for element n to contribute to DM, an eigenvector should exist, which is non-zero on both unit number n and exit unit i. Thus, we conclude that eigenvectors of $\hat{A}$ should be delocalized for broader impact of elements on the decision. This is not surprising in view of the mentioned analogy with the diffusion problem. In case if matrix $\hat{A}$ is not symmetric, the Gram matrix may be non-diagonal and (50) cannot be used. However, the off-diagonal elements of $\hat{G}$ are usually smaller than diagonal ones, due to uncorrelated sign changes, when (48) is computed with $\alpha \neq \beta$. Therefore, (50) may apply approximately.

IV. ANALYSIS USING STIMULATION

Stimulations with electric current add a new degree of freedom to DMA, thus leading to more effective ways of finding decision makers. There are two great advantages of the stimulation method. First, it only involves stimulation of a single neuron, therefore no simultaneous multiple-electrode measurements are required. Second, the knowledge of network connectivity is not needed to solve the problem. In this section we study our simple networks and find what stimulation strategies are consistent with our earlier definitions, such as Eq. (14).

We will use continuous model for concreteness (section III.D). Consider the output variable y. It is a linear function of the inputs. It is also a function, which contains noise components, variable from trial to trial. The noise components were acquired from all units in different degree. Since noise in each unit is gaussian, the output variable is described by gaussian distribution too

$$\rho(y) = \frac{1}{\sqrt{2\pi\sigma^2(y)}} e^{-(y-\bar{y}(s))^2/2\sigma^2(y)} \tag{51}$$

In each trial a random value of y is obtained, according to distribution (51). The response of the system is equal to 1 if y is positive, and 0 otherwise. The probability to obtain response equal to 1 to given stimulus s is given by the error function (Abramowitz and Stegun, 1972)

$$p_1(s) = \int_0^\infty \rho(y)dy = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\bar{y}(s)}{\sigma(y)\sqrt{2}}\right)\right], \tag{52}$$

whereas the probability of zero response is

$$p_0(s) = \int_{-\infty}^0 \rho(y)dy = \frac{1}{2}\left[1 - \operatorname{erf}\left(\frac{\bar{y}(s)}{\sigma(y)\sqrt{2}}\right)\right]. \tag{53}$$

Both probabilities depend upon the mean response to stimulus $\bar{y}(s)$ and the standard deviation $\sigma(y)$. Therefore the electric stimulation strategies may be based on affecting either the former or the latter. We now consider both of these strategies and show that affecting the mean response may provide misleading results, while changing the variance of response allows estimating contributions to DM consistently with our previous definitions. Thus, strategies of stimulation based on standard deviation of the output variable are *always* correct in our simple model, independently on the topology of the network. This may seem a trivial consequence of definition (14), but we will discuss it here for the sake of comparison of two strategies and optimizing them.

We start with the strategies of stimulation, which affect the mean response $\bar{y}(s)$. In our simple model this may be accomplished by injecting a tonic input current into a unit number i. Mathematically it is accomplished by adding extra stimulus s_i to this unit in Eq. (31). Note that in biological systems the stimulating current is alternating with constant amplitude (Salzman et al., 1992). The mean response is shifted by the stimulation, i.e.

$$\Delta\bar{y} = \frac{\partial\bar{y}}{\partial s_i} s_i, \tag{54}$$

where s_i is the magnitude of injected tonic current. This leads to observable changes in the probability p_1

$$\Delta p_1(i) = \frac{\partial p}{\partial\bar{y}}\frac{\partial\bar{y}}{\partial s_i} s_i. \tag{55}$$

Here $\Delta p_1(i)$ is the change in probability of correct responses after unit number i is electrically stimulated. Can $\Delta p_1(i)$ be a measure of DM?

We notice that $\Delta p_1(i)$ can be either positive or negative. This depends on the sign of derivative $\partial\bar{y}/\partial s_i$, which is positive for excitatory pathway from unit i to the output and negative for inhibitory pathway. Since contribution to DM ought to be positive, we cannot assume simply that $DM_i \sim \Delta p_1(i)$. The correct expression, which we provide here without derivation is

$$DM_i \sim \overline{\eta_i^2}\left[\Delta p_1(i)\right]^2. \tag{56}$$

This equation is understood in proportional sense, since DM_i should be normalized to ensure that $\sum_i DM_i = 1$. Our investigations show that this expression is accurate for trees and is consistent with both our earlier definitions (9) or (14). Remarkably, it employs quantities, which can be measured in a single-electrode experiment. Indeed, the amplitude of noise $\overline{\eta_i^2}$ can be found from autocorrelation of unit's response, using (44); and $\Delta p_1(i)$ is determined from behavioral changes in response to single-unit stimulation. This equation thus provides an approach potentially useful in practice. Does this relationship work for networks of arbitrary connectivity?

Figure 12B shows a counterexample, in which a unit is stimulated, which results in *no* change in probability of correct response (we consider trials in which 1 is the correct response throughout this section). This is because there are two pathways, leading from this unit to the exit, one positive and one negative. They have equal strength, and, therefore, compensate each other. On the other hand, unit number one *does* participate is DM, because if a non-stationary stimulation/stimulus is applied, its effect on the decision is not zero. Thus, (56) and tonic stimulation method cannot be applied to arbitrary circuits, such as shown in Figure 12B, to accurately reveal decision makers.

Is there a stimulation method for finding decision making components in arbitrary networks? The method follows directly from the definition (14) [or (35), which is

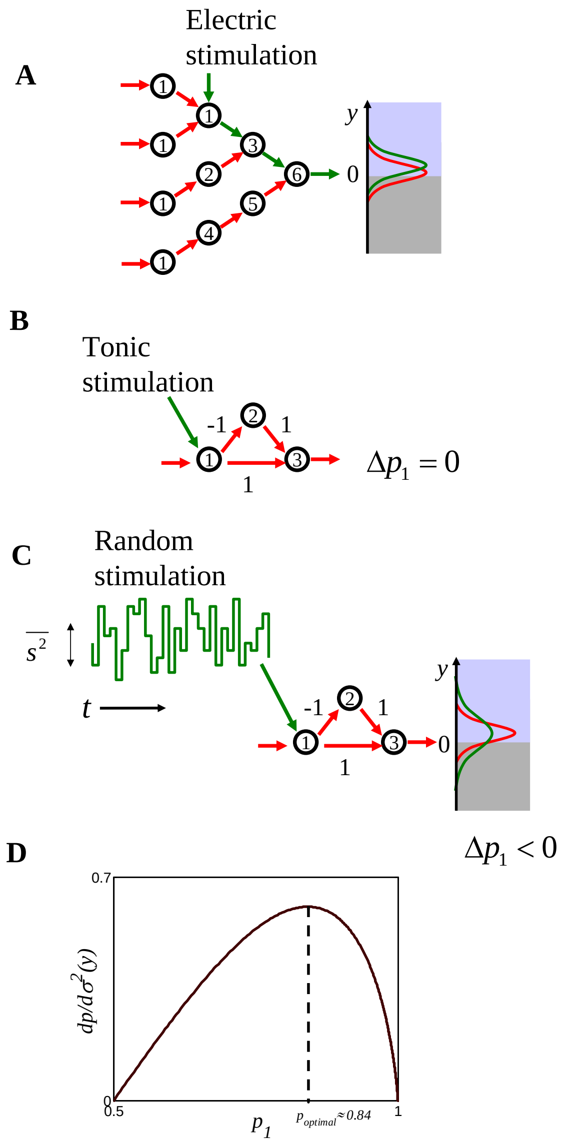

FIG. 12 Finding decision makers using electric stimulation. **A** and **B**, tonic stimulation; **C**, random stimulation. **A**, tonic stimulation for trees results in shift in probability, which leads to correct estimation of decision making units. **B**, example of a circuit for which tonic stimulation leads to incorrect estimation of decision making, since it does not lead to the shift in probability. **C**, stimulation with a random current leads to correct estimate of decision making for networks with *any* connectivity. **D**, for optimal performance in random stimulation paradigm, the task should be set so that the probability of correct responses is close to $p_{optimal} \approx 0.84$.

equivalent]. Indeed, when stimulating current is a temporal white noise, the output variable y acquires a larger variance (Figure 12C). Hence, the derivative of output variance, entering (35) can be calculated operationally, by injecting a distracter current. More precisely, if the variance of stimulating current applied to unit i is $\overline{s_i^2}$ the derivative entering definition (35) is

$$\frac{\partial \sigma^2(y)}{\partial \overline{\eta_i^2}} = \frac{\Delta \sigma^2(y)}{\overline{s_i^2}}. \tag{57}$$

In practice one has no access to the variable y, so one cannot measure directly the change in variance $\Delta\sigma^2(y)$. Instead, one could measure the change in the probability of correct responses under the influence of distracting current. Indeed, from (52) we obtain

$$\Delta p_1(i) = \frac{\partial p_1}{\partial \sigma^2(y)} \Delta \sigma^2(y) \tag{58}$$

Combining the last two equations we obtain for the important derivative

$$\frac{\partial \sigma^2(y)}{\partial \overline{\eta_i^2}} = \frac{\Delta p_1(i)}{\overline{s_i^2}} \left(\frac{\partial p_1}{\partial \sigma^2(y)} \right)^{-1} \tag{59}$$

Since the probability of correct responses always decreases under the influence of distracters, the derivative $\partial p_1/\partial\sigma^2(y)$ is a negative constant. It is the same for all units. We arrive therefore to the expression for contributions to DM, which follows from (35)

$$DM_i \sim -\overline{\eta_i^2} \frac{\Delta p_1(i)}{\overline{s_i^2}} \tag{60}$$

Here $\Delta p_1(i)$ is the decrease in probability of correct responses produced by electric stimulation with variance of the random current equal to $\overline{s_i^2}$. The variance of noise on each unit $\overline{\eta_i^2}$ can be found from autocorrelation using (45). This procedure works for any topology in our simplified model. It should be noted here that if noise is not entirely white or cannot be considered white, (45) cannot be used directly and should be replaced by an expression reflecting the spectral characteristics of noise appropriate for the system under investigation. Thus, if noise is provided by other parts of the network, its dynamic features may be more complex. Therefore, (45) may not apply directly to the 'hidden pathway' example given in the end of section II.B.

The procedure, which we just described, permits further optimization. Indeed, imagine that the probability of correct responses is exactly ½. Adding distracting stimulation current will not change this probability, i.e. $\Delta p_1(i) = 0$ no matter what unit is stimulated. In the opposite limiting case when $p_1 \approx 1$, the effect of distracter on performance is exponentially small. Hence, behavioral response to stimulation has an optimum between $p_1 = 1/2$ and 1. To find the optimum we observe from (58) that Δp_1 is maximum for the same variation in $\Delta\sigma^2(y)$ when $\partial p_1/\partial\sigma^2(y)$ is maximum. We therefore plot the latter derivative as a function of p_1 in Figure 12D. We indeed observe a maximum at the value of probability of correct responses close to

$$p_{optimal} \approx 0.841 \tag{61}$$

To summarize, the following scenario describes algorithm for finding contributions to DM using random stimulation.

Scenario 3: Assume that we *do not know* the network connectivity $\hat{A}$ and output metrics vector $\vec{v}$; but we know autocorrelation for each unit $X_{ii}(t_1, t_2)$. The steps below allow finding the decision makers.

1. Prepare stimulus so that the probability of correct responses is close to the value given by (61).

2. Stimulate one unit with random current, whose variance is $\overline{s_i^2}$, and measure the decrease in probability of correct responses Δp_1.

3. Record autocorrelation and evaluate noise variance $\overline{\eta_i^2}$ for this unit using (45).

4. Find contribution to DM for this unit using equation (60).

5. Repeat steps 1 through 4 for all units in the system.

6. Normalize contributions to DM so that $\sum_i DM_i = 1$.

V. DISCUSSION

In this work we defined decision makers in networks, which behave in a well-defined fashion. As with any definition, there is certain degree of arbitrariness in our study, since this is the first mathematical study of this sort. We had to make choices about the features of decision making we were attempting to describe as well as about the way they were quantified. We demonstrated these features in a set of examples. Future studies will show if these features can be used as a basis of a more complete model-independent theory.

In this study we postulated that variability and noise, causally linked to decisions, are the chief descriptors of DM. Although this point may seem paradoxical we suggest three arguments in its favor. First, variability may reflect additional information needed to make a decision in case of uncertainty. Such may be inputs from other modalities, memories, or some other relevant modulatory inputs, supplying e.g. emotional condition of the subject or changing utility values (Figure 8). Second, many behaviors, such as C-start escape responses in fish (Eaton and Emberley, 1991) and other organisms (Glimcher, 2003), have stochastic character. This makes the task of pursuer more difficult. Such unpredictable behaviors are reproduced in our model if the sensory input is weak or in the small signal-to-noise ratio case. Third, the goal of DM is to dissipate sensory information, as suggested in the introduction (Figure 1A), whereby an analog multi-dimensional stimulus space is reduced to a discrete space of several decisions. We argue that this transformation is facilitated by noise.

We have studied the problem of finding decision making units in networks of various connectivities. This path took us from simple linear chains, for which the information-theoretical (IT) approach was found to be effective, to trees, and, finally to an alternative definition of decision makers, based on propagation of noise in networks. This latter definition is valid in networks of arbitrary topology. All these approaches are equivalent, when they can be compared, but include progressively broader classes of networks. As a practical application for the alternative definition we considered the problem of electric stimulation in the surrogate networks and showed a way of determining DM contributions for arbitrary networks using stimulation with random current. Our findings are summarized in Figures 13 and 14.

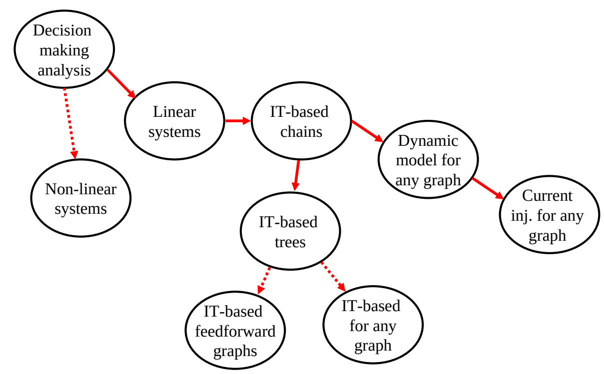

FIG. 13 The cluster of problems covered in this study. Solid/dashed arrows show the derivations performed here/yet to be confirmed or denied. IT stands for information-theoretical.

Although we studied networks of complex connectivity, the model describing a single network element was quite simple. Not all of the units are linear, of course, since DM is a non-linear task (Figure 1A). However, our model is essentially based on linear elements. The motivation for this model is that it is easy to analyze. The study of simple models is a necessary step before analysis proceeds any further. Once the methodological issues are resolved for simpler models, complex non-linear systems can be studied in the same paradigm. One of important questions resolved here is that a completely linear element can be a decision maker, despite the presence of non-linear units in the network. Thus, nonlinearity is not a necessary attribute of DM. This question would be impossible to answer for more realistic system, since in practice all units contain nonlinearities.

Decision making task, as formulated in Figure 1A, is similar to general object discrimination task. Representation of motor response in our model is not distinguishable mathematically from the representation of abstract object/decision category (Horwitz and Newsome, 1998; Shadlen and Newsome, 2001). The latter does not necessarily lead to a motor command. Thus, our analysis may uncover the identities of units responsible for categorization of sensory inputs. In terms of this analysis

Method	Requires knowledge of connectivity	Applies to any connectivity	Requires intrusion into system	Significance
IT-based	Yes	No	No	Is used to prove noise-based definition
Dynamic models	Yes	Yes	No	Solves problem for arbitrary connectivity
Current injection	No	Yes	Yes	Does not demand the knowledge of connectivity

FIG. 14 Comparison between different approaches studied here.

we emphasize the distinction between units representing the object category and the units in which this representation is actually formed. The former are analogous to motor units in the decision task, while the latter are similar to decision makers. As follows from this study, the analysis is dependent upon the topology of the network involved. For simple linear sensory chains our conclusion is that the *first* unit, spatially or temporally, in which the representation of the object is correlated with final outcome of the discrimination process, is responsible for casting the stimulus in one of the abstract classes. In case of recurrent networks a more detailed quantitative analysis is needed to draw conclusions about identities of categorizing units. Thus, DMA may find a broader use in identifying units representing abstract object's percepts.

A special care should be taken in distinguishing the DM task from the sensory discrimination task. It may occur that in the same experiment these tasks are performed by different populations of neurons. An example is given by (Salinas and Romo, 1998). They discovered a population of M1 neurons responding differentially to two categories of tactile stimuli. Some of these neurons did not respond, when the same behavior was guided by visual cues. This observation is consistent with these neurons performing sensory discrimination of tactile stimuli, while some other population making decisions about the actual motor response. Our mathematical analysis is general enough to include both of these functions. Thus, if correlations with motor response are studied, it will result in the decision makers; while when the correlations with percepts are investigated, DMA should provide the identities of discriminating elements.

We suggest that DMA may be relevant to other biological systems. Possible applications may include the analysis of molecular networks, such as genetic regulatory or protein binding networks; finding decision makers in compartmental models of dentritic trees (Poirazi and Mel, 2001); studies of neural networks and structural networks of connectivities between different brain areas; and analysis of social networks.

VI. CONCLUSION

In this study we define network elements responsible for making decisions. We obtain two equivalent definitions. According to one, decisions are made by elements, in which correlations with the decision are first formed. According to the second definition, decision making activity is measured by the impact of variability in given unit on the response. We give examples of network motifs, especially potent from decision making prospective, such as fan-out hubs and recurrent loops. The latter can function as temporal integrators of sensory inputs. We also study how electric stimulations can reveal decision making components. We conclude that stimulations with time-varying random current produce correct results for all network topologies.

APPENDIX A: The linear chain model.

Here we solve a more general version of linear chain model than considered in the text. The responses of neighboring neurons are related linearly

$$x_i = C_{i-1} x_{i-1} + \eta_i \tag{A1}$$

This is a generalization of (1). The response of the nth unit is

$$x_n = \sum_{i=1}^{n} \alpha_{ni} \eta_i + \alpha_{n0} x_0 \tag{A2}$$

where coefficients $\alpha_{ni} = C_{n-1} C_{n-2} \ldots C_i$, $\alpha_{nn} = 1$. The external signal x_0 is assumed to be zero in this appendix, due to (5). For the last element in the chain we have

$$x_N = \sum_{i=1}^{N} \alpha_{Ni} \eta_i . \tag{A3}$$

Comparing (A2) and (A3) we conclude that

$$x_N = \alpha_{Nn} x_n + \xi , \tag{A4}$$

where ξ is a variable, which describes noise in the networks downstream from unit n. It is, thus, uncorrelated with x_n. This is where tree-like topology enters our solution, since in case of loops, x_n and ξ are correlated. Our goal now is to calculate MI between the decision variable $d = H(x_N)$ and x_n. We will use the definition for MI

$$MI(d, x_n) = \sum_{d=0,1} \int_{-\infty}^{\infty} dx_n \rho\left(d, x_n\right) \log_2 \left[\frac{\rho\left(d, x_n\right)}{\rho(d)\rho(x_n)} \right] \tag{A5}$$

Here $\rho\left(d\right) = 1/2$, since there is no signal;

$$\rho\left(x_n\right) = \exp\left(-x_n^2 / 2\overline{x_n^2}\right) / \left(2\pi \overline{x_n^2}\right)^{1/2} \tag{A6}$$

and

$$\rho(d;x_n) = \frac{\rho(x_n)}{2}\left[1 \pm \mathrm{erf}\left(\frac{\alpha_{Nn}x_n}{\sigma(\xi)\sqrt{2}}\right)\right]. \tag{A7}$$

The upper/lower sign is assumed for $d = 0$ or 1 in (A7); $\sigma(\xi)$ is the standard deviation of Gaussian variable ξ defined in (A4). The expression for MI (A5) results in

$$\begin{aligned} &MI_n = M(s_n) \\ &M(s_n) = \tfrac{1}{\sqrt{\pi}}\int\limits_{-\infty}^{\infty} dz e^{-z^2}\left[1+\mathrm{erf}(zs_n)\right]\log_2\left[1+\mathrm{erf}(zs_n)\right] \\ &s_n = \sigma(\alpha_{Nn}x_n)/\sigma(\xi). \end{aligned} \tag{A8}$$

MI is therefore a function of signal-to-noise ratio s_n. Inversely,

$$s_n^2 = \frac{\alpha_{Nn}^2\overline{x_n^2}}{\overline{\xi^2}} = \left[M^{-1}(MI_n)\right]^2 \tag{A9}$$

On the other hand, (A4) leads to

$$\overline{x_N^2} = \alpha_{Nn}^2\overline{x_n^2} + \overline{\xi^2} \tag{A10}$$

Solving (A9) and (A10) with respect to $\alpha_{Nn}^2\overline{x_n^2}$ we have

$$\frac{\alpha_{Nn}^2\overline{x_n^2}}{\overline{x_N^2}} = \frac{\left[M^{-1}(MI_n)\right]^2}{1+\left[M^{-1}(MI_n)\right]^2} \equiv F(MI_n) \tag{A11}$$

Function M^{-1} here is inverse to M defined in (A8). Function $F(MI)$ numerically calculated from (A8) and (A11) is shown in Figure 5. Lastly, we recall that variances $\alpha_{Nn}^2\overline{x_n^2}$ are related to the strength of noise $\overline{\eta_i^2}$ through (A2). We have

$$\alpha_{Nn}^2\overline{x_n^2} = \sum_{i=1}^{n}\alpha_{Ni}^2\overline{\eta_i^2} \tag{A12}$$

Eqs. (A11) and (A12) are used below to prove a variety of statements about function $F(MI)$ used in the main text.

1. In the uniform noise example $F(MI)$ is a linear function of position in the chain.

In this case $C_1 = \ldots = C_{N-1} = 1$, and, consequently, $\alpha_{N1} = \ldots = \alpha_{NN} = 1$. Noise variance is the same on every node, i.e. $\overline{\eta_i^2} \equiv \eta^2$. As follows from (A12) $\overline{x_n^2} = \eta^2 n$, which results in

$$F(MI_n) = n/N \tag{A13}$$

It follows that contributions to DM defined by (9) are the same for all units.

2. In the 'loud' neuron example the contributions of units upstream from the strong link are larger by a factor of K^2 than contribution from the downstream units.

In this case $\alpha_{1\ldots k} = K$, while $\alpha_{k+1\ldots N} = 1$, assuming that the link from unit k to $k+1$ is strengthened. In the example in the text $k = 5$ [cf. (10)]. Eq. (A12) leads us to the values for variances of responses

$$\alpha_{Nn}^2\overline{x_n^2} = \begin{cases} \eta^2K^2n, & n \le k \\ \eta^2K^2k + \eta^2(n-k), & n > k \end{cases} \tag{A14}$$

Applying (A11) we obtain the expression for $F(MI)$

$$F(MI_n) = \begin{cases} \dfrac{K^2n}{N-k+K^2k}, & n \le k \\ \dfrac{n-k+K^2k}{N-k+K^2k}, & n > k \end{cases}, \tag{A15}$$

which is a piece-wise linear function of n. Eq. (9) determines contributions to DM as

$$DM_n = \begin{cases} \dfrac{K^2}{N-k+K^2k}, & n \le k \\ \dfrac{1}{N-k+K^2k}, & n > k \end{cases} \tag{A16}$$

This confirms that the upstream units ($n \le k$) are K^2 times more potent than the downstream ones ($n > k$).

3. Two definitions of contribution to DM using derivative of $F(MI)$ (9) and the impact of noise (14) are equivalent.

Let us start by determining decision makers from definition (14). According to (A3)

$$\overline{x_N^2} = \sum_{i=1}^{N}\alpha_{Ni}^2\overline{\eta_i^2}. \tag{A17}$$

Definition (14) gives

$$DM_i \propto \overline{\eta_i^2}\frac{\partial\overline{x_N^2}}{\partial\overline{\eta_i^2}} = \alpha_{Ni}^2\overline{\eta_i^2}. \tag{A18}$$

After normalization we obtain

$$DM_i = \frac{\alpha_{Ni}^2\overline{\eta_i^2}}{\overline{x_N^2}}. \tag{A19}$$

Let us derive the same result from (9). As follows from (A11)

$$\begin{aligned} &F(MI_n) - F(MI_{n-1}) = \frac{1}{\overline{x_N^2}}\left(\alpha_{Nn}^2\overline{x_n^2} - \alpha_{N\ n-1}^2\overline{x_{n-1}^2}\right) \\ &= \frac{\alpha_{Nn}^2}{\overline{x_N^2}}\left(\overline{x_n^2} - C_{n-1}^2\overline{x_{n-1}^2}\right) = \frac{\alpha_{Nn}^2\overline{\eta_n^2}}{\overline{x_N^2}} \end{aligned} \tag{A20}$$

This proves the equivalence of (9) and (14), since the result is identical to (A19).

APPENDIX B: Connection between discrete- and continuous-time models.

In this section we show that the discrete-time model can be derived from continuous-time model. Starting from equation (37) for the unit responses in the continuous case we obtain the relation for solutions at two different time points separated by the time-interval τ, analogous to (19) in the discrete-time description. Then we show that in the limiting case $\tau \to 0$ two descriptions are equivalent.

From (37) we obtain

$$\vec{x}(t+\tau) = e^{-\hat{A}\tau}\vec{x}(t) + \int\limits_{t}^{t+\tau} e^{-\hat{A}(t+\tau-t')}\vec{\eta}(t')dt'. \quad \text{(B1)}$$

This equation can be rewritten as $\vec{x}(t+\tau) = \hat{C}\vec{x}(t)+\vec{\eta}'(t)$, where

$$\hat{C} = e^{-\hat{A}\tau} \approx \hat{I} - \hat{A}\tau. \quad \text{(B2)}$$

Thus it has the same form as (19). Using (32) we obtain that the new noise cross-correlation matrix

$$\hat{\mathcal{N}}' = \int\limits_{0}^{\tau} e^{-\hat{A}t'}\hat{\mathcal{N}}e^{-\hat{A}^T t'}dt'. \quad \text{(B3)}$$

The solution of the continuous-time problem satisfies the equations of the discrete-time model for an arbitrarily large time interval τ, but the new noise cross-correlation matrix $\hat{\mathcal{N}}'$ is non-diagonal in this case. In the limiting case $\tau \to 0$ it becomes diagonal. Indeed (B3) implies that in this limit

$$\hat{\mathcal{N}}' = \hat{\mathcal{N}}\tau \quad , \quad \text{(B4)}$$

which is diagonal by the definition of the continuous-time model. Here we kept only terms linear in τ. Thus, in this limit the matrix $\hat{\mathcal{N}}'$ is diagonal as needed in our formulation of discrete-time model. One can also derive (41) from (28) using (B2) and (B4) and taking the limit $\tau \to 0$.

References

[1] Abramowitz M, Stegun IA (1972) Handbook of mathematical functions with formulas, graphs, and mathematical tables, 10th printing, 1972, with corrections. Edition. Washington: U.S. Govt. Print. Off.

[2] Eaton RC, Emberley DS (1991) How stimulus direction determines the trajectory of the Mauthner-initiated escape response in a teleost fish. J Exp Biol 161:469-487.

[3] Efetov K (1997) Supersymmetry in disorder and chaos. Cambridge [England] ; New York: Cambridge University Press.

[4] Glimcher PW (2003) The neurobiology of visual-saccadic decision making. Annu Rev Neurosci 26:133-179.

[5] Gold JI, Shadlen MN (2002) Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. Neuron 36:299-308.

[6] Horwitz GD, Newsome WT (1998) Neurophysiology: sensing and categorizing. Curr Biol 8:R376-378.

[7] Mazurek ME, Roitman JD, Ditterich J, Shadlen MN (2003) A role for neural integrators in perceptual decision making. Cereb Cortex 13:1257-1269.

[8] Poirazi P, Mel BW (2001) Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. Neuron 29:779-796.

[9] Robinson DA (1989) Integrating with neurons. Annu Rev Neurosci 12:33-45.

[10] Roitman JD, Shadlen MN (2002) Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J Neurosci 22:9475-9489.

[11] Romo R, Salinas E (2003) Flutter discrimination: neural codes, perception, memory and decision making. Nat Rev Neurosci 4:203-218.

[12] Salinas E, Romo R (1998) Conversion of sensory signals into motor commands in primary motor cortex. J Neurosci 18:499-511.

[13] Salzman CD, Murasugi CM, Britten KH, Newsome WT (1992) Microstimulation in visual area MT: effects on direction discrimination performance. J Neurosci 12:2331-2355.

[14] Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J Neurophysiol 86:1916-1936.