

Article

Intrinsic Computation of a Monod-Wyman-Changeux Molecule

Sarah Marzen 

Physics of Living Systems Group, Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; semarzen@mit.edu

Received: 12 July 2018; Accepted: 10 August 2018; Published: 11 August 2018



Abstract: Causal states are minimal sufficient statistics of prediction of a stochastic process, their coding cost is called statistical complexity, and the implied causal structure yields a sense of the process' "intrinsic computation". We discuss how statistical complexity changes with slight changes to the underlying model— in this case, a biologically-motivated dynamical model, that of a Monod-Wyman-Changeux molecule. Perturbations to kinetic rates cause statistical complexity to jump from finite to infinite. The same is not true for excess entropy, the mutual information between past and future, or for the molecule's transfer function. We discuss the implications of this for the relationship between intrinsic and functional computation of biological sensory systems.

Keywords: statistical complexity; intrinsic computation; excess entropy

1. Introduction

Intrinsic computation [1] is a theory of how a dynamical system "intrinsically" computes. In short, one makes a minimal maximally predictive model (or ϵ -machine) of the process generated by a dynamical system. States of the ϵ -machine are called "causal states", although these states are normally not causal in the sense of Ref. [2]. Certain words are forbidden, in that those words can never be seen. The words that are seen and thus accepted by the ϵ -machine constitute the ϵ -machine's language, in a nod to the computation performable by finite and infinite automata. The "memory stored by the process", the statistical complexity, is taken to mean the coding cost of the ϵ -machine's states.

One interesting hypothesis is that the ϵ -machine's structure provides a guide to the "functional" computation of the corresponding dynamical system. Functional computation—biologically-relevant computation, e.g., transformation of information with fitness consequences—might include everything from estimating past input [3,4] to predicting future input [5] to performing logical computations on input [6]. A more rigorous definition of "functional computation" remains an open problem; here, we merely list examples of quantities that can be identified with functional computations. As of yet, no link between intrinsic and functional computation has been found.

Here, we investigate the intrinsic computation and two functional computations (ligand concentration transduction and low-pass filtering) of a Monod-Wyman-Changeux (MWC) molecule, a widely-used model of a biological sensor [7–9]. This is the first time that the intrinsic computation of an MWC molecule—which here is limited to the ϵ -machine structure, the statistical complexity, and the excess entropy—has been calculated. The calculational techniques used here can be applied to study intrinsic computation of a more general class of biological sensors than previously studied.

We find that certain arbitrarily small perturbations to the underlying MWC molecule can lead to arbitrarily large perturbations in the process' intrinsic structure but lead to arbitrarily small changes in the stated functional computations. To the author's knowledge, this is the first such example in the literature. These results therefore suggest that causal structure and functional computation are orthogonal characterizations of a process, at least for oft-considered functional computations.

However, intrinsic computation could be taken to include several newer structure-related information-theoretic measures of a process, including excess entropy [10–12]. These newer measures do not suffer the same sensitivity as the ϵ -machine and statistical complexity, suggesting that these measures might help characterize functional computation.

Section 2 reviews the definition of ϵ -machine and MWC molecules. Section 3 explores how variations in kinetic rates change the molecule's intrinsic and functional computation. Section 4 discusses future research directions for intrinsic computation.

2. Background

The subject of interest here is a continuous-time, discrete-event process. However, for reasons explained later, for characterization of statistical complexity we will consider time-binning the process and treating it as a discrete-time, discrete-event process. Hence, we are also interested in discrete-time, discrete-event processes.

First, we discuss discrete-time, discrete-event processes. We code these processes as $\dots, x_{-1}, x_0, x_1, \dots$ where x_i is the i -th symbol appearing. The past \overleftarrow{x} (with corresponding random variable \overleftarrow{X}) is taken to be \dots, x_{-2}, x_{-1} while the future \overrightarrow{x} (with corresponding random variable \overrightarrow{X}) is taken to be x_0, x_1, x_2, \dots

Next, we discuss continuous-time, discrete-event processes. We code these processes as $\dots, (x_{-1}, \tau_{-1}), (x_0, \tau_0), (x_1, \tau_1), \dots$ where x_i is the i -th symbol, appearing for a total duration τ_i . We enforce $x_i \neq x_{i+1}$ so as to ensure a unique coding. The present is said to occur sometime during the presentation of x_0 , and so we denote the past $\overleftarrow{(x, \tau)}$ (with corresponding random variable $\overleftarrow{(X, T)}$) as $\dots, (x_{-1}, \tau_{-1}), (x_0, \tau_+)$ and the future $\overrightarrow{(x, \tau)}$ (with corresponding random variable $\overrightarrow{(X, T)}$) as $(x_0, \tau_-), (x_1, \tau_1), \dots$ where $\tau_+ + \tau_- = \tau_0$.

Section 2.1 reviews the definition of causal states, statistical complexity, the continuous-time ϵ -machine, and the mixed-state simplex. Section 2.2 reviews the dynamical models of Monod-Wyman-Changeux molecules used here.

We assume knowledge of information theory at the level of Ref. [13], but we briefly review definitions here. When X is a discrete random variable with probability distribution $p(x)$, then its entropy is $H[X] = -\sum_x p(x) \log p(x)$; when X is a continuous random variable with probability density function $\rho(x)$, then the differential entropy is $H[X] = -\int \rho(x) \log \rho(x) dx$; and when X is a mixed random variable (as is the case here), the entropy $H[X]$ is given by Ref. [14]. Entropy can be thought of as a measure of uncertainty. Conditional entropy of X conditioned on random variable Y is $H[X|Y] = \langle H[X|Y = y] \rangle_y$, and the mutual or shared information $I[X; Y]$ between two random variables X and Y is merely $I[X; Y] = H[X] - H[X|Y]$.

2.1. Causal States \mathcal{S} , Statistical Complexity C_μ , the ϵ -Machine, and the Mixed-State Simplex

Consider the equivalence relation \sim_ϵ that clusters two semi-infinite pasts, \overleftarrow{x} and \overleftarrow{x}' , together if $\Pr(\overrightarrow{X} | \overleftarrow{X} = \overleftarrow{x}) = \Pr(\overrightarrow{X} | \overleftarrow{X} = \overleftarrow{x}')$ —that is, if the two pasts are equivalent from the standpoint of prediction. The corresponding clusters are causal states σ , which are realizations of the random variable for the causal states \mathcal{S} . The statistical complexity C_μ is simply their coding cost, $C_\mu = H[\mathcal{S}]$. In short, causal states \mathcal{S} are minimal sufficient statistics of prediction; the statistical complexity $C_\mu = H[\mathcal{S}]$ is the coding cost of those causal states [15]; and the ϵ -machine is the minimal maximally predictive model constructed from those causal states [16].

The same constructions apply when considering continuous-time, discrete-event processes. In that case, the equivalence relation \sim_ϵ clusters two semi-infinite pasts $\overleftarrow{(x, \tau)}$, $\overleftarrow{(x, \tau)'}$ together if $\Pr(\overrightarrow{(X, T)} | \overleftarrow{(X, T)} = \overleftarrow{(x, \tau)}) = \Pr(\overrightarrow{(X, T)} | \overleftarrow{(X, T)} = \overleftarrow{(x, \tau)'})$. Again, the clusters are causal states σ , realizations of the random variable for causal states \mathcal{S} .

For what follows, we must define two terms: mixed state simplex, and unifilar. Consider any hidden Markov model whose hidden state at time t is a random variable; now consider the probability distribution over hidden states given observations \overleftarrow{x} . Each one of these conditional probability

distributions is a mixed state, and it lies in the mixed state simplex, the set of all possible probability distributions over hidden states. The hidden Markov model is unifilar when, given one's hidden state at time t and an observation at time t , one knows exactly which hidden state comes next at $t + 1$. There is a connection between unifilarity and the mixed state simplex: when the hidden Markov model under study is unifilar, then the mixed states will lie at the edge of the simplex. (This is not true for nonunifilar hidden Markov models.) The causal states are just the mixed states of the minimal (potentially nonunifilar) generative model.

The causal states of discrete-time processes are usually uncountably infinite. When this is the case, then the box-counting dimension of the mixed state presentation in the mixed state simplex is nonzero. Let's unpack this statement. Suppose that a (potentially nonunifilar) Hidden Markov model with states g generates the observed discrete-time process. Then we use $p(g|\overleftarrow{x})$ to denote the probability over hidden states in the generative model given past output. Typically, $p(g|\overleftarrow{x})$ is in the interior of the mixed state simplex—the space of probability distributions over hidden states. The box-counting dimension of the mixed state presentation is obtained by gridding the mixed state simplex by cubes of side length ϵ , counting the number of non-empty cubes N_ϵ (which contain coarse-grainings of histories) [17], and then calculating the scaling of N_ϵ with ϵ —so the box-counting dimension is $\lim_{\epsilon \rightarrow 0} \frac{\log N_\epsilon}{\log \frac{1}{\epsilon}}$. A cube is considered non-empty when there is at least one history that leads to a mixed state in the cube. When there are countable causal states for a discrete-time process, the box-counting dimension is 0.

The causal states inherit a dynamic, and the ϵ -machine of a process is the pairing of causal states together with that dynamic. For discrete-time, discrete-event processes, tractable ϵ -machines are merely countable unifilar Hidden Markov models [16], where unifilarity implies that the next hidden state is determined uniquely by the previous hidden state and the present emitted symbol. For continuous-time, discrete-event processes, tractable ϵ -machines can (for instance) take the form of joined conveyor belts [18]. Continuous-time causal states are then usually accompanied by labeled transition operators $\mathcal{O}^{(x)}$, and the list of labeled transition operators specifies the continuous-time ϵ -machine. A “tractable ϵ -machine” is one for which C_μ is finite, with the exception of the ϵ -machines of continuous-time periodic processes (which are tractable but which correspond to infinite C_μ).

2.2. Monod-Wyman-Changeux Molecules

A Monod-Wyman-Changeux (MWC) molecule has two configurations, active (A) and inactive (I), and n binding sites. Each configuration can bind any number of molecules, from 0 to n . This gives a total of 2^{n+1} possible states. If binding sites are indistinguishable, then a simplified model can be made based on the symmetry in binding sites so that there are only $2(n + 1)$ distinguishable possible states: it can be either active or inactive, with any number of binding sites occupied by ligand molecules. As our argument holds for any n , we focus on the case that $n = 1$. The four states of the corresponding MWC molecule— $\{A_0, A_1, I_0, I_1\}$, standing for active/inactive (A/I) with either 0 or 1 ligands bound as written in the subscript—are shown in Figure 1, along with allowed transitions.

We denote the probability distribution of being in various states as

$$\vec{p} = \begin{pmatrix} p(A_0) \\ p(A_1) \\ p(I_0) \\ p(I_1) \end{pmatrix}. \quad (1)$$

This probability distribution evolves via the master equation

$$\frac{d\vec{p}}{dt} = M(c(t))\vec{p} \quad (2)$$

where

$$M(c) = \begin{pmatrix} -(f_T + f_{Ac}) & b_A & b_T & 0 \\ f_{Ac} & -(f'_T + b_A) & 0 & b'_T \\ f_T & 0 & -(b_T + f_{Ic}) & b_I \\ 0 & f'_T & f_{Ic} & -(b'_T + b_I) \end{pmatrix} \quad (3)$$

with $f_T, f_A, b_A, b_T, f'_T, b'_T$ are kinetic parameters. In fixed ligand concentration c , Equation (2) is solved as

$$\vec{p}(t) = e^{M(c)t} \vec{p}(0) \quad (4)$$

where $\vec{p}(0)$ is the initial probability distribution over the MWC molecule's states.

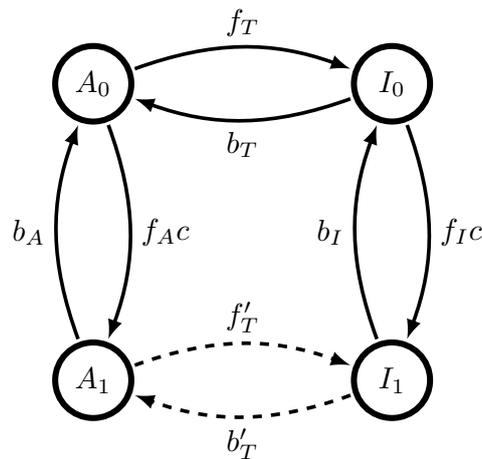


Figure 1. A dynamical single-site Monod-Wyman-Changeux molecule, with kinetic rates as shown. States marked A_i are active with i bound ligand molecules, while states marked I_i are inactive with i bound ligand molecules. When transitioning from states A_0, A_1 , A is emitted, while when transitioning from states I_0, I_1 , I is emitted.

3. Results

We suppose that we are only allowed to see whether or not the MWC molecule is active or inactive, as would be true for most experimental observations of ligand-gated ion channels. This is a key constraint, as otherwise, the minimal generative model would be the minimal maximally predictive model and none of the discrepancies described here would arise. In what follows, we explore the effects of kinetic rates on intrinsic and functional computation.

Intrinsic computation as it was originally defined included the ϵ -machine and statistical complexity, and today includes other information measures, such as the excess entropy [10–12]. The number of functional computation-related quantities is unbounded, but we focus on two here due to their presence in the literature: the binding curve, or the probability of the MWC molecule being active as a function of ligand concentration; and the transfer function, or how the MWC molecule responds to sinusoidal perturbations of the ligand concentration.

Our argument will essentially be a proof by contradiction. We will start by assuming that there is some relationship between at least one aspect of intrinsic computation and at least one aspect of functional computation. If there were a relationship between these two quantities, then we should not be able to change kinetic rates so that one quantity changes by an arbitrarily small amount and the other by an arbitrarily large amount. (If so, these kinetic rates would then be of incredible importance to the process' causal architecture, say, but of vanishingly small importance to the so-called functional computations.) We will then show in the following analysis that arbitrarily small perturbations in the kinetic rates f'_T, b'_T induce arbitrarily large perturbations to the ϵ -machine and statistical complexity, but induce arbitrarily small perturbations to excess entropy and the functional

computations considered here. We therefore conclude that if there is a relationship between intrinsic computation and functional computation for these kinds of molecules, it will more likely come from excess entropy (or other more recently-studied information measures of time series [19]) than from statistical complexity or the ϵ -machine. We discuss the possibilities of finding functional computations that are sensitive to arbitrarily small increases in f'_T, b'_T in Section 4.

3.1. Intrinsic Computation

If $b'_T, f'_T > 0$, then there is no “sync word”—that is, no string of observed past symbols that uniquely determines the underlying present state of the MWC molecule. (Note again that this would not be the case if we were allowed to observe the full state, and not just whether or not the MWC molecule is active or inactive.) This has important consequences for C_μ and for the process’ ϵ -machine.

To analyze C_μ , we move to the discrete-time domain, so as to avoid the interpretational difficulties with differential entropy [18]. By observing the process every Δt for Δt much smaller than any inherent time constant in the problem, the process is turned into a discrete-time process. The transition probabilities of this new process have corresponding labeled transition matrices $T^{(x)} = e^{M^{(x)}\Delta t}$, which are approximated to lowest order in Δt by $T^{(x)} = I^{(x)} + M^{(x)}\Delta t$, where x is an emitted symbol:

$$M^{(A)}(c) = \begin{pmatrix} -(f_T + f_{Ac}) & b_A & 0 & 0 \\ f_{Ac} & -(b_A + f'_T) & 0 & 0 \\ f_T & 0 & 0 & 0 \\ 0 & f'_T & 0 & 0 \end{pmatrix} \text{ and}$$

$$M^{(I)}(c) = \begin{pmatrix} 0 & 0 & b_T & 0 \\ 0 & 0 & 0 & b'_T \\ 0 & 0 & -(b_T + f_{Ic}) & b_I \\ 0 & 0 & f_{Ic} & -(b_I + b'_T) \end{pmatrix}$$

and

$$I^{(A)} = \begin{pmatrix} I_{2 \times 2} & 0_{2 \times 2} \\ 0_{2 \times 2} & 0_{2 \times 2} \end{pmatrix}$$

$$I^{(I)} = \begin{pmatrix} 0_{2 \times 2} & 0_{2 \times 2} \\ 0_{2 \times 2} & I_{2 \times 2} \end{pmatrix}.$$

The causal states correspond to the mixed states $\Pr(\mathcal{S} | \overleftarrow{X} = \overleftarrow{x})$ defined by either $(a, 1 - a, 0, 0)$ or $(0, 0, 1 - b, b)$. When $b'_T, f'_T > 0$, all a 's and b 's are allowed; there are an uncountable infinity of these causal states because there is no sync word. As such, the ϵ -machine is uncountable and intractable, and statistical complexity is very likely infinite. However, when $b'_T = f'_T = 0$, then only a countable set of a 's and b 's are allowed according to the discrete-time analogue of Theorem 1 of Ref. [18]. In particular, the causal states are identified as both the present configuration (active or inactive) and the number of time steps since last configuration switch. As a result, the ϵ -machine is countably infinite and tractable, and statistical complexity is finite, though it increases with $\log \frac{1}{\Delta t}$ [20].

This can be seen more directly by considering the mixed-state presentation’s box-counting dimension h_0 when the process is turned into a discrete-time process with small time resolution $\Delta t = 0.01$. We coarse-grain mixed-state simplex into cubes of side-length ϵ , and count the number of non-empty boxes N_ϵ , as described in Ref. [17,21]. The scaling of N_ϵ with ϵ reveals the box-counting dimension h_0 of the mixed-state presentation, via $N_\epsilon \sim (1/\epsilon)^{h_0}$. Figure 2 shows the scaling of N_ϵ with ϵ for an MWC molecule with and without $f'_T = b'_T = 0$. When $f'_T = b'_T = 0, h_0 = 0$; when $f'_T, b'_T > 0$, the box-counting dimension $h_0 > 0$.

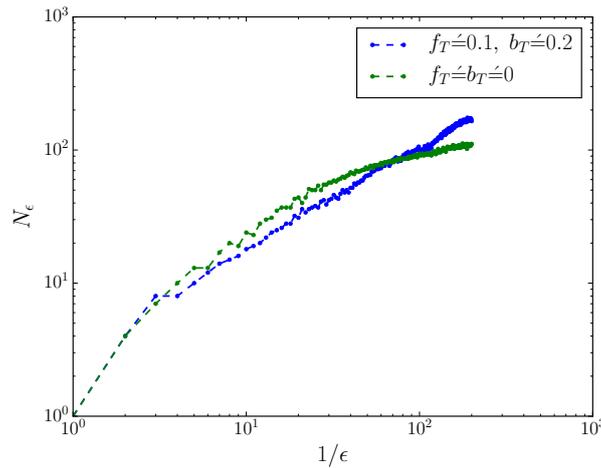


Figure 2. Box-counting dimension of the mixed-state presentation changes drastically with f'_T, b'_T . For both processes, we have: $f_T = 1.0, f_{Ac} = 2.9, b_A = 3.4, b_T = 3, f_{Ic} = 4, b_I = 2$. The process with nonzero f'_T, b'_T has a scaling of $\log N_\epsilon \sim \log(1/\epsilon)$ and thus a nonzero box-counting dimension $h_0 > 0$, whereas the process with $f'_T = b'_T = 0$ has a scaling of $\log N_\epsilon \sim \log \log(1/\epsilon)$ and thus a box-counting dimension $h_0 = 0$.

The reason for the former fact lies in Theorem 1 of Ref. [18]. When $f'_T = b'_T = 0$, the dynamic MWC molecule of Figure 1 generates a semi-Markov process, a restricted version of the unifilar hidden semi-Markov processes analyzed in Ref. [18]. This is true even when there is more than one ligand binding site. Causal states are characterized by x , whether or not the MWC molecule is presently active, and τ_+ , the time since the MWC molecule last switched between activities. The now-tractable ϵ -machine takes the form shown in Figure 3.

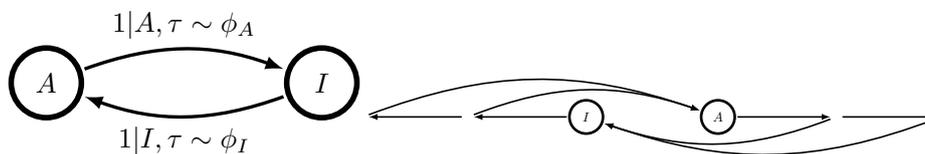


Figure 3. At left, a generative model of the process generated by the MWC molecule in fixed ligand concentration c of Figure 1 with $f'_T = b'_T = 0$. The dwell time distributions $\phi_A(t)$ and $\phi_I(t)$ are given in Equations (8) and (9). At right, the corresponding topological ϵ -machine. While emitting A , one moves along the “conveyor belt” starting with state A to the left; while emitting I , one moves along the conveyor belt starting with state I to the right. To switch the letter that one is emitting, one jumps to the other conveyor belt. The states along the conveyor belt to the left correspond to the time that one has been inactive, and the states along the conveyor belt to the right correspond to the time that one has been active.

In the continuous-time limit, which one can derive by considering the limit of the discrete-time process considered above with appropriate renormalization (e.g., compare Ref. [20] to Ref. [22]), all probability distributions over mixed states become probability density functions. Continuous-time statistical complexity can be defined using the entropy of mixed random variables [18,22], though differential entropy does have some troubling properties mentioned in those references. Given the analysis above, there is likely a singular limit in the continuous-time statistical complexity as the kinetic rates b'_T, f'_T tend to 0.

Not all structure-based characterizations of a process lack robustness in this way, as different structure-based metrics pick up on different kinds of structure. To show this, we now compare the statistical complexity C_μ and the excess entropy $\mathbf{E} = I[(X, \mathcal{T}); (X, \mathcal{T})]$ [10–12] when $f'_T = b'_T = 0$. (We calculate \mathbf{E} of the continuous-time process, as the excess entropy of the discrete-time process

converges to that of the continuous-time process in the $\Delta t \rightarrow 0$ limit [20].) The latter can be calculated via $\mathbf{E} = I[\mathcal{S}^+; \mathcal{S}^-]$ [23,24], while $C_\mu = H[\mathcal{S}^+]$, and so calculation of both merely requires the joint distribution $p(\sigma^+, \sigma^-)$. For that, we need $\phi_{A/I}(t)$, the dwell time distributions of activity and inactivity. Note that emission of an A implies that one has just landed in A_0 , and similarly, emission of an I implies that one has just landed in I_0 . Hence, $\phi_A(t)$ is the first-passage time distribution to state I_0 in which one starts in A_0 ; similarly, $\phi_I(t)$ is the first-passage time distribution to state A_0 in which one starts in I_0 . To aid with the calculation, we recall the labeled transition matrices of Equation (5) when $f'_T = b'_T = 0$. The matrix $M^{(A)}$ includes only the transitions between various active conformations and the only transition from active to inactive, $A_0 \rightarrow I_0$. Therefore, the probability of not having stayed in active states given that one started in A_0 after a time t is given by

$$1 - \Phi_A(t) = (\hat{e}_3 + \hat{e}_4)^\top \vec{p}(t), \quad \frac{d\vec{p}}{dt} = M^{(A)}(c)\vec{p}(t), \quad \vec{p}(0) = \hat{e}_1, \tag{5}$$

where \hat{e}_k is the vector with elements $\delta_{i,k}$. Hence, the survival function $\Phi_A(\tau) = \int_\tau^\infty \phi_A(\tau')d\tau'$, the probability that one stays in the active conformation (one of A_0, A_1) after time t given that one started in A_0 , can be calculated via

$$\Phi_A(t) = 1 - (\hat{e}_3 + \hat{e}_4)^\top e^{M^{(A)}(c)t}\hat{e}_1. \tag{6}$$

Similarly,

$$\Phi_I(t) = 1 - (\hat{e}_1 + \hat{e}_2)^\top e^{M^{(I)}(c)t}\hat{e}_3. \tag{7}$$

After differentiation, we find that

$$\begin{aligned} \phi_A(t) &= -\frac{d\Phi_A(t)}{dt} \\ &= (\hat{e}_3 + \hat{e}_4)^\top M^{(A)}(c)e^{M^{(A)}(c)t}\hat{e}_1 \end{aligned} \tag{8}$$

$$\begin{aligned} \phi_I(t) &= -\frac{d\Phi_I(t)}{dt} \\ &= (\hat{e}_1 + \hat{e}_2)^\top M^{(I)}(c)e^{M^{(I)}(c)t}\hat{e}_3. \end{aligned} \tag{9}$$

Examples of $\phi_A(t)$ for various ligand concentrations c and kinetic rates are shown in Figure 4.

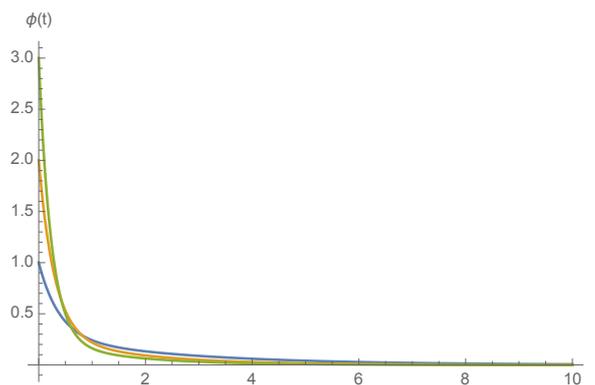


Figure 4. $\phi_A(t)$ for $f_A = b_A = 1.0$ and $f_T = 1.0$ (blue), $f_T = 2.0$ (orange), and $f_T = 3.0$ (green), calculated using Equation (8).

From Lemma 1 of Ref. [18], we find that the statistical complexity of this semi-Markov process is given by

$$C_\mu = H_b(p(A)) - \sum_{x \in \{A,I\}} p(x) \int_0^\infty (\mu_x \Phi_x(\tau)) \log(\mu_x \Phi_x(\tau)) d\tau, \tag{10}$$

where $p(A) = \frac{\mu_I}{\mu_A + \mu_I}$, $\mu_x = 1 / \int_0^\infty \Phi_x(\tau) d\tau$, and $H_b(x) := -x \log x - (1 - x) \log(1 - x)$. Figure 5 shows how C_μ smoothly varies with changes in f_T, b_T for $f_T, b_T > 0$. Statistical complexity is maximized at small kinetic rates, $f_T, b_T \rightarrow 0$; when those kinetic rates are small, dwell time distributions have longer tails, and the memory required to losslessly predict increases. Interestingly, if either f_T or b_T is exactly 0, then the generated process emits only A or I (see Figure 1) and thus has $C_\mu = 0$. In other words, the limits $f_T, b_T \rightarrow 0$ are singular, as were the limits $f'_T, b'_T \rightarrow 0$.

We now wish to calculate $\mathbf{E} = I[(X, \mathcal{T}); (X, \overline{\mathcal{T}})]$ which is $\mathbf{E} = I[\mathcal{S}^+; \mathcal{S}^-]$ [23], and so

$$\mathbf{E} = H[\mathcal{S}^-] - H[\mathcal{S}^- | \mathcal{S}^+]. \tag{11}$$

As a semi-Markov process is causally reversible, we have

$$H[\mathcal{S}^-] = H[\mathcal{S}^+] = C_\mu \tag{12}$$

as given in Equation (10). Furthermore, the reverse-time causal states are the pair (x_-, τ_-) (the time to next symbol and present symbol) while the forward-time causal states are still the pair (x_+, τ_+) (the time since last symbol and present symbol) [18], so that $x_+ = x_-$ almost surely, implying that $H[X_- | X_+] = 0$. Hence,

$$\begin{aligned} H[\mathcal{S}^- | \mathcal{S}^+] &= H[X_-, \mathcal{T}_- | X_+, \mathcal{T}_+] \\ &= H[X_- | X_+, \mathcal{T}_+] + H[\mathcal{T}_- | X_-, X_+, \mathcal{T}_+] \\ &= H[\mathcal{T}_- | X_0, \mathcal{T}_+] \end{aligned} \tag{13}$$

where x_0 is just the present symbol. We then note that

$$p(\tau_- | x_0, \tau_+) = \frac{\phi_{x_0}(\tau_+ + \tau_-)}{\Phi_{x_0}(\tau_+)}, \tag{14}$$

as was derived in Ref. [22] for a continuous-time renewal process, but the same derivation holds for the semi-Markov process. It is then straightforward to show that excess entropy is

$$\mathbf{E} = H[X] + \sum_x p(x) \mathbf{E}[\phi_x(t)] \tag{15}$$

where

$$\mathbf{E}[\phi_x(t)] = \int_0^\infty \int_0^\infty \mu_x \phi_x(t + t') \log \frac{\phi_x(t + t')}{\mu_x \Phi_x(t) \Phi_x(t')} dt dt'. \tag{16}$$

Figure 5b shows how excess entropy \mathbf{E} varies with f_T, b_T . Interestingly, \mathbf{E} varies in opposition to C_μ , attaining its lowest values at low values of f_T and b_T . Hence, the singular limits $f_T, b_T \rightarrow 0$ that plague C_μ are not singular limits for \mathbf{E} . Nor are $f'_T, b'_T \rightarrow 0$ singular limits for \mathbf{E} , as arbitrarily small values of f'_T, b'_T lead to arbitrarily small perturbations to the trajectory distribution, and thus arbitrarily small perturbations to the mutual information between past and future.

How different would the results be for $n > 1$, i.e., when the number of binding sites of the MWC molecule exceeded 1? In short, we would expect the same qualitative trends and singular limits. In this more general case, we allow an active MWC molecule with k ligands bound to transition to an inactive MWC molecule with k ligands bound, and vice versa, both with rates f'_T and b'_T , as for $n = 1$. Meanwhile, also as for $n = 1$, the active MWC molecule with no ligands bound can transition to the inactive MWC molecule with no ligands bound with rate f_T , and the reverse transition can occur with rate b_T . The observed process for fixed ligand concentration would still be semi-Markov when $f'_T = b'_T = 0$, as was true for $n = 1$. Then, decreases in f_T, b_T would lead to longer dwell times in active and inactive states, thereby increasing the statistical complexity C_μ ; and the dwell time distributions would become closer to exponential, decreasing the excess entropy \mathbf{E} . When f'_T, b'_T become small but

nonzero, all pasts are causal states, and so C_μ shoots to infinity, while \mathbf{E} (because it is a function of trajectory distributions) barely changes.

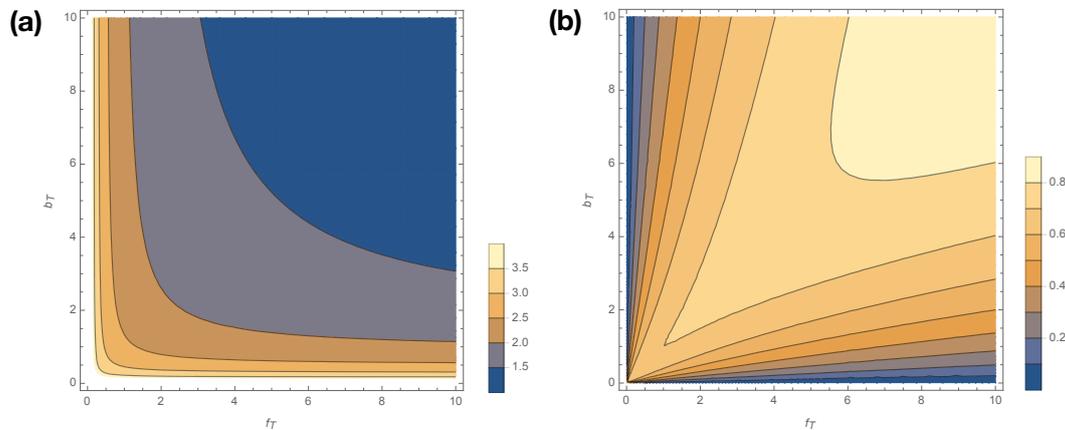


Figure 5. Contour plot of C_μ (a) and \mathbf{E} (b) as a function of f_T , b_T when $f'_T = b'_T = 0$, $f_A c = f_I c = b_A = b_I = 1$.

There are some well-known examples of how arbitrarily large ϵ -machines can still have arbitrarily small excess entropies, e.g., the almost fair coin. Indeed, Ref. [23] defined crypticity as the difference between statistical complexity C_μ and excess entropy \mathbf{E} . The dynamical MWC molecule described above adds another such example to the literature, finding not only that a familiar process can have arbitrarily large crypticity, but that C_μ and \mathbf{E} can be anti-correlated with respect to underlying kinetic rates, as is true for the process generated by the parametrized Simple Nonunifilar Source [25]. There are also examples in the literature of processes with uncountable ϵ -machines and nonzero box-counting dimensions of their mixed-state presentation, e.g., the Cantor process in Ref. [26].

However, the dynamical MWC molecule is more than just an example of a process with potentially arbitrarily large crypticity or an uncountable ϵ -machine; it is also an example of how arbitrarily small changes to a generative model can lead to arbitrarily large changes in the causal structure of a process. Of course, it may be obvious to those familiar with intrinsic computation that sometimes, arbitrarily small perturbations in transition probabilities of a generative model can lead to arbitrarily large perturbations in ϵ -machine structure. However, to the author’s knowledge, the above MWC molecule example is the first such example in the literature.

3.2. Functional Computation

Monod-Wyman-Changeux (MWC) molecules have been used to model everything from ligand-gated ion channels to gene regulation [8]. The functional computations that an MWC molecule is thought to perform include transduction of ligand concentration and low-pass filtering of input [7].

Let $\text{eig}_0(M(c))$ be the normalized eigenvector of eigenvalue 0 of matrix $M(c)$, normalized so that $1^\top \text{eig}_0(M(c)) = 1$; and let $p_{eq,A}(c)$ be the equilibrium probability of being in state A . The MWC molecule’s ability to convey the ligand concentration via its activity is a static property, relying only on how the equilibrium distribution

$$\begin{aligned}
 p_{eq,A}(c) &= p_{eq,A_0}(c) + p_{eq,A_1}(c) \\
 &= (\hat{e}_1 + \hat{e}_2)^\top \text{eig}_0(M(c))
 \end{aligned}
 \tag{17}$$

varies with kinetic rates. An observer that can only see whether or not the MWC molecule is active can discern, to some extent, the external ligand concentration c . Such a situation might occur, for instance, for the nicotinic acetylcholine receptors at the neuromuscular junction that transduce information about whether or not a muscle fiber should seize, based on acetylcholine concentration. Though

$\text{eig}_0(M(c))$ in principle might be a non-smoothly varying function of f'_T, b'_T , a Mathematica calculation finds that $\text{eig}_0(M(c))$ and thus $p_{eq,A}(c)$ varies smoothly with kinetic rates f'_T, b'_T :

$$\text{eig}_0(M(c)) \propto \begin{pmatrix} b_A b_I b_T + b_A b_T b'_T + b_A b'_T f_I c + b_I b_T f'_T \\ (b_I b_T f_A + b_T b'_T f_A + b'_T f_A f_I c + b'_T f_T f_I) c \\ b_A b_I f_T + b_A f_T b'_T + f'_T b_I f_A c + b_I f_T f'_T \\ (b_A f_T f_I + b_T f_A f'_T + f'_T f_I f_A c + f'_T f_T f_I) c \end{pmatrix},$$

where the normalization constant is chosen so that $1^\top \text{eig}_0(M(c)) = 1$. The smooth variation of $p_{eq,A}(c)$ with respect to f'_T, b'_T is depicted for a random choice of kinetic rates in Figure 6.

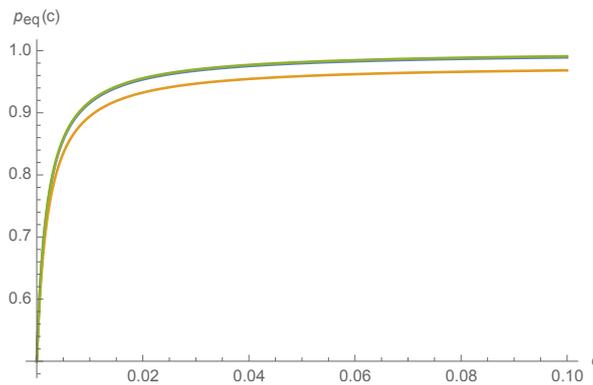


Figure 6. Probability of being in the active state, $p_{eq,A}(c)$, as a function of ligand concentration c , for $f_T = 1, f_A = 100, b_A = 0.1$ and $b_T = 1, f_I = 1, b_I = 1$, and: $f'_T = b'_T = 0$ (blue); $f'_T = 0.01$ and $b'_T = 0$ (orange); and $f'_T = 0$ and $b'_T = 10$ (green), almost overlaying the blue.

The MWC molecule is a low-pass filter of ligand concentration c . Suppose that $c(t) = c_0 + \delta c \sin \omega t$, where δc is small. Then $p_{eq,A}(t)$ will also take the form $p_{eq,A} = p_{eq,A}(c_0) + G(\omega) \delta c \sin \omega t + O(\delta c^2)$, where $G(\omega)$ is the transfer function. This transfer function therefore characterizes the dynamical response of the MWC molecule to fluctuations in the ligand concentration. From Equation 49 of Ref. [8], we find that the transfer function $G(\omega)$ is

$$G(\omega) = (\hat{e}_1 + \hat{e}_2)^\top (i\omega I - M_0)^{-1} M_1 \text{eig}_0(M(c_0)) \tag{18}$$

where

$$M_0 = \begin{pmatrix} -f_T & b_A & b_T & 0 \\ 0 & -(f'_T + b_A) & 0 & b'_T \\ f_T & 0 & -b_T & b_I \\ 0 & f'_T & 0 & -(b'_T + b_I) \end{pmatrix} \text{ and}$$

$$M_1 = \begin{pmatrix} -f_A & 0 & 0 & 0 \\ f_A & 0 & 0 & 0 \\ 0 & 0 & -f_I & 0 \\ 0 & 0 & f_I & 0 \end{pmatrix}.$$

A series expansion not shown here confirms that $G(\omega)$ varies smoothly with kinetic rates f'_T, b'_T , as would be expected from the realization that all expressions in Equation (18) are smoothly varying with f'_T, b'_T . To illustrate this, the magnitude of the transfer function, $|G(\omega)|$, is plotted in Figure 7 for a randomly chosen initial concentration of $c_0 = 1$.

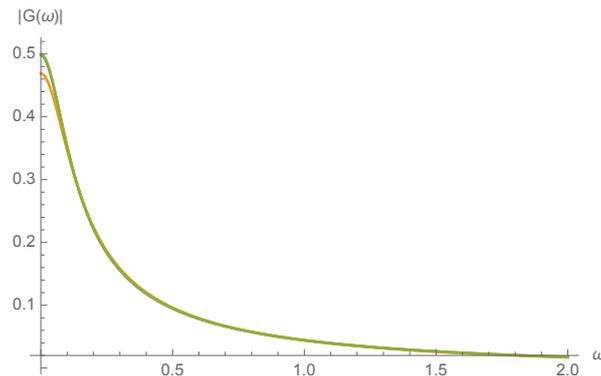


Figure 7. Transfer function $G(\omega)$ as a function of input frequency ω at randomly chosen initial concentration c_0 for $f_T = 1$, $f_A = 100$, $b_A = 0.1$ and $b_T = 1$, $f_I = 1$, $b_I = 1$, and: $f'_T = b'_T = 0$ (blue); $f'_T = 0.01$ and $b'_T = 0$ (orange); and $f'_T = 0$ and $b'_T = 10$ (green), almost overlaying the blue.

Again, it is worth commenting on how these results would vary with larger n , i.e., a larger number of potential ligands bound to the MWC molecule. We consider the dynamical model for this more complex MWC molecule as specified in Section 3.1. Just as for the case when $n = 1$, the eigenvector of eigenvalue 0 for this larger MWC molecule's rate matrix is a continuous function of kinetic rates f_T , b_T and f'_T , b'_T ; as a result, both the binding curve and the transfer function vary smoothly with these rates.

4. Discussion

Our overarching aim here was to study the link between intrinsic computation and functional computation by focusing on a popular model of a biological sensor—the Monod-Wyman-Changeux (MWC) molecule. While studying its intrinsic computation, we found interesting singular limits for C_μ . In particular, we found that statistical complexity was infinite and that all pasts were causal states when two of the kinetic rates were nonzero, $f'_T, b'_T \neq 0$, no matter how small f'_T, b'_T ; and we found that statistical complexity was zero when $f_T = b_T = 0$ but nonzero and arbitrarily large for arbitrarily small $f_T, b_T > 0$. While studying the MWC molecule's functional computation and its process' excess entropy E [10–12], we found no such singular limits with respect to these kinetic rates.

The reason for this is that the studied functional computations and excess entropy is a continuous function of trajectory distributions alone, while statistical complexity must be written in terms of a distribution of causal states, which can vary in a non-continuous manner with the trajectory distribution. As a result, statistical complexity (and the mixed state presentation's box-counting dimension h_0) and the ϵ -machine are incredibly sensitive to a particular type of process structure, which includes but is not limited to forbidden words. On the other hand, the studied functional computations and excess entropy are smoothly varying functions of the generative model's kinetic rates.

From the study of the MWC molecule alone, we can conclude that a restrictive definition of intrinsic computation, e.g., only causal structure, does not necessarily provide a guide to the functional computation of a dynamical system, at least for the functional computations considered here. Accordingly, high statistical complexity does not imply biological function.

It's well worth emphasizing that the functional computations listed here are far from an exhaustive list of all possible functional computations, and so future research might uncover a functional computation that depends sensitively on the process' causal structure. Also, even if no such functional computation is identified, the sensitivity of causal structure to certain changes in the generative model might be considered by some to be an interesting feature, and not a bug, perhaps as a case study in how limited computational resources yield innovation [1]. However, for those wishing to study functional computation only, the extreme sensitivity of statistical complexity to particular types of process structure might prove to be a bug rather than a feature.

However, even then, the ϵ -machine finds use. In more recent years, intrinsic computation has been expansively defined to include a study of other structure-related information-theoretic statistics of a process besides statistical complexity [21]. This list includes but is not limited to excess entropy [10–12] as studied here, bound information rate [27], and predictive rate-distortion functions [28,29]. The last is particularly notable here in that predictive rate-distortion includes statistical complexity and excess entropy as limiting cases. On the whole, these quantities enjoy the label “information anatomy” [19] or the more broadly-construed “informational architecture”. Most of these quantities are smoothly varying functions of the transition probabilities in the minimal generative model and thus trajectory distribution, and so are not as sensitive to the underlying structure of the process as is statistical complexity. (The statistical complexity can vary discontinuously with transition probabilities in the minimal generative model, but varies smoothly with the transition probabilities in the minimal maximally predictive model; the structure of the minimal maximally predictive model, the ϵ -machine, can change by an arbitrarily large amount with arbitrarily small changes to the minimal generative model.) Those that are not so sensitive to the process’ structure are often easily calculable from the process’ ϵ -machine [29,30]. In the future, these quantities might provide interesting statistics with which to interpret the functional computation performed by biological or social systems, e.g., as in Ref. [31].

Funding: The author acknowledges generous funding from the U.C. Berkeley Chancellor’s Fellowship, the National Science Foundation Graduate Research Fellowship, and the MIT Physics of Living Systems Fellowship.

Acknowledgments: The author would like to thank James P. Crutchfield for clarifying discussions, as well as two anonymous referees.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Crutchfield, J.P. The calculi of emergence: Computation, dynamics, and induction. *Phys. D Nonlinear Phenom.* **1994**, *75*, 11–54. [[CrossRef](#)]
2. Pearl, J. *Causality*; Cambridge University Press: Cambridge, UK, 2009.
3. White, O.L.; Lee, D.D.; Sompolinsky, H. Short-term memory in orthogonal neural networks. *Phys. Rev. Lett.* **2004**, *92*, 148102. [[CrossRef](#)] [[PubMed](#)]
4. Ganguli, S.; Huh, D.; Sompolinsky, H. Memory traces in dynamical systems. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 18970–18975. [[CrossRef](#)] [[PubMed](#)]
5. Creutzig, F.; Globerson, A.; Tishby, N. Past-future information bottleneck in dynamical systems. *Phys. Rev. E* **2009**, *79*, 041925. [[CrossRef](#)] [[PubMed](#)]
6. McCulloch, W.S.; Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [[CrossRef](#)]
7. Martins, B.M.; Swain, P.S. Trade-offs and constraints in allosteric sensing. *PLoS Comput. Biol.* **2011**, *7*, e1002261. [[CrossRef](#)] [[PubMed](#)]
8. Marzen, S.; Garcia, H.G.; Phillips, R. Statistical mechanics of Monod–Wyman–Changeux (MWC) models. *J. Mol. Biol.* **2013**, *425*, 1433–1460. [[CrossRef](#)] [[PubMed](#)]
9. Changeux, J.P. 50 years of allosteric interactions: the twists and turns of the models. *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 819–829. [[CrossRef](#)] [[PubMed](#)]
10. Grassberger, P. Toward a quantitative theory of self-generated complexity. *Int. J. Theor. Phys.* **1986**, *25*, 907–938. [[CrossRef](#)]
11. Bialek, W.; Nemenman, I.; Tishby, N. Predictability, complexity, and learning. *Neural Comput.* **2001**, *13*, 2409–2463. [[CrossRef](#)] [[PubMed](#)]
12. Bialek, W.; Nemenman, I.; Tishby, N. Complexity through Nonextensivity. *Phys. A Stat. Mech. Its Appl.* **2001**, *302*, 89–99. [[CrossRef](#)]
13. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley-Interscience: New York, NY, USA, 2006.
14. Nair, C.; Prabhakar, B.; Shah, D. On entropy for mixtures of discrete and continuous variables. *arXiv* **2006**, arXiv:cs/0607075.

15. Crutchfield, J.P.; Young, K. Inferring statistical complexity. *Phys. Rev. Lett.* **1989**, *63*, 105–108. [[CrossRef](#)] [[PubMed](#)]
16. Shalizi, C.R.; Crutchfield, J.P. Computational mechanics: Pattern and prediction, structure and simplicity. *J. Stat. Phys.* **2001**, *104*, 817–879. [[CrossRef](#)]
17. Marzen, S.E.; Crutchfield, J.P. Nearly maximally predictive features and their dimensions. *Phys. Rev. E* **2017**, *95*, 051301. [[CrossRef](#)] [[PubMed](#)]
18. Marzen, S.E.; Crutchfield, J.P. Structure and randomness of continuous-time, discrete-event processes. *J. Stat. Phys.* **2017**, *169*, 303–315. [[CrossRef](#)]
19. James, R.G.; Ellison, C.J.; Crutchfield, J.P. Anatomy of a bit: Information in a Time series observation. *Chaos Interdiscip. J. Nonlinear Sci.* **2011**, *21*, 037109. [[CrossRef](#)] [[PubMed](#)]
20. Marzen, S.; DeWeese, M.R.; Crutchfield, J.P. Time resolution dependence of information measures for spiking neurons: Scaling and universality. *Front. Comput. Neurosci.* **2015**, *9*, 105. [[CrossRef](#)] [[PubMed](#)]
21. Crutchfield, J.P. Between order and chaos. *Nat. Phys.* **2012**, *8*, 17–24. [[CrossRef](#)]
22. Marzen, S.; Crutchfield, J.P. Informational and causal architecture of continuous-time renewal processes. *J. Stat. Phys.* **2017**, *168*, 109–127. [[CrossRef](#)]
23. Crutchfield, J.P.; Ellison, C.J.; Mahoney, J.R. Time's barbed arrow: Irreversibility, crypticity, and stored information. *Phys. Rev. Lett.* **2009**, *103*, 094101. [[CrossRef](#)] [[PubMed](#)]
24. Ellison, C.J.; Mahoney, J.R.; Crutchfield, J.P. Prediction, retrodiction, and the amount of information stored in the present. *J. Stat. Phys.* **2009**, *136*, 1005–1034. [[CrossRef](#)]
25. Marzen, S.; Crutchfield, J.P. Informational and causal architecture of discrete-time renewal processes. *Entropy* **2015**, *17*, 4891–4917. [[CrossRef](#)]
26. Upper, D.R. Theory and Algorithms for Hidden Markov Models and Generalized Hidden Markov Models. Ph.D. Thesis, University of California, Berkeley, CA, USA, 1997.
27. Abdallah, S.A.; Plumbley, M.D. A measure of statistical complexity based on predictive information. *arXiv* **2010**, arXiv:1012.1890.
28. Still, S.; Crutchfield, J.P.; Ellison, C.J. Optimal causal inference: Estimating stored information and approximating causal architecture. *Chaos Interdiscip. J. Nonlinear Sci.* **2010**, *20*, 037111. [[CrossRef](#)] [[PubMed](#)]
29. Marzen, S.E.; Crutchfield, J.P. Predictive rate-distortion for infinite-order Markov processes. *J. Stat. Phys.* **2016**, *163*, 1312–1338. [[CrossRef](#)]
30. Crutchfield, J.P.; Ellison, C.J.; Riechers, P.M. Exact complexity: The spectral decomposition of intrinsic computation. *Phys. Lett. A* **2016**, *380*, 998–1002. [[CrossRef](#)]
31. Palmer, S.E.; Marre, O.; Berry, M.J., II; Bialek, W. Predictive information in a sensory population. *arXiv* **2013**, arXiv:1307.0225.



© 2018 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).