

# Thermodynamic framework for information in nanoscale systems with memory

J. Ricardo Arias-Gonzalez<sup>a)</sup>

*Instituto Madrileño de Estudios Avanzados en Nanociencia, C/Faraday 9, Cantoblanco, 28049 Madrid, Spain and CNB-CSIC-IMDEA Nanociencia Associated Unit “Unidad de Nanobiotecnología,” Cantoblanco, 28049 Madrid, Spain*

(Received 15 September 2017; accepted 7 November 2017; published online 29 November 2017)

Information is represented by linear strings of symbols with memory that carry errors as a result of their stochastic nature. Proofreading and edition are assumed to improve certainty although such processes may not be effective. Here, we develop a thermodynamic theory for material chains made up of nanoscopic subunits with symbolic meaning in the presence of memory. This framework is based on the characterization of single sequences of symbols constructed under a protocol and is used to derive the behavior of ensembles of sequences similarly constructed. We then analyze the role of proofreading and edition in the presence of memory finding conditions to make revision an effective process, namely, to decrease the entropy of the chain. Finally, we apply our formalism to DNA replication and RNA transcription finding that Watson and Crick hybridization energies with which nucleotides are branched to the template strand during the copying process are optimal to regulate the fidelity in proofreading. These results are important in applications of information theory to a variety of solid-state physical systems and other biomolecular processes. *Published by AIP Publishing.* <https://doi.org/10.1063/1.5004793>

## I. INTRODUCTION

Information comes in linear, stochastic chains with memory in biological, nanoscale processes such as replication, transcription, and translation. Molecular subunits are added directionally and on a one-by-one basis to form DNA, RNA, and proteins, respectively.<sup>1</sup> Likewise, current computers assemble information by configuring bits linearly and sequentially.<sup>2</sup> Finally, human language involves ordering symbols to compose or copy messages.<sup>3</sup> Common to all these phenomena is the presence of memory effects, which consist of interactions between the newest inserted object, either a particle or a symbol or both, and its nearest neighbors in the chain. Memory imposes mathematical correlations of an object over its previous neighbors and physically involves either positive or negative feedback of the past over the present.<sup>4</sup>

Errors are inherent to all these processes, a consequence of their stochasticity. In biology, fidelity is regulated to balance the maintenance of genetic identity and the species ability to evolve/adapt.<sup>5</sup> Many of these errors carry no consequences for the inheritance because they are filtered out in the complex, genetic information flow that goes from DNA to proteins. In language reading, we also tend to filter errors out or, if recognized, to edit them. Information technologies can also disregard a few errors by applying mathematical algorithms. Proofreading of information is a strategy to recognize errors and edition/correction mechanisms aim at reducing them by removing wrong symbols and incorporating new ones.<sup>6,7</sup>

Information theory and thermodynamics are converging to each other<sup>8</sup> in their interpretation of phenomena that involve *order* because physical entropy and information entropy may appear together in many systems like the biomolecular ones.<sup>9,10</sup> In this regard, several efforts have been made to use DNA as a practical, high-capacity, and low maintenance information storage drive.<sup>11,12</sup> Concepts like proofreading, edition, and correction, which can be comprised under the most general paradigm of revision, are becoming common across disciplines due to their intuitive meaning. Although they are assumed to improve information accuracy, they may worsen the contents of a message or, in general, they may increase the entropy of a linear, stochastic chain. In fact, several questions can be posed as follows: (i) When are errors sufficiently unimportant to be disregarded? (ii) When can proofreading identify errors? (iii) When is edition effective in substituting wrong objects by correct ones?

Information feedback has only been addressed either by Markov processes, which condition the future to only the present and discard the past, or by introducing phenomenological *Ansätze*, which are case-dependent. Strong efforts on non-Markovian dynamics have been placed in the understanding of quantum open systems,<sup>13</sup> but there is to date no thermodynamic theory that accounts exactly for the full memory of a system as it evolves reversibly, neither in classical systems nor in their quantum realm. In addition, the protocol by which the system evolves has to be considered when ensemble-average thermodynamics are addressed, especially when memory effects are present.<sup>14</sup> Such cases are common in the nanoscale, for small systems,<sup>15</sup> namely, those for which the energy exchanges are smaller or similar to the thermal level.

<sup>a)</sup>Author to whom correspondence should be addressed: ricardo.arias@imdea.org

Here, we present a formalism to analyze physical information systems in the presence of thermal fluctuations. We show its usefulness by finding conditions to determine when the revision of a linear chain that conveys information is effective. We end up applying it to nucleic acids' information transfer processes, which are biomolecular reactions in which nucleotides are incorporated sequentially by a protein whose operation determines a certain copying direction and a mechanism, both of them responsible for chain stability and information fidelity.

## II. ANALYSIS

A microstate of the information system will be specified by a sequence of values,  $\nu = \{x_1, \dots, x_i, \dots, x_n\}$ , taken by the physical subunits of the chain and represented by random variables  $X_i = x_i$  ( $x_i \in \mathcal{X}$ ,  $\mathcal{X}$  being the alphabet of the symbols or domain of the variables with cardinality  $|\mathcal{X}|$ , and  $i = 1, \dots, n$ ); see Fig. 1. The probability of a sequence is a function of its energy  $E_\nu = \sum_{i=1}^n E_i$ , which is a sum over the energies of the objects in the chain, from  $x_1$  to  $x_n$ . Due to memory effects, the energy of each object in the chain is in turn a function of the previous objects, namely,  $E_i = E(x_i; x_{i-1}, \dots, x_1)$ . These partial energy functions contain the relative interactions of every object,  $x_i$ , with its previous ones,  $\{x_{i-1}, \dots, x_1\}$ . We will treat sequences  $\nu$  as directional, stochastic chains with memory.<sup>14</sup>

The mechanism and external conditions that determine the protocol by which a system evolves are characterized by the so-called control parameter,  $\lambda$ , which may actually be a set of parameters that describe the state of the thermal bath and the constraints over the system. For a system represented by a chain, the protocol indicates, among others, the way the chain is constructed. A material one-dimensional chain may be constructed by incorporating objects on a one-by-one basis and directionally; it can also be constructed, e.g., by incorporating more than one object at a time and/or by alternating senses at each step and even by including editions, which imply removing objects. For the particular protocol of directional stepwise construction, say from left to right in Fig. 1, in material chains such as DNA, we will use  $\lambda = D$ .<sup>14</sup> Typical copying systems, either natural like DNA replication and RNA transcription or artificial like tape-based technologies, generate copies stepwisely in one sense and corrections by removing symbols in the opposite direction.

### A. Theory

We next extend the concept of thermodynamic function to individual chains that have been constructed by a microscopically reversible process. We will assign to each single-chain thermodynamic function, “A,” a sequence,  $\nu$ , and a protocol,

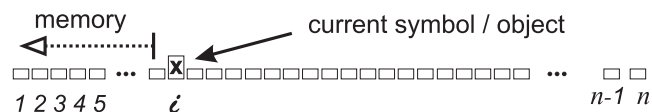


FIG. 1. Sketch of a stochastic sequence of  $n$  symbols/objects for which the memory at each position  $i$  is constituted by the  $i - 1$  previous symbols/objects.

$\lambda$ , by which that microstate has been assembled. We will use the following notation:  $A_\nu^{(\lambda)}$ .

The chains constructed under the same protocol can be treated by using expected values,

$$A^{(\lambda)} \equiv \langle A_\nu^{(\lambda)} \rangle_\lambda = \sum_{\nu=1}^N p_\nu^{(\lambda)} A_\nu^{(\lambda)}, \quad (1)$$

where  $N = |\mathcal{X}|^n$  is the number of configurations, which is the result of combining  $n$  events and  $|\mathcal{X}|$  possibilities for each event, and  $p_\nu^{(\lambda)}$  is defined according to protocol  $\lambda$  by using the corresponding sequence-dependent partition function<sup>14</sup>

$$p_\nu^{(\lambda)} = \frac{e^{-\beta E_\nu}}{Z_\nu^{(\lambda)}}, \quad (2)$$

such that  $\sum_{\nu=1}^N p_\nu^{(\lambda)} = 1$ , where  $\beta = 1/kT$ ,  $T$  being the temperature and  $k$  being the Boltzmann constant. As a directional chain,<sup>14</sup> the partition function  $Z_\nu^{(\lambda)}$  has to be evaluated according to each protocol  $\lambda$  and with respect to a certain pathway  $\nu$ , i.e., calculating individual probabilities according to the available configurations at each time step,

$$Z_\nu^{(\lambda)} \equiv \sum_{\nu'(\lambda)=1}^N \exp(-\beta E_{\nu'\nu}), \quad (3)$$

where  $E_{\nu'\nu}$  is the two-sequence energy,<sup>14</sup>

$$E_{\nu'\nu} \equiv \sum_{i=1}^n E(x'_i; x_{i-1}, \dots, x_1), \quad (4)$$

and subindex  $\nu'(\lambda)$  in the sigma symbol reminds that the sum over the multiple  $x'_i$  variables, which are correlated due to memory effects, has to be evaluated according to the constraints imposed by the protocol.

The equilibrium statistical physics is formulated by using the standard partition function, namely,

$$p_\nu = \frac{\exp(-\beta E_\nu)}{Z}, \quad Z = \sum_{\nu=1}^N \exp(-\beta E_\nu). \quad (5)$$

In the formalism given by Eq. (5), no constraints or rules have been imposed to the calculations. This formalism does not restrict how to access each final configuration. It comprises all the possible pathways to access all the possible configurations in the limit of no friction. In particular, it contains directional pathways, based on a step-after-step object incorporation and other protocol-dependent pathways in which adjacent objects are not necessarily incorporated with a temporal order. Therefore, this formalism comprises all the protocol-dependent pathways addressed by Eqs. (2)–(4), with which, by the way, all the possible configurations are accessed too.

Partition function  $Z$  does not make any assumption on a particular protocol and therefore it comprises all the possibilities for all the protocols.<sup>14</sup> In fact, sequence-dependent and equilibrium partition functions fulfill the following relations:

$$\left\langle \frac{1}{Z_\nu^{(\lambda)}} \right\rangle = \frac{1}{Z}, \quad (6)$$

$$\langle Z_\nu^{(\lambda)} \rangle_\lambda = Z, \quad (7)$$

which are trivially demonstrated from the definition of  $p_\nu$  and  $p_\nu^{(\lambda)}$ .<sup>14</sup> These equations are valid for every protocol  $\lambda$ , thus indicating that the equilibrium partition function is an average over all possible sequences for a fixed protocol, independent of which the protocol is. In other words,  $\langle p_\nu^{(\lambda)}/p_\nu \rangle = 1$  and  $\langle p_\nu/p_\nu^{(\lambda)} \rangle_\lambda = 1$ .

The pathways represented by Eq. (5) can be envisioned as a linear combination of forward and backward pathways in which objects are incorporated in the forward direction and removed in the backward one. In this picture, a decision is taken at each step as whether to incorporate a new object or remove the previous one(s). Such a decision is based on what the formerly incorporated objects are, hence representing a feedback mechanism. As already reported,<sup>14</sup> such a picture thus naturally comprises the concept of revision, in which backward recognition of errors (proofreading) and their substitution by new objects (edition) take place. We can actually associate the process represented by Eqs. (2) and (3) with *writing* and that represented by Eq. (5) with *revision*.

We now define the thermodynamic potentials, “ $U$ ,” “ $F$ ,” and “ $S$ ,” for single chains, namely, the *sequence- and protocol-dependent internal energy*, *Helmholtz free energy*, and *entropy*,

$$U_\nu^{(\lambda)} \equiv E_\nu, \quad (8)$$

$$F_\nu^{(\lambda)} \equiv -kT \ln Z_\nu^{(\lambda)}, \quad (9)$$

$$S_\nu^{(\lambda)} \equiv -k \ln p_\nu^{(\lambda)} \quad (10)$$

fulfilling

$$F_\nu^{(\lambda)} = U_\nu^{(\lambda)} - TS_\nu^{(\lambda)}, \quad (11)$$

which is the energy conservation. These potentials characterize the microstates of the system at  $i = n$ .

Note that  $E_\nu$  is independent of the protocol but not  $F_\nu^{(\lambda)}$  or  $S_\nu^{(\lambda)}$ . From a microcanonical point of view, these functions can be understood as the thermodynamic potentials for fixed energy  $E_\nu$ . When memory is extended to the complete history of the system, there is in general a one-to-one relationship between sequences and energies, but when memory is limited to a finite number of previous objects, a degeneration of sequences with the same energy  $E_\nu$  arises, what renders statistical meaning to the entropy within the framework of the microcanonical ensemble. Namely, the entropy can be expressed in the form of the Boltzmann formula “ $S = k \ln \omega$ ” by identifying  $\omega_\nu^{(\lambda)} = 1/p_\nu^{(\lambda)}$ .

The ensemble-average thermodynamic potentials can be constructed by taking expected values [see Eq. (1)] on Eqs. (8)–(10),

$$U^{(\lambda)} \equiv \langle E_\nu \rangle_\lambda, \quad (12)$$

$$F^{(\lambda)} \equiv -kT \langle \ln Z_\nu^{(\lambda)} \rangle_\lambda, \quad (13)$$

$$S^{(\lambda)} \equiv -k \langle \ln p_\nu^{(\lambda)} \rangle_\lambda. \quad (14)$$

$U^{(\lambda)}$ ,  $F^{(\lambda)}$ , and  $S^{(\lambda)}$  do not depend on the sequence, but they do depend on the protocol. Therefore, they will appear under the names of *protocol-dependent internal energy*, *Helmholtz*

*free energy*, and *entropy*, respectively. These potentials characterize the microstates of the system at time  $i = n$ .

The energy conservation for ensemble-average phenomena can also be expressed in terms of protocol-dependent potentials as

$$F^{(\lambda)} = U^{(\lambda)} - TS^{(\lambda)}, \quad (15)$$

which arise by formally taking expected values on Eq. (11); see Appendix A.

It is easy to see that equilibrium thermodynamics is a particular case of the above formalism. Namely, using Eq. (5) for the partition function and the probabilities, the energy conservation,  $F = U - TS$ , appears naturally by dropping the protocol superindex  $\lambda$  on Eqs. (12)–(14). Potentials  $U$ ,  $F$ , and  $S$  characterize the equilibrium states of the system.

Noteworthy, thermodynamic functions  $A_\nu^{(\lambda)}$ , which characterize the microstates of a system, and  $A^{(\lambda)}$ , which characterize the states of the system, are different from the equilibrium state functions,  $A$ , which are independent of time and of both protocols and pathways.

The internal energy for protocol-dependent stochastic chains fulfills

$$U^{(\lambda)} = - \left\langle \frac{\partial}{\partial \beta} \ln Z_\nu^{(\lambda)} \right\rangle_\lambda. \quad (16)$$

Likewise, the entropy for protocol-dependent, stochastic chains follows the law

$$S^{(\lambda)} = - \left\langle \frac{\partial}{\partial T} F_\nu^{(\lambda)} \right\rangle_\lambda. \quad (17)$$

Equations (16) and (17), while immediate in their demonstrations, are not obvious (Appendix A). Their equilibrium analogs appear, respectively, as particular cases within this formalism,

$$U \equiv \langle E \rangle = - \left\langle \frac{\partial}{\partial \beta} \ln Z \right\rangle = - \frac{\partial}{\partial \beta} \ln Z, \quad (18)$$

$$S \equiv -k \langle \ln p_\nu \rangle = - \left\langle \frac{\partial}{\partial T} F \right\rangle = - \frac{\partial}{\partial T} F. \quad (19)$$

To conclude this subsection, we derive an expression that relates the sequence-dependent partition function with the respective sequence-dependent energies (Appendix A), which can be useful from an experimental point of view. For each sequence,  $\nu$ , the sequence-dependent partition function can be estimated if knowledge of the energies  $E_{\mu\nu}$  exists through the following relation:

$$\frac{Z_\nu^{(\lambda)}}{Z} = \left\langle e^{-\beta(E_{\mu\nu} - E_\mu)} \right\rangle, \quad (20)$$

where the expected value is taken over all sequences  $\mu \equiv \mu(\lambda) = 1, \dots, N$  with the equilibrium probability distribution  $p_\mu$ , Eq. (5). It is in general difficult to apply Eq. (20) when interactions extend over many neighbors because energies  $E_{\mu\nu}$  involve many combinations for a defined protocol. When sufficient knowledge on the system is gathered, for example, in DNA replication, it is possible to restrict the elements of the energy data set.<sup>9,16</sup>

## B. Effective revision

We next consider inequalities between protocol-dependent and equilibrium statistics. The Kullback-Leibler distances are positive as expected,

$$D(p_\nu || p_\nu^{(\lambda)}) = \left\langle \ln \frac{Z_\nu^{(\lambda)}}{Z} \right\rangle \geq 0, \quad (21)$$

$$D(p_\nu^{(\lambda)} || p_\nu) = \left\langle \ln \frac{Z}{Z_\nu^{(\lambda)}} \right\rangle_\lambda \geq 0, \quad (22)$$

where we have applied Jensen's inequality<sup>4</sup> and Eqs. (6) and (7). Concerning the ensemble-average thermodynamic functions, the following inequalities hold:

$$F^{(\lambda)} \geq F \equiv -\frac{1}{\beta} \ln Z, \quad (23)$$

$$S - S^{(\lambda)} \geq \frac{1}{T} (U - U^{(\lambda)}). \quad (24)$$

Inequality (23) follows from the application of Jensen's inequality<sup>4</sup> to  $-\ln x$ , which is a convex function of  $x$ , and the use of Eq. (7). The fact that  $F^{(\lambda)} - F \geq 0$  physically means that a final state achieved under statistical equilibrium (i.e., by going through all pathways under all possible protocols) is always more stable than when the same state has been achieved under all the possible pathways but defined by only one specific protocol  $\lambda$ .

Inequality (24) is derived from the subtraction of Eq. (15) from  $F = U - TS$  and the use of inequality (23). The entropy of a chain can increase or decrease with respect to the statistical equilibrium entropy value (i.e.,  $S - S^{(\lambda)} \leq 0$ ); it depends on the protocol  $\lambda$ . In fact, certain protocols can decrease the entropy largely at the cost of high dissipations under non-equilibrium conditions.<sup>14,17</sup> Inequality (24) guarantees that if the ensemble-average internal energy of the chain achieved under protocol  $\lambda$  is lower than that achieved under statistical equilibrium (namely, if  $U - U^{(\lambda)} \geq 0$ ), the final entropy decreases below the statistical equilibrium value through this protocol (namely, then  $S - S^{(\lambda)} \geq 0$ ).

From the ensemble-average energy conservation, Eq. (15), it is clear that  $U \geq F$  and that  $U^{(\lambda)} \geq F^{(\lambda)}$ , which mean that the useful energies,  $F$  or  $F^{(\lambda)}$ , are always lower than the total energies,  $U$  or  $U^{(\lambda)}$ , respectively, due to the entropic term. From these inequalities, it follows that

$$U^{(\lambda)} \geq F, \quad (25)$$

which sets a minimal boundary for the ensemble-average internal energy of the system. Former results on DNA replication in the reversible limit<sup>9</sup> are compatible with inequalities (23)–(25). In particular, the statistical equilibrium entropy of the stochastic chain was lower than the entropy achieved under a directional construction protocol, and the same trend was observed for the internal energy, both of them to a strength compatible with inequality (24) at the temperature of that study.

We are now in the position to formulate conditions for effective revision:

**Condition 1** (effective revision necessary and sufficient condition).

$$\text{Effective revision} \iff S \leq S^{(\lambda)}, \quad (26)$$

$$\text{ineffective revision} \iff S > S^{(\lambda)}. \quad (27)$$

**Condition 2** (effective revision necessary condition).

$$\text{Effective revision} \Rightarrow U \leq U^{(\lambda)}. \quad (28)$$

**Condition 3** (ineffective revision sufficient condition).

$$U > U^{(\lambda)} \Rightarrow \text{Ineffective revision}. \quad (29)$$

The equalities take place when there exist no interactions with the previous neighbors (see the independence limit theorem<sup>14</sup>).

These conditions are consequences of Eq. (24). In the following, we apply this formalism to biomolecular processes that manage information, showing coherence with the above formalism and finding consequences to the associated natural systems.

## III. GENETIC INFORMATION TRANSFER

Nucleic acids are the genetic information carriers: linear polymers with writing and reading directionality that are common to all living beings on Earth. Their alphabet is a set of four symbols in contrast to the binary alphabet used in computer and communication sciences to date. Information is stored and transmitted by the arrangement of complementary data lists into double-helix, long molecules. These structures are flexible platforms that interact with proteins, these interactions being at the core of information processing in cells.<sup>18</sup> Understanding genetic information storage and transfer is not only interesting from a biological point of view but also inspiring nanotechnologies for recognition and for information and communication.<sup>11,12</sup>

We next illustrate the above theory with two natural processes in which information and thermodynamics appear together: DNA replication and RNA transcription. In these biophysical processes, nucleotides of four types are chained into a DNA/RNA strand according to a DNA template sequence. These four nucleotide types are univocally identified to four alphabet symbols: A, C, G, and T/U (which stand for adenine, cytosine, guanine, and thymine/uracil, the last of which corresponds to either DNA or RNA and to either replication or transcription). A double-stranded DNA copolymer is generated in DNA replication, and a DNA-RNA hybrid is generated during RNA transcription, a copying process which leads to the conversion of DNA sequences to the RNA "language" in the cells. Then, a DNA/RNA strand represents a sequence of symbols that contain messages (genes that code for proteins), and, at the same time, this sequence is a linear string of particles (molecular subunits) whose stability is related to a low total free energy of the copolymerized strand, and that is ultimately related to a low number of copying errors. In these systems, the definitions of errors and certainties were shown to be intimately linked to the energies of incorporating an A, a C, a G, or a T/U.<sup>9</sup>

Replication and transcription are enzymatic reactions that involve an energy source responsible for driving these

processes in a certain direction, either forward or backward.<sup>10</sup> The mechanisms used by DNA and RNA polymerases, their respective enzymes, naturally introduce a time arrow.

### A. Toy model

We will use the toy model developed in a previous study,<sup>14</sup> which was later on applied to explain the energetic contribution of fidelity to the thermodynamic efficiency of polymerases.<sup>10</sup> We propose partial energy functions with linear dependence on the energies of each object in the absence of neighboring interactions,

$$E_i(x_i; x_{i-1}, \dots, x_1) = \sum_{j=1}^i \kappa(i-j; \alpha) E_j(x_j), \quad (30)$$

where  $\kappa$  is a kernel function in the energy domain. This kernel can be positive or negative, thus representing positive or negative feedback, respectively.  $\alpha$  is a real parameter that addresses the neighboring-interaction strength: it increases when interactions weaken and fulfills  $\alpha \rightarrow +\infty$  in the limit of no interactions (independently distributed, *id*, random variables). The total energy of a sequence is

$$E_v = \sum_{i=1}^n \sum_{j=1}^i \kappa(i-j; \alpha) E_j(x_j). \quad (31)$$

Based on the behavior of the hybridization energies of the four nucleotide bases<sup>16,19</sup> for complementary strands, we will use the following energy spectrum:<sup>14</sup>

$$E_i(x_i) = \begin{cases} -E, & x_i \text{ correct,} \\ +E, & x_i \text{ error,} \end{cases} \quad (32)$$

where  $E$  is a real, positive number. Additionally, we propose that at each position  $i$ , there is only one so-called correct  $x_i$ , the rest being errors, a simplification that does not consider the mechanisms of spontaneous point mutations.<sup>20,21</sup>

The standard, directional, and *id* entropies were already calculated<sup>14</sup> and are provided in Appendix B, Eqs. (B1) and (B2), along with other auxiliary functions. In the following, we show the expressions of the rest of the thermodynamic potentials of the information chain.

The internal energies for the standard, directional, and *id* processes are, respectively,

$$U(\beta, E, |\mathcal{X}|; \alpha) = -E \sum_{i=1}^n a_i(\alpha) \Lambda_i(\beta E, |\mathcal{X}|; \alpha), \quad (33)$$

$$U^{(D)}(\beta, E, |\mathcal{X}|; \alpha) = -EA(\alpha) \Lambda^{(id)}(\beta E, |\mathcal{X}|), \quad (34)$$

$$U^{(id)}(\beta, E, |\mathcal{X}|) = -nE \Lambda^{(id)}(\beta E, |\mathcal{X}|), \quad (35)$$

where  $\Lambda_i$  and  $\Lambda^{(id)}$  are factors with absolute values between 0 and 1 given by Eqs. (B3) and (B4), respectively, and  $a_i$  and  $A$  are partial and total sums over the kernel functions given by Eqs. (B5) and (B6), respectively (see Appendix B).

The standard, directional, and *id* Helmholtz free energies are

$$F(\beta, E, |\mathcal{X}|; \alpha) = -\frac{1}{\beta} \ln Z(\beta E, |\mathcal{X}|; \alpha), \quad (36)$$

$$F^{(D)}(\beta, E, |\mathcal{X}|; \alpha) = -\frac{1}{\beta} \ln Z^{(id)}(\beta E, |\mathcal{X}|) - E \Lambda^{(id)}(\beta E, |\mathcal{X}|) (A(\alpha) - n), \quad (37)$$

$$F^{(id)}(\beta, E, |\mathcal{X}|) = -\frac{1}{\beta} \ln Z^{(id)}(\beta E, |\mathcal{X}|), \quad (38)$$

where  $Z$  and  $Z^{(id)}$  are given by Eqs. (B7) and (B9), respectively (see Appendix B).

### B. Effective revision in genetic information transfer

Both DNA replication and RNA transcription have proof-reading mechanisms<sup>22,23</sup> although they are more normal and effective in the former process, where edition improves fidelity by 10–10<sup>3</sup> fold with respect to the raw copying reaction. As previously introduced, DNA and RNA polymerases are the enzymes responsible for each process, which are nanomachines able to correlate the current nucleotide incorporation at  $i$ , see Fig. 1, with the previously formed base-pairs. The memory is thus related to the number of base-pairs in the resulting double-stranded polymer that each protein covers, which is different for DNA and RNA polymerases.

To approach the information transfer, we will use two kernel functions, one with hyperbolic attenuation and the other with exponential attenuation, Eqs. (C4) and (C5), respectively, in the influence of the previous neighbors over the current incorporation (Appendix C). These kernels, with appropriate neighboring-interaction strength and positive feedback in the above toy model, were shown to be sufficient to derive the order of magnitude of error rates in DNA replication,<sup>14</sup> hence, we will apply them to test the effective proofreading conditions.

Figure 2 shows the behavior of the entropy difference,  $\Delta S \equiv S - S^{(D)}$ , as a function of the energy invested per nucleotide,  $E$ , in building the replicate or transcript strand. As expected, according to condition 1, this increment is lower than zero for positive feedback regimes, Eq. (26), in agreement with DNA replication and transcription where the polymerase makes it possible to increase fidelity by proofreading. When the feedback is negative,  $\Delta S$  becomes positive, which, consistent with the first condition, as expressed by Eq. (27), implies that proofreading within this adverse coupling regime actually worsens the fidelity of the initially *written* replicate or transcript strand.

Asymptotic behaviors of the thermodynamic potentials within the toy model are given in Appendix C. In general, the higher the energy invested at each incorporation, the lower the  $S$  and  $S^{(D)}$ , tending to zero asymptotically. However, for negative feedback regimes, if the influence of the previously formed base-pairs in the energy of the current incorporation is higher than the energy contrast  $E$ , the entropy may not decrease below a certain threshold [see Fig. 2(b) for  $\alpha = 1.5$ ] because the positive value of a correct incorporation is balanced by a negative correlation.

Figure 3 shows the behavior of the internal energy difference,  $\Delta U \equiv U - U^{(D)}$ . The increments are negative for

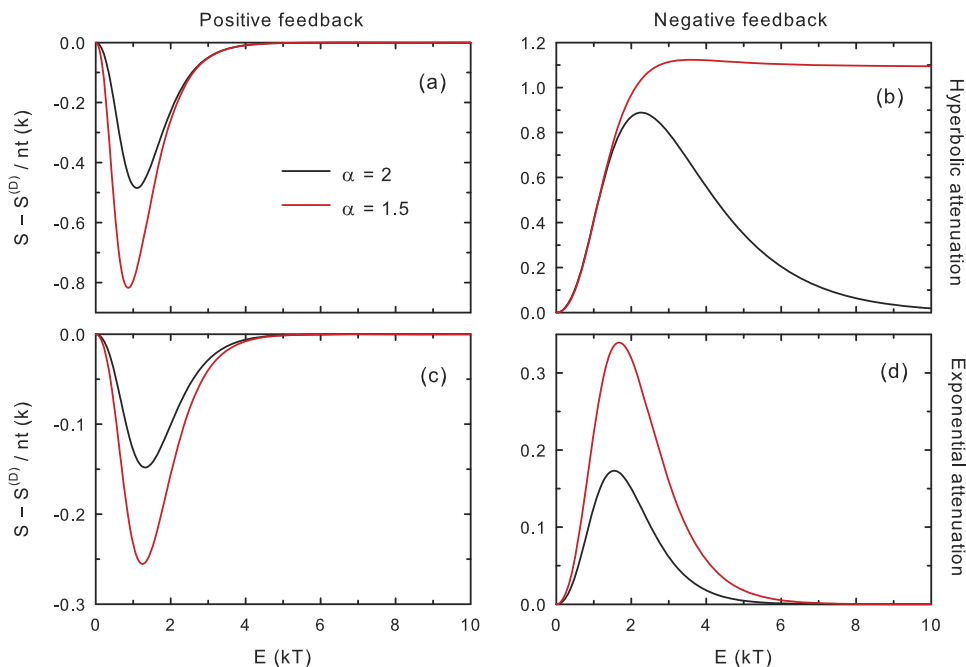


FIG. 2. Entropy difference per nucleotide (nt) as a function of the energy contrast between correct and wrong insertions. The curves in (a) and (b) have been calculated by imposing a hyperbolic attenuation for the influence of the past over the present, Eq. (C4), and those in (c) and (d) have been calculated by using an exponential attenuation, Eq. (C5) (Appendix C). [(a) and (c)] Positive feedback. [(b) and (d)] Negative feedback. Results are shown for  $n = 1000$  and two coupling strengths,  $\alpha$ .

positive feedback and positive for negative feedback with the exception of the curve for  $\alpha = 1.5$ , negative feedback, and hyperbolic coupling. For the last case, the correlations among previous neighbors become very important with respect to the hybridization energy alone, thus reversing the internal energy difference trend. In any case, all the data follow Eq. (24) and observe conditions 2 and 3, Eqs. (28) and (29). More in depth, the fact that  $\Delta U < 0$  is a necessary condition for effective revision, panels (a) and (c) in Figs. 2 and 3, but not sufficient, panel (b) in the same figures with  $\alpha = 1.5$ , and the fact that  $\Delta U > 0$  is a sufficient condition for ineffective revision, panels (b) and (d). DNA replication and RNA transcription give rise to double helices with higher fidelities upon proofreading, according to Figs. 2(a) and 2(c), which are related to lower internal energies, Figs. 3(a) and 3(c).

The free energy differences,  $\Delta F \equiv F - F^{(D)}$ , are plotted in Fig. 4. As predicted by Eq. (23), these increments are always negative and they can be used as control calculations. Another control calculation is obtained by using Eq. (25), which for our case states that  $U^{(D)} \geq F$ . Although not shown here for the sake of brevity, the simulations also obey this inequality.

The three potential differences increase when parameter  $\alpha$  decreases because the lower this parameter, the higher the interactions with the previous neighbors, making correlations stronger and thus the positive or negative feedback influence in the revision.

A common and important feature of these two copolymerization processes is also reflected in Figs. 2–4: The energies for which the maximum differences between writing and

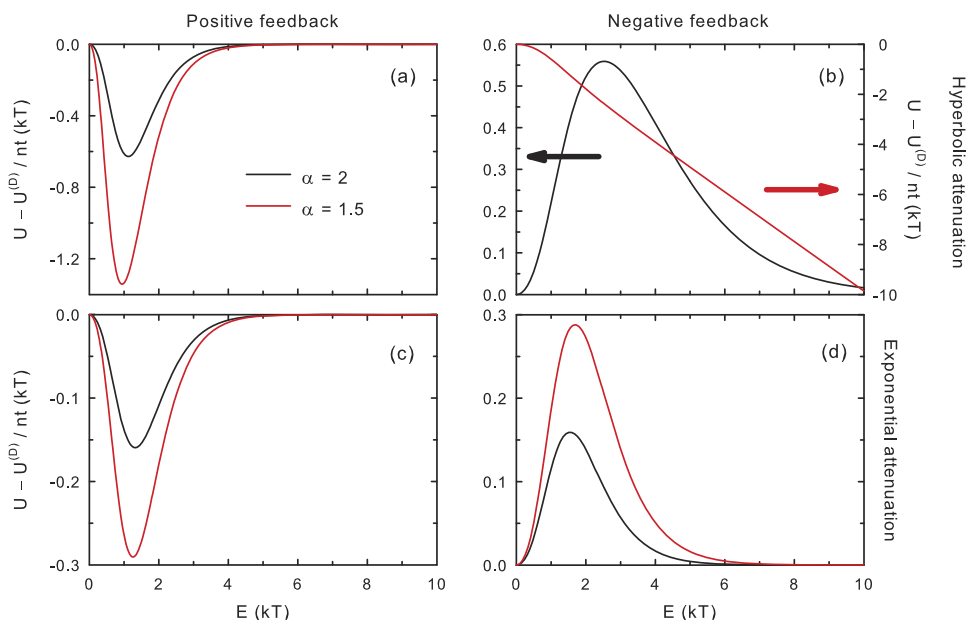


FIG. 3. The same as in Fig. 2 for the internal energy difference per nucleotide (nt).

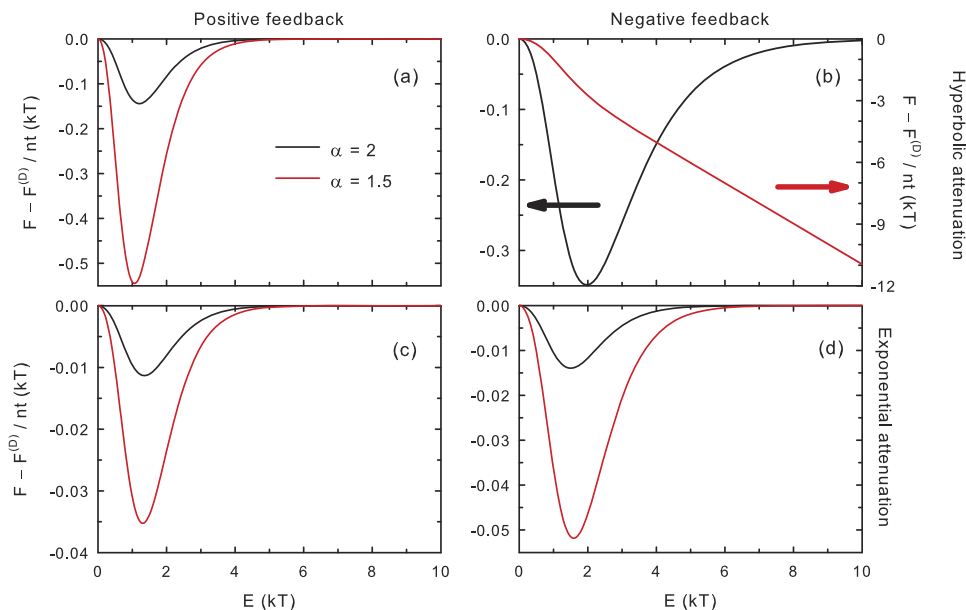


FIG. 4. The same as in Fig. 2 for the free energy difference per nucleotide (nt).

edition take place for the three thermodynamic potential differences are  $E \approx 1 - 3 kT$ . This energy level corresponds to the hybridization energy contrast between correct (Watson-Crick) and wrong base-pairs in both DNA and DNA/RNA hybrid double helices.<sup>16</sup> More in depth, Watson-Crick unions (namely, A-T, T-A, C-G, or G-C) involve an energy release between  $-1.5$  and  $-3 kT$ , whereas wrong incorporations normally involve energy absorptions between  $+1$  and  $+2 kT$ . Hybridization energies, as derived from melting experiments,<sup>16</sup> depend on the nature of the particular base-pair and exhibit a Markovian behavior, that is, a dependence on the previously formed base-pair.<sup>16</sup> This behavior—which, to build a sequence-average model, we have not included in the energy spectrum of Eq. (32)—contrasts with the non-Markovianity involved in the proofreading mechanisms of editing enzymes, for which the memory length depends on the structural fitting of the protein to the resulting nucleic acid double-stranded polymer (see Refs. 9, 24, and 25 for the case of DNA replication).

The peaks in the potential differences therefore imply that hybridization energies found in nature for the double-helix arrangement are optimal for discrimination protocols, as used by editing proteins during proofreading. The position of the critical points, in particular, of these maximum/minimum increments, is slightly different for the distinct potential increments since they are solutions of different characteristic equations, as shown in Appendix D. In addition, these positions depend slightly on whether the feedback is positive or negative and on the coupling strength,  $\alpha$ , and the alphabet cardinality,  $|\mathcal{X}|$ .

#### IV. CONCLUSIONS

Systems represented by strings of physical objects subject to thermal fluctuations convey information and accept a thermodynamic description. We have presented a theoretical framework to analyze their construction in the presence of a driving protocol and by addressing memory effects, inherent

to information management. To do that, we have introduced sequence- and protocol-dependent functions, including thermodynamic potentials, that characterize the non-Markovian states of the system. The canonical ensemble—with fixed temperature,  $T$ , and number (or density) of objects/symbols,  $n$ , in the chain—uses protocol-dependent potentials, whereas the microcanonical ensemble—with fixed pathway energy,  $E_V$ , besides  $T$  and  $n$ —uses both sequence- and protocol-dependent potentials. Within this framework, it has been possible to discern between writing and edition processes and to find conditions for effective proofreading.

This theory has been applied to DNA replication and transcription finding that the energies involved in the formation of double-stranded copolymers with genetic information are optimal for making error discrimination and correction effective. A high energy contrast between correct symbols/objects and errors is wished to increase the certainty in information transfer, but since edition requires the excision of the wrong symbolic objects, it might make the overall process energetically expensive. Double-stranded nucleic acids involve hybridization energies above but near the thermal level, which, besides making edition cost-effective, allows the optimization of fidelity by non-equilibrium reactions in which the polymerase enzyme enhances the energetic contrast between correct and wrong incorporations. Watson-Crick base-pairs can thus be envisioned as an evolutionary adaptation, among others, for making edition processes developed by polymerase enzymes not just effective from an information viewpoint but also thermodynamically efficient.

These natural systems inspire a strategy in the construction of artificial devices for storing and transferring information in the nanoscale: use of (i) energies above but near the thermal level to stabilize the information chain and (ii) nanometric heads that, by working out of equilibrium, increase the energetic contrast between correct and wrong symbolic objects in the writing and edition processes, which is required to optimize fidelity.

The toy model herein developed for DNA replication and RNA transcription is minimal to represent the main information and thermodynamic features of these reactions. This model may be, however, tuned to characterize these copying processes in the presence of protocols with step-dependent negative/positive feedback or under sequence-dependent effects so that it can be used to characterize specific polymerases.

## ACKNOWLEDGMENTS

This work was supported by the Spanish Ministry of Economy and Competitiveness (Grant No. MAT2015-71806-R).

## APPENDIX A: PROOFS

We next demonstrate the expressions of Sec. II A.

*Proof* to Eq. (15): From Eqs. (1)–(4) and (12)–(14), it follows that

$$\begin{aligned} S^{(\lambda)} &= -k \sum_{\nu=1}^N p_{\nu}^{(\lambda)} \ln p_{\nu}^{(\lambda)} = -k \sum_{\nu=1}^N p_{\nu}^{(\lambda)} (-\beta E_{\nu} - \ln Z_{\nu}^{(\lambda)}) \\ &= k (\beta \langle E_{\nu} \rangle_{\lambda} + \langle \ln Z_{\nu} \rangle_{\lambda}) = k\beta (U^{(\lambda)} - F^{(\lambda)}). \end{aligned}$$

*Proof* to Eq. (16): From Eqs. (12) and (1),

$$\begin{aligned} \langle E_{\nu} \rangle_{\lambda} &= \sum_{\nu=1}^N p_{\nu}^{(\lambda)} E_{\nu} = \sum_{\nu=1}^N \frac{e^{-\beta E_{\nu}}}{Z_{\nu}^{(\lambda)}} E_{\nu} \\ &= - \sum_{\nu=1}^N \frac{1}{Z_{\nu}^{(\lambda)}} \frac{\partial}{\partial \beta} e^{-\beta E_{\nu}} = - \sum_{\nu=1}^N \frac{\partial}{\partial \beta} p_{\nu}^{(\lambda)} \\ &= - \sum_{\nu=1}^N p_{\nu}^{(\lambda)} \frac{\partial}{\partial \beta} \ln Z_{\nu}^{(\lambda)} = - \left\langle \frac{\partial}{\partial \beta} \ln Z_{\nu}^{(\lambda)} \right\rangle_{\lambda}. \end{aligned}$$

*Proof* to Eq. (17): From Eqs. (15), (13), (16), and (9), in this order, it follows that

$$\begin{aligned} S^{(\lambda)} &= -\frac{1}{T} (F^{(\lambda)} - U^{(\lambda)}) \\ &= -\frac{1}{T} \left\langle -kT \ln Z_{\nu}^{(\lambda)} + \frac{\partial}{\partial \beta} \ln Z_{\nu}^{(\lambda)} \right\rangle_{\lambda} \\ &= - \left\langle \frac{\partial}{\partial T} F_{\nu}^{(\lambda)} \right\rangle_{\lambda}. \end{aligned}$$

*Proof* to Eq. (20): From Eqs. (3) and (5), it is straightforward that

$$\begin{aligned} \frac{Z_{\nu}^{(\lambda)}}{Z} &= \frac{Z_{\nu}^{(\lambda)} - Z}{Z} + 1 = \frac{1}{Z} \sum_{\mu^{(\lambda)=1}}^N (e^{-\beta E_{\mu\nu}} - e^{-\beta E_{\mu}}) + 1 \\ &= \frac{1}{Z} \sum_{\mu^{(\lambda)=1}}^N e^{-\beta E_{\mu}} (e^{-\beta(E_{\mu\nu}-E_{\mu})} - 1) + 1 \\ &= \sum_{\mu^{(\lambda)=1}}^N p_{\mu} (e^{-\beta(E_{\mu\nu}-E_{\mu})} - 1) + 1 \\ &= \left\langle e^{-\beta(E_{\mu\nu}-E_{\mu})} \right\rangle. \end{aligned}$$

## APPENDIX B: PARTITION FUNCTIONS, ENTROPIES, AND AUXILIARY FUNCTIONS FOR THE TOY MODEL

The entropy for the standard, directional, and *id* processes is

$$\begin{aligned} S(\beta E, |\mathcal{X}|; \alpha) &= k \ln Z(\beta E, |\mathcal{X}|; \alpha) \\ &\quad - k\beta E \sum_{i=1}^n a_i(\alpha) \Lambda_i(\beta E, |\mathcal{X}|; \alpha), \end{aligned} \quad (\text{B1})$$

$$\begin{aligned} S^{(D)}(\beta E, |\mathcal{X}|) &= S^{(id)}(\beta E, |\mathcal{X}|) = k \ln Z^{(id)}(\beta E, |\mathcal{X}|) \\ &\quad - kn\beta E \Lambda^{(id)}(\beta E, |\mathcal{X}|), \end{aligned} \quad (\text{B2})$$

where factors  $\Lambda_i$  and  $\Lambda^{(id)}$  are

$$\Lambda_i(\beta E, |\mathcal{X}|; \alpha) = \frac{e^{\beta E a_i(\alpha)} - (|\mathcal{X}| - 1)e^{-\beta E a_i(\alpha)}}{e^{\beta E a_i(\alpha)} + (|\mathcal{X}| - 1)e^{-\beta E a_i(\alpha)}}, \quad (\text{B3})$$

$$\Lambda^{(id)}(\beta E, |\mathcal{X}|) = \frac{e^{\beta E} - (|\mathcal{X}| - 1)e^{-\beta E}}{e^{\beta E} + (|\mathcal{X}| - 1)e^{-\beta E}} \quad (\text{B4})$$

and partial and total sums of the kernel,  $\kappa$ , are, respectively,

$$a_i(\alpha) = \sum_{j=1}^i \kappa(i-j; \alpha), \quad (\text{B5})$$

$$A(\alpha) = \sum_{i=1}^n a_i(\alpha). \quad (\text{B6})$$

The standard, sequence-dependent, and *id* partition functions read

$$\begin{aligned} Z &= \prod_{i=1}^n Z'_i \\ &= \prod_{i=1}^n (e^{\beta E a_i(\alpha)} + (|\mathcal{X}| - 1)e^{-\beta E a_i(\alpha)}), \end{aligned} \quad (\text{B7})$$

$$\begin{aligned} Z_{\nu} &= \prod_{i=1}^n Z_i = \prod_{i=1}^n (e^{\beta E} + (|\mathcal{X}| - 1)e^{-\beta E}) \\ &\quad \times \exp\left(-\beta \sum_{j=1}^{i-1} \kappa(i-j; \alpha) E_j(x_j)\right), \end{aligned} \quad (\text{B8})$$

$$Z^{(id)} = (e^{\beta E} + (|\mathcal{X}| - 1)e^{-\beta E})^n. \quad (\text{B9})$$

It is important to note that while index  $i$  runs over the ordered sequence positions,  $1, \dots, n$ , in Eq. (B8), this index is strictly not related to sequence positions in Eq. (B7).

## APPENDIX C: ASYMPTOTIC BEHAVIORS IN THE TOY MODEL

To illustrate representative physical systems, we use kernel functions with the following properties:

$$\lim_{j \rightarrow i} \kappa(i-j; \alpha) = 1, \quad (\text{C1})$$

$$\lim_{\alpha \rightarrow +\infty} \kappa(i-j; \alpha) = \delta_{ij}, \quad (\text{C2})$$



$$\lim_{i-j \rightarrow +\infty} \kappa(i-j; \alpha) = 0. \quad (\text{C3})$$

In particular, we have used the following kernels with hyperbolic or exponential attenuation, respectively, for simulations (see Figs. 2–4):

$$\kappa(i-j; \alpha) = \begin{cases} 1, & j = i, \\ +(-)1/(i-j+1)^\alpha, & j < i \end{cases} \quad (\text{C4})$$

and

$$\kappa(i-j; \alpha) = \begin{cases} 1, & j = i, \\ +(-)e^{-\alpha(i-j)}, & j < i. \end{cases} \quad (\text{C5})$$

They can be used for the positive (negative) feedback coupling regimes. In addition, they are bounded for  $\alpha > 1$ .<sup>14</sup>

In the limit of no interactions, it is fulfilled that

$$\lim_{\alpha \rightarrow +\infty} a_i(\alpha) = 1, \quad \lim_{\alpha \rightarrow +\infty} A(\alpha) = n, \quad (\text{C6})$$

$$\lim_{\alpha \rightarrow +\infty} \Lambda_i(\beta E, |\mathcal{X}|; \alpha) = \Lambda^{(id)}(\beta E, |\mathcal{X}|), \quad (\text{C7})$$

$$\lim_{\alpha \rightarrow +\infty} Z = \lim_{\alpha \rightarrow +\infty} Z_\nu = Z^{(id)}. \quad (\text{C8})$$

Then, the standard and protocol-dependent potentials converge to the case of independent variables,

$$\lim_{\alpha \rightarrow +\infty} S = \lim_{\alpha \rightarrow +\infty} S^{(D)} = S^{(id)}, \quad (\text{C9})$$

$$\lim_{\alpha \rightarrow +\infty} U = \lim_{\alpha \rightarrow +\infty} U^{(D)} = U^{(id)}, \quad (\text{C10})$$

$$\lim_{\alpha \rightarrow +\infty} F = \lim_{\alpha \rightarrow +\infty} F^{(D)} = F^{(id)}. \quad (\text{C11})$$

In the limit of low energies with respect to the thermal level,

$$\lim_{\beta E \rightarrow 0} \Lambda_i = \lim_{\beta E \rightarrow 0} \Lambda^{(id)} = \frac{2 - |\mathcal{X}|}{|\mathcal{X}|}, \quad (\text{C12})$$

$$\lim_{\beta E \rightarrow 0} Z = \lim_{\beta E \rightarrow 0} Z_\nu = \lim_{\beta E \rightarrow 0} Z^{(id)} = |\mathcal{X}|^n. \quad (\text{C13})$$

Therefore,

$$\begin{aligned} \lim_{\beta E \rightarrow 0} S &= \lim_{\beta E \rightarrow 0} S^{(D)} = \lim_{\beta E \rightarrow 0} S^{(id)} \\ &= kn \ln |\mathcal{X}|, \end{aligned} \quad (\text{C14})$$

$$\lim_{\beta E \rightarrow 0} U = \lim_{\beta E \rightarrow 0} U^{(D)} = \lim_{\beta E \rightarrow 0} U^{(id)} = 0, \quad (\text{C15})$$

$$\begin{aligned} \lim_{\beta E \rightarrow 0} F &= \lim_{\beta E \rightarrow 0} F^{(D)} = \lim_{\beta E \rightarrow 0} F^{(id)} \\ &= -\frac{1}{\beta} n \ln |\mathcal{X}|, \end{aligned} \quad (\text{C16})$$

which represent the maximum uncertainty.

The equalities in the thermodynamic potentials and partition functions in these two limits, Eqs. (C8)–(C11) and (C13)–(C16), are expected according to the independence limit theorem.<sup>14</sup> Besides, consistent with the same theorem, for  $\alpha > 1$  and sufficiently large,  $a_i > 0$ ,  $A > 0$ ,  $\lim_{\beta E \rightarrow +\infty} \Lambda_i = 1$  ( $\lim_{\beta E \rightarrow +\infty} \Lambda^{(id)} = 1$ ), and the limits

$$\begin{aligned} \lim_{\beta E \rightarrow +\infty} S &= \lim_{\beta E \rightarrow +\infty} S^{(D)} \\ &= \lim_{\beta E \rightarrow +\infty} S^{(id)} = 0, \end{aligned} \quad (\text{C17})$$

$$\begin{aligned} \lim_{\beta E \rightarrow +\infty} U &= \lim_{\beta E \rightarrow +\infty} U^{(D)} \\ &= \lim_{\beta E \rightarrow +\infty} -EA = -\infty, \end{aligned} \quad (\text{C18})$$

$$\lim_{\beta E \rightarrow +\infty} U^{(id)} = \lim_{\beta E \rightarrow +\infty} -En = -\infty, \quad (\text{C19})$$

$$\begin{aligned} \lim_{\beta E \rightarrow +\infty} F &= \lim_{\beta E \rightarrow +\infty} F^{(D)} \\ &= \lim_{\beta E \rightarrow +\infty} -EA = -\infty, \end{aligned} \quad (\text{C20})$$

$$\lim_{\beta E \rightarrow +\infty} F^{(id)} = \lim_{\beta E \rightarrow +\infty} -En = -\infty \quad (\text{C21})$$

hold for both positive feedback and negative feedback since the high contrast between correct insertions and errors leads asymptotically to a total certainty. For the kernels of Eqs. (C4) and (C5), it suffices that  $\alpha \geq 1.75$  for both the positive and negative feedback regimes.<sup>14</sup>

These results remain if the feedback is positive with independence of the actual value of  $\alpha$ , but if the feedback is negative and  $\alpha$  is not large enough, then  $a_i < 0$ ,  $A < 0$ , and  $\lim_{\beta E \rightarrow +\infty} \Lambda_i = -1$  ( $n$  and  $i$  sufficiently large), modifying some of the previous limits

$$\lim_{\beta E \rightarrow +\infty} S = kn \ln (|\mathcal{X}| - 1), \quad (\text{C22})$$

$$\lim_{\beta E \rightarrow +\infty} U = \lim_{\beta E \rightarrow +\infty} EA = -\infty, \quad (\text{C23})$$

$$\lim_{\beta E \rightarrow +\infty} U^{(D)} = \lim_{\beta E \rightarrow +\infty} -EA = +\infty, \quad (\text{C24})$$

$$\begin{aligned} \lim_{\beta E \rightarrow +\infty} F &= -\frac{1}{\beta} n \ln (|\mathcal{X}| - 1) \\ &+ \lim_{\beta E \rightarrow +\infty} EA = -\infty, \end{aligned} \quad (\text{C25})$$

$$\lim_{\beta E \rightarrow +\infty} F^{(D)} = \lim_{\beta E \rightarrow +\infty} -EA = +\infty. \quad (\text{C26})$$

In these conditions, the asymptotic behavior of the potential differences for  $\beta E \rightarrow +\infty$  yields

$$\lim_{\beta E \rightarrow +\infty} S - S^{(D)} = kn \ln (|\mathcal{X}| - 1), \quad (\text{C27})$$

$$\lim_{\beta E \rightarrow +\infty} U - U^{(D)} = \lim_{\beta E \rightarrow +\infty} 2EA = -\infty, \quad (\text{C28})$$

$$\begin{aligned} \lim_{\beta E \rightarrow +\infty} F - F^{(D)} &= -\frac{1}{\beta} n \ln (|\mathcal{X}| - 1) \\ &+ \lim_{\beta E \rightarrow +\infty} 2EA = -\infty, \end{aligned} \quad (\text{C29})$$

which is the case for  $\alpha = 1.5$ , hyperbolic attenuation, Eq. (C4), and negative feedback, as shown in Figs. 2–4(b). If  $\alpha > 1$  and large enough or the feedback is positive, the three potential differences tend to 0; see the rest of the curves in Figs. 2–4. For a binary alphabet,  $\Delta S$  always tend to zero asymptotically. This alphabet actually confers a symmetry for  $E < 0$  and  $E > 0$  in the potential differences.

#### APPENDIX D: CRITICAL POINTS IN EFFECTIVE REVISION WITHIN THE TOY MODEL

Critical points for  $S - S^{(D)}$ ,  $U - U^{(D)}$ , and  $F - F^{(D)}$  as functions of the energy  $E$  are given by the following equations, respectively:

$$0 = \beta E \left[ \sum_{i=1}^n a_i^2 (1 - \Lambda_i^2) - n (1 - \Lambda^{(id)^2}) \right], \quad (D1)$$

$$- \sum_{i=1}^n a_i \Lambda_i + A \Lambda^{(id)}$$

$$= \beta E \left[ \sum_{i=1}^n a_i^2 (1 - \Lambda_i^2) - A (1 - \Lambda^{(id)^2}) \right], \quad (D2)$$

$$- \sum_{i=1}^n a_i \Lambda_i + A \Lambda^{(id)}$$

$$= \beta E (1 - \Lambda^{(id)^2}) (n - A). \quad (D3)$$

It is straightforward to see that  $E = 0$  is a critical point for the three potential differences. Other critical points like those near  $\beta E = 1$ , see Figs. 2–4, or the asymptotic critical point for  $\beta E \rightarrow +\infty$  may appear depending on the behavior of the memory kernel. Critical points near  $\beta E = 1$  appear at different positions in  $E$  for the three potentials since Eqs. (D1)–(D3) are different.

- <sup>1</sup>C. Bustamante, C. Cheng, and Y. X. Mejia, *Cell* **144**, 480 (2011).
- <sup>2</sup>C. H. Bennett, *Int. J. Theor. Phys.* **21**, 905 (1982).
- <sup>3</sup>C. E. Shannon, *Bell Syst. Tech. J.* **27**(379-423), 623656 (1948).
- <sup>4</sup>T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, 1991).
- <sup>5</sup>F. Bernardi and J. Ninio, *Biochimie* **60**, 1083 (1978).
- <sup>6</sup>J. J. Hopfield, *Proc. Natl. Acad. Sci. U. S. A.* **71**, 4135 (1974).
- <sup>7</sup>J. Ninio, *Biochimie* **57**, 587 (1975).
- <sup>8</sup>R. Landauer, *Phys. Today* **44**(5), 23 (1991).
- <sup>9</sup>J. R. Arias-Gonzalez, *PLoS One* **7**, e42272 (2012).
- <sup>10</sup>J. R. Arias-Gonzalez, *Sci. Rep.* **7**, 7566 (2017).
- <sup>11</sup>G. M. Church, Y. Gao, and S. Kosuri, *Science* **337**, 1628 (2012).
- <sup>12</sup>N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. LeProust, B. Sipo, and E. Birney, *Nature* **494**, 77 (2013).
- <sup>13</sup>H.-P. Breuer, E.-M. Laine, J. Piilo, and B. Vacchini, *Rev. Mod. Phys.* **88**, 021002 (2016).
- <sup>14</sup>J. R. Arias-Gonzalez, *J. Chem. Phys.* **145**, 185103 (2016).
- <sup>15</sup>C. Bustamante, J. Liphardt, and F. Ritort, *Phys. Today* **58**(7), 43 (2005).
- <sup>16</sup>J. SantaLucia, Jr. and D. Hicks, *Annu. Rev. Biophys. Biomol. Struct.* **33**, 415 (2004).
- <sup>17</sup>D. Andrieux and P. Gaspard, *Proc. Natl. Acad. Sci. U. S. A.* **105**, 9516 (2008).
- <sup>18</sup>J. R. Arias-Gonzalez, *Integr. Biol.* **6**, 904 (2014).
- <sup>19</sup>D. A. Erie, T. D. Yager, and P. H. von Hippel, *Annu. Rev. Biophys. Biomol. Struct.* **21**, 379 (1992).
- <sup>20</sup>O. O. Brovarets' and D. M. Hovorun, *RSC Adv.* **5**, 99594 (2015).
- <sup>21</sup>O. O. Brovarets' and D. M. Hovorun, *RSC Adv.* **5**, 66318 (2015).
- <sup>22</sup>B. Ibarra, Y. R. Chemla, S. Plyasunov, S. B. Smith, J. M. Lazaro, M. Salas, and C. Bustamante, *EMBO J.* **28**, 2794 (2009).
- <sup>23</sup>J. F. Sydow and P. Cramer, *Curr. Opin. Struct. Biol.* **19**, 732 (2009).
- <sup>24</sup>E. C. Friedberg and P. L. Fischhaber, *eLS* (John Wiley & Sons Ltd., 2006), Vol. 69, p. 497.
- <sup>25</sup>T. A. Kunkel, *J. Biol. Chem.* **279**, 16895 (2004).