

Kinetic vs. energetic discrimination in biological copying

Pablo Sartori [‡] and Simone Pigolotti [‡]

1

Abstract

Copying information is a fundamental task of biological systems that has to be performed at a finite temperature. There is agreement that this fact alone implies a lower limit on the error rate. However, contrasting results have been obtained regarding how this limit is approached. For instance, it is not clear when it can be achieved in a slow, quasi-adiabatic regime or in a fast and highly dissipative one. In this paper, by means of analytical calculations and numerical simulations, we unravel a common feature of stochastic copying systems: the existence of two radically different copying modes. The first is based on different kinetic barriers, and is characterized by a high speed and high dissipation close to the lowest possible error. The second is based on energy differences between right and wrong copies, and is characterized by the fact that minimum copying error can be achieved at low speed and low dissipation. In models characterized by a single copying step, we demonstrate that these modes are alternative, i.e. they cannot be mixed to further reduce the minimum error. However, the two modes can be combined in multi-step reactions, such as in models implementing error correction through a proofreading pathway. By analyzing experimentally measured kinetic rates of two seemingly similar DNA polymerases, T7 and Pol γ , we argue that one of them operates in the kinetic and the other in the energetic regime.

Keywords: polymerization | proofreading | non-equilibrium | cellular information processing

1 Introduction

Living organisms need to process information in a fast and reliable way in order to perform biological functions. Copying information is a task of particular relevance, as it is required for the replication of the genetic code, which ensures heritability; and

^{*}Max Planck Institute for the Physics of Complex Systems. Noethnitzer Strasse 38, 01187, Dresden, Germany

[†]Corresponding author, simone.pigolotti@upc.edu. Dept. de Física i Eng. Nuclear, Universitat Politècnica de Catalunya Edif. GAIA, Rambla Sant Nebridi s/n, 08222 Terrassa, Barcelona, Spain

the transcription of DNA into mRNA and its translation into a chain of aminoacids, ultimately folding to form a protein. In general, biological copying is carried on by thermodynamic machines (such as polymerases and exonucleases) that operate at a finite temperature, so that a certain amount of errors is unavoidable. However, reliability is fundamental in the examples above, since a large error rate η can result in the costly (or harmful) production of a non-functional protein. To this aim, cells have developed complex mechanisms to reduce copying errors to values as low as $\eta \sim 10^{-4}$ for protein transcription-translation [1] and $\eta \sim 10^{-10}$ for DNA replication [2]. Such mechanisms include several discrimination steps [1, 2], as well as alternative pathways to undo wrong copies as in kinetic proofreading [2, 4], or even more complex error correcting schemes such as backtracking [5]. Despite the complexity of these copying schemes, some of their underlying thermodynamic principles can be understood by means of simple stochastic models.

A key quantity to understand the thermodynamics of copying is the dissipation per copied bit ΔS . The reason is that information processing and thermodynamics are linked by Landauer's principle [6], which establishes that energy is dissipated when information is erased. As later clarified by Bennett [7], information can actually be copied adiabatically. However, in the biophysics literature, where the typical problem of interest is copolymerization, contrasting results have been obtained regarding this issue. The copying scheme proposed in Hopfield's seminal proofreading paper [2] reaches its minimum error in an adiabatic limit of zero velocity v and zero dissipation [8]. However, in a seemingly similar model proposed few years later by Bennett [1], later termed copolymerization model [10, 11], the minimum error is achieved in a highly dissipative regime, where both the velocity and the dissipation diverge. In an adiabatic copying scheme, high copying fidelity comes at the cost of low copying speed, while in a very irreversible one it comes at the expense of high dissipation of chemical energy.

Some of the biological literature has favoured the viewpoint that the minimum error is achieved in near equilibrium conditions, see for example [8]. This view is however not unanimous [12]. Furthermore, recent biophysical literature supports a highly dissipative minimum error limit [13, 10, 14, 11]. Such contrasting results persist when copying schemes are generalized to include kinetic proofreading. While all models agree on the fact that, at low error rates, proofreading schemes are highly dissipative, the proofreading model in [1] dissipates systematically less energy than the corresponding copying scheme. This clearly contrasts with the outcome of Hopfield's model, in which at low errors dissipation occurs only in the proofreading step [2]. A recent study shows how the copying step in a proofreading scheme can be made faster, at the cost of a slightly higher error rate [15].

In this paper, we show how some of these contrasting points of view can be rationalized by making a clear distinction between two different modes of discrimination, that we term *energetic discrimination* and *kinetic discrimination*. In the former mode, an energy difference between the correct and wrong base matching is used to discriminate them. In this case, the minimum error is achieved when copies are produced adiabatically, with a low copying speed and corresponding low dissipation. Conversely, the latter mode uses an energy barrier for discrimination, resulting in a large dissipation and velocity when approaching the minimum error. Kinetic discrimination is at the core of the copying schemes in [1, 10, 11]. Instead, the original proofreading model [2] can be seen as the combination of an energetic discrimination step, corresponding to a reversible copy, followed by a kinetic discrimination step in the proofreading pathway. Remarkably, by analyzing experimentally measured rates of DNA duplication, we find

that copying in two seemingly similar systems (T7 and Pol γ) occur one in the kinetic and the other in the energetic regime.

We begin our study by presenting a simple model for copying a single bit of information in the spirit of those proposed in [7] and later in [16, 17, 18], see scheme in Fig 1a. In this model, a bio-machine such as a polymerase binds and unbinds monomers of different species to a template, trying to match it. We then move to the case of copolymerization (Fig. 1b), where a polymerase has to grow a polymer chain to match a template strand. The same binding/unbinding rates defined for the single bit case here represent rates of incorporation/removal of monomers. Finally, we discuss two different proofreading schemes, 1c and 1d, where the polymerase is assisted by an exonuclease that tends to remove wrong monomers. The two models differ by the absence or presence of an intermediate state in the copying process.

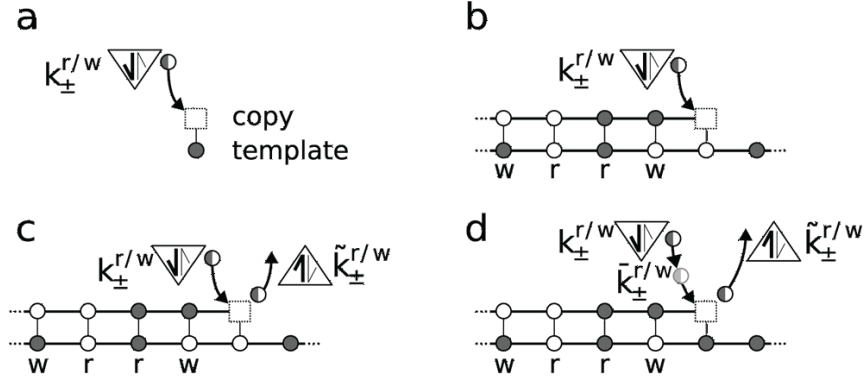


Fig. 1: Copying and proofreading schemes. a) Stochastic copying of a single bit. The lower-vertex triangle represents a bio-machine, as a polymerase, binding or unbinding one of two kinds of monomers to match a template. This reversible process is characterised by the rates of addition and removal of right and wrong matches $k_{\pm}^{r/w}$. b) Copolymerization. A template strand (bottom) is copied into a new strand (top) with the assistance of a polymerase. The addition and removal rates $k_{\pm}^{r/w}$ are biased by the template towards addition of right matches (r) and removal of wrong ones (w). c) & d) Proofreading schemes. In addition to a copying polymerase there is an uncopied exonuclease, represented by an upper-vertex triangle and characterized by the rates $\tilde{k}_{\pm}^{r/w}$. The exonuclease tends to remove monomers with a bias towards wrong copies. The difference between c) and d) is the absence or presence of an unbiased intermediate state in the monomer incorporation by the polymerase, characterised by rates $\bar{k}_{\pm}^{r/w}$.

2 Results

2.1 Copying strategies of a stochastic bit

We start our analysis with a simple example in which a single bit of information has to be copied stochastically. The copying machine is described as a three-states system. Two of the states correspond to the physical condition in which the right (r) or wrong (w) molecule is attached to the machine. We also consider a “blank” and distinct third state ($\emptyset$), representing the state of the system before a matching is done.

To help physical intuition and following [1], we define the parameters of the model in terms of the free energy differences among the three states. A sketch of the free energy landscape of the model is presented in Fig. 2a. Despite its simplicity, the interest of this model is that discrimination between the right and the wrong state can be achieved in two ways: via the difference in barrier height δ , or via the difference in energy levels of the final states γ . The energy difference ϵ can be thought as a chemical driving; while it does not affect discrimination in this model, it will play a crucial role in copolymerization.

The four rates k_+^r , k_-^r , k_+^w and k_-^w , connect the unbound state with the right and wrong states respectively. They can be written from the energy barriers in Fig. 2a by means of Kramer’s formula

$$k_+^r = \omega e^{\epsilon+\delta} ; k_-^r = \omega e^{\delta} ; k_+^w = \omega e^{\epsilon} ; k_-^w = \omega e^{\gamma}. \quad (1)$$

where we also introduced an overall rate unit ω . The three-state master equation governing the dynamics of the probabilities p_r and p_w of finding the system in the right or wrong state reads

$$\begin{aligned} \dot{p}_r &= (1 - p_r - p_w)k_+^r - k_-^r p_r \\ \dot{p}_w &= (1 - p_r - p_w)k_+^w - k_-^w p_w \end{aligned} \quad (2)$$

where the equation for the probability of the unbound state has been eliminated by means of the normalization condition. To study how the discrimination mechanisms affects the copying accuracy, we study the dynamics when the system is initially prepared in the unbound state, corresponding to the initial condition $p_r(t=0) = p_w(t=0) = 0$. The explicit analytical solution of the system of equations (2) with such initial condition and for a generic choice of the parameters is presented in *Supplementary Information (SI)*.

We now focus on the dynamics of the error, which is defined in terms of the state of the system as $\eta(t) = p_r(t)/[p_r(t) + p_w(t)]$. The importance of $\eta(t)$ is that, to maximize the copying accuracy, the copying reaction should be arrested when the error is at its minimum value. Indeed, in many concrete cases, dynamics of the form (2) will be carried over only for a limited time. After this time, the copying machine will terminate its task, and the system will be quenched into either a right or wrong copy outcome. This can be thought as arbitrarily increasing the energy of state $\emptyset$ in the diagram of Fig. 2a, so that the values of p_r , p_w and thus the error η will be frozen, see e.g. [7]. For example, in an enzymatic reaction, the termination event corresponds to the transformation of one of the two bound states into a product. In [7], where bits are encoded in one-domain ferromagnets, termination corresponds to the decoupling from an external transverse field.

The essential features of this copying system can be understood by discussing the behavior of the error in the limiting cases of very short time and very long times.

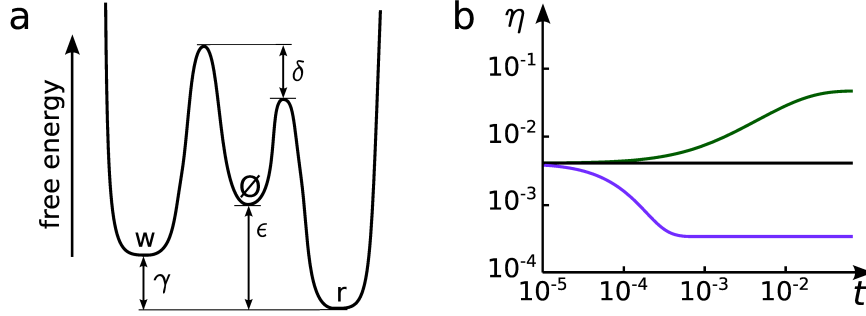


Fig. 2: a) Energy diagram for the single bit copying model. The barrier height difference δ , energy difference between right and wrong states γ , and the chemical driving ϵ are shown. b) Time-evolution of the error in a single bit copy for three parameter choices which represent the qualitatively different regimes: $\gamma < \delta$ (green curve, kinetic discrimination, $\gamma = 3$ and $\delta = 5.5$), $\gamma > \delta$ (blue curve, energetic discrimination, $\gamma = 8$ and $\delta = 5.5$), and the limiting case $\gamma = \delta = 5.5$ (black curve). The other parameters are $\epsilon = 5$ and $\omega = 4$.

At short times $t \ll 1/\omega$ the population of the right and wrong states will be simply determined by the transition probabilities to these states (forward kinetic rates) multiplied by the time, i.e. $p_r = tk_+^r$ and $p_w = tk_+^w$. To shorten the notation, we define the function $f(x) = e^{-x}/(1 + e^{-x})$ mapping energy differences into the corresponding errors. The short-time error is then given by $\eta(t \rightarrow 0) = f(\delta)$. In the opposite limiting case of $t \gg 1/\omega$, the system will reach equilibrium and the ratio of probabilities can be determined by imposing detailed balance, $p_w/p_r = e^\gamma$. This yields $\eta(t \rightarrow \infty) = f(\gamma)$.

For arbitrary times, one can analytically show that $\eta(t)$ is a monotonic function of time for any choice of the rates (see SI). This means that if $f(\delta) < f(\gamma)$ (i.e., if $\delta > \gamma$) the error is monotonically increasing over time from $f(\delta)$ to $f(\gamma)$. On the other hand, if $f(\delta) > f(\gamma)$ (i.e., if $\delta < \gamma$) the error is monotonically decreasing from $f(\delta)$ to $f(\gamma)$. Looking at Eq. 1, one can easily show that these conditions are equivalent to $k_-^r > k_-^w$ and $k_-^r < k_-^w$ respectively, independent of the choice of the on-rates.

The implications of this are the following. When the barrier difference δ is larger than the energy difference γ , the error is increasing with time, so that an optimal accuracy in the copying scheme requires stopping the binding-unbinding process as fast as possible. We refer to this as a *kinetic discrimination* regime. Conversely, when $\gamma > \delta$ the error is decreasing with time and to obtain the minimum error the reaction has to reach equilibrium before being terminated, which requires an arbitrarily large time. We refer to this as *energetic discrimination*. In the limiting case $\delta = \gamma$ the error is independent of time, so that the termination time is irrelevant for the purpose of minimizing the error. These three behaviors are exhibited in Fig. 2b for a particular choice of parameters.

At the simple level of this model, the parameter space can be divided in two regions, depending on whether small error can be achieved via kinetic discrimination (and thus fast copying) or energetic discrimination (quasi-static copying). The possibility of improving the error rate by using both mechanisms is ruled out. For the sake of

simplicity, we did not model explicitly the termination step. Notice however that in an energetic discrimination scheme, the quench can be performed slowly, and thus without dissipating in principle [7]. In the kinetic scheme, the quench has to be fast and will be highly dissipative. We will show in the following that similar features are present in more complex copying systems.

2.2 Kinetic and energetic discrimination in copolymerization

Copolymerization can be thought of as an extension of the model of the previous section to the case where many bits are copied in series. In copolymerization, a polymerase continuously adds and removes monomers to a tip of the growing copy strand, trying to match them with the monomers on a template strand (see Fig. 1b). Because of thermal fluctuations, this matching process exhibits errors, which will affect the free energy of the chain. However, existing studies assumed for simplicity that chains incorporating different amounts of errors were iso-energetic [1, 13, 10, 11]. By relaxing this assumption, we explore the possibility of a trade-off between kinetic and energetic discrimination, as well as studying the differences between these two copying schemes.

The copolymerization process is defined by the incorporation and removal rates of right $k_{\pm}^r$ and wrong $k_{\pm}^w$ matching monomers. We assume that the rates depends only on whether there is a match with the underlying template, and not on the specific type of monomer (this assumption is relaxed in [10]). In analogy with the previous section, the rates are defined by the energy landscape of Fig. 2a and Eq. 1. This implies that the chemical driving for right and wrong bases are given by ϵ and $\epsilon - \gamma$ respectively. Also in this case, discrimination can occur through the barrier height difference δ , or the higher energy γ that a wrong matching of monomers has over a right matching. Following [1], we call $[\&]$ an arbitrary state of the growing chain in the presence of the template, and $[\&r]$ and $[\&w]$ the states in which a right or wrong base has been added. At steady state there will be net fluxes of incorporation of the correct, $J_r = [\&]k_{+}^r - [\&r]k_{-}^r$, and wrong, $J_w = [\&]k_{+}^w - [\&w]k_{-}^w$, monomers. By imposing the relations between the fluxes and the error, $J_r = (1 - \eta)v[\&]$ and $J_w = \eta v[\&]$, being v the velocity of copy, the steady state solution can be explicitly calculated.

The two main quantities on which we focus are the average copying velocity v , and the rate of entropy production $\dot{S}$. The latter quantity is a measure of dissipation that can be also thought as the rate of energy that has not been used in performing chemical work [19]. In particular, we want to study their dependence on the error η . It is straightforward to show that $v = k_{+}^r - (1 - \eta)k_{-}^r + k_{+}^w - \eta k_{-}^w$ [1, 11]. The entropy production rate $\dot{S}$ can be calculated and is

$$\dot{S} = v\Delta S = v(1 - \eta)\epsilon + v\eta(\epsilon - \gamma) + vH(\eta) \quad (3)$$

where ΔS is the dissipation per step, and $H(\eta) = -\eta \log(\eta) - (1 - \eta) \log(1 - \eta)$ is the Shannon entropy corresponding to an error rate η . The first and the second contributions to the entropy production in the right hand side of Eq. (3) correspond to the distinct chemical driving forces of right and wrong bases, multiplied by the flux of right and wrong incorporated bases. The fact that in our case right and wrong bases have different drivings is the key difference with previous studies. As discussed in [1, 10, 11], the last term of Eq. (3) corresponds to the information entropy growth due to incorporation of errors, hence information, into the chain.

The behavior of the entropy production per copied base ΔS and the velocity v as a function of the error rate η are illustrated in Fig.3a and 3c, for a fixed value of γ and

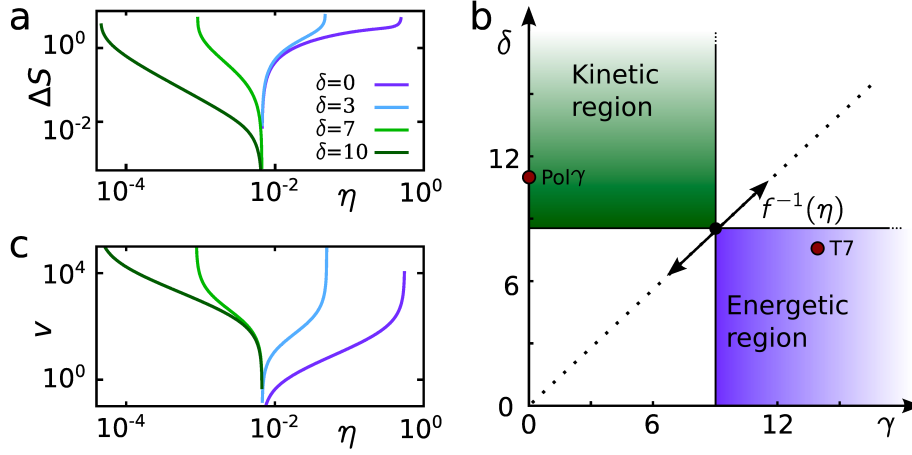


Fig. 3: a) Dissipation per step in the copolymerization model. In all curves $\gamma = 5$, while δ varies as shown in the legend. All curves tend to zero at the same value $\eta = f(\gamma) \approx 6.7 \cdot 10^{-3}$. Blue curves are in the energetic discrimination region, $\gamma > \delta$, while green curves are in the kinetic region $\delta > \gamma$. b) Phase diagram in the γ - δ plane, showing the parameter values compatible with a given error rate $\eta \sim f(-7) \approx 9 \cdot 10^{-4}$. The disconnected kinetic discrimination and energetic discrimination regions are shown. Estimated values of (γ, δ) for the Pol γ and T7 polymerases are also shown. c) Behavior of the average copying velocity v for the same parameter choices as in a).

several values of δ . Tuning γ and δ corresponds to varying the range of errors η being physically admissible. In particular, one finds the following condition on the error

$$\min[f(\delta), f(\gamma)] < \eta < \max[f(\delta), f(\gamma)], \quad (4)$$

where $f(x)$ has been defined in the previous section. A detailed derivation of formula (4) is presented in *SI*. To better understand the implications of the dissipation and velocity curves in Fig. 3 and formula (4), it is useful to first discuss two limiting cases: the highly dissipative limit $\eta \rightarrow f(\delta)$, and the adiabatic limit $\eta \rightarrow f(\gamma)$.

In the limit $\eta \rightarrow f(\delta)$, the chemical driving ϵ , the reaction velocity v and the dissipation per step ΔS diverge (see Fig. 3a and 3c, where for all choices of δ the curves diverge as $\eta \rightarrow f(\delta)$). From the explicit expression of the chemical driving (presented in *SI*) it can be shown that it diverges as $\epsilon \sim -\log|\eta - f(\delta)|$. Substituting this expression into the equation for the dissipation per step, Eq. (3), shows that also ΔS diverges as

$$\Delta S \sim \eta \epsilon \sim \eta \log|\eta - f(\delta)| \quad \text{for } \eta \rightarrow f(\delta). \quad (5)$$

Note that in this regime, as the chemical driving is very high, $\epsilon \gg 1$, the Shannon entropy term $H(\eta)$ in Eq. (3) is negligible, and the dissipation is dominated by the chemical terms. As an effect of the strong driving, the velocity also diverges as $|\eta - f(\delta)|^{-1}$.

Conversely, in the limit $\eta \rightarrow f(\gamma)$, both the reaction velocity v and the dissipation per step ΔS vanish, as the system approaches an adiabatic regime (see Fig. 3a and 3c, where all curves collapse to zero for $\eta = f(\gamma)$). This means that all the chemical energy is invested in copying the information, none being wasted. The chemical driving in this limit is equal to

$$\epsilon = \log(1 - \eta) = \log[1 - f(\gamma)] < 0 \quad \text{for} \quad \eta = f(\gamma). \quad (6)$$

Here ϵ is small and *negative*, in order to compensate for the effective positive driving caused by the Shannon entropy term in Eq. (3). The entropy production due to the error incorporation in the chain is balanced with negative chemical drivings in order to give small error in an adiabatic regime where $\Delta S = 0$ and $v = 0$.

From the inequality (4), we can explore the values of γ and δ that are compatible with a given error η . A given copy error η can be obtained with different choices of the discrimination parameters δ and γ by changing the driving ϵ . It is straightforward to derive from (4) that the values of γ and δ must satisfy either of the two conditions $\gamma < f^{-1}(\eta) < \delta$ or $\delta < f^{-1}(\eta) < \gamma$, where $f^{-1}(\eta) = \log(1 + 1/\eta)$ is the inverse of the function $f(x)$. These conditions define the two disconnected regions of the phase space in Fig. 3b. We then define the *kinetic discrimination region* of parameter space as the one characterized by $\delta > \gamma$, and the *energetic discrimination region* as the one where $\delta < \gamma$. Different points in the phase space for a desired copy error η are characterized by different values of the entropy production per copied base ΔS , as well as different copying velocities v .

A main new outcome of this analysis is that, depending on the sign of $\delta - \gamma$, discrimination has a drastically different behavior in proximity of the minimum error. In the kinetic region, the minimum error is attained when $\eta \rightarrow f(\delta)$. This implies a trade-off between small error and high entropy production, and thus high chemical energy spending. As the error reaches its lowest possible value $f(\delta)$, both quantities diverge. Note however that even in the kinetic region such monotone relation between error, dissipation and velocity only holds for small errors, since at larger errors the term $H(\eta)$ is not negligible and causes a non-monotonic behavior of ΔS [1]. Conversely, in the energetic region, accurate copying comes at the cost of the copying velocity, which goes to zero at the minimum error $\eta = f(\gamma)$.

We mention here that the copolymerization model can be thought of as a steady-state version of Hopfield's model (see SI for details of the mapping). Indeed, a main difference between Hopfield's copying model [2] and the copolymerization model [1] is that the former excludes forward discrimination (i.e. $\delta = 0$ in our language), but considers the different energy of the bound states corresponding to the different bases. The latter assumes the bound states to be isoenergetic ($\gamma = 0$) but reachable via different energy barriers.

2.3 Discrimination in DNA polymerase copying

To show the biological relevance of our framework, we now analyze and compare two specific examples: DNA replication of the phage T7 [2, 21], and replication of human cells DNA carried by the polymerase Pol γ [20].

A recent study of T7 [20] points at the rapid unbinding of nucleotides by strong and asymmetric backward rates as the leading mechanism determining low copying errors. We derive the copolymerization rates of T7 from the experimental values in [20]. To this aim, we consider the nucleotide binding to be very fast and at equilibrium,

with dissociation constants $K_r = 28\mu\text{M}$ and $K_w = 200\mu\text{M}$ for right and wrong base matching. We study the error rate in a range of nucleotide concentrations, as they can sensibly vary depending on the conditions [20]. Considering nucleotide concentrations in a range of $[dNTP] \sim 0.5 - 50\mu\text{M}$ we obtain the binding states $1/(1 + K_{r/w}/[dNTP])$ of right and wrong base matching. Multiplying these by the forward rates (360Hz and 0.2Hz for right and wrong bases respectively) we obtain $k_+^{r/w}$. The experimentally measured backward rates are $k_-^r \approx 2\text{Hz}$ and $k_-^w \approx 0.04\text{Hz}$ [20]. Substituting these values in the formula for the error rate results in an error range $\eta \sim 10^{-6} - 10^{-4}$, in agreement to the estimate in [2]. Notice that usual estimates of the error assume linear binding, approximation valid for low $[dNTP]$, which in our case corresponds to the lowest end of the error range. The resulting polymerization velocities are $v \sim 5 - 250\text{bps}$ (copied bases per second), in agreement with the saturation polymerization rate measured in [20]. By inverting Eqs.(1), we can infer that the energetic and kinetic discrimination factors are $\gamma \approx 14$ and $\delta \approx 8$ respectively. Since $\gamma > \delta$, we conclude that the polymerase operates in the energetic regime (see Fig. 3b).

A recent study [10] analyzed DNA duplication by Pol γ with a variant of the copolymerization model, where different monomer species are characterised by different rates. Parameter values were taken from measurements in [21]. Agreement with experimental data was obtained under the assumption that the copy be iso-energetic with respect to the error rate, that is $\gamma = 0$. In order to compare with the T7 case, we simplify their analysis by considering an average over the different monomer species. To this aim, we first average the dissociation constants of right and wrong bases, and multiply them by the average polymerization constants of right and wrong bases. Using the same chemical driving determined for T7, $\epsilon \approx 5$, we obtain a range of error rates $\eta \sim 10^{-5} - 10^{-3}$, which in the limit of low $[dNTP]$ agrees with the estimates in [21, 10]. The resulting kinetic discrimination parameter is $\delta \approx 11$. As $\gamma = 0$, the polymerase Pol γ lies in the kinetic discrimination region of parameter space (Fig. 3b).

The estimates above indicate that the two polymerases operate in the two different regimes proposed in the previous section. In particular, Fig 3b shows the location in phase space of the operating regimes of the two polymerases. While the two polymerases are characterized by a similar error rate, the fact that they operate in different regimes implies that they experience different trade-offs when some external parameter is varied. For example, a smaller error of the T7 polymerase is achieved when reducing the chemical driving, which in practice can be obtained by lowering the nucleotide concentration. Reducing $[dNTP]$ reduces both the error and the dissipation ΔS , at the cost of a smaller polymerization speed v . This situation is qualitatively similar to that of the blue curves in Figs. 3a and 3c. Conversely, error reduction for Pol γ requires a stronger driving, hence dissipation, as already observed in [10]. This in turn gives a faster polymerization rate, in analogy with the green curves of Figs. 3a and 3c.

Let us now discuss the robustness of this conclusion with respect to the assumptions made and the uncertainties in the measures. In the case of T7, our estimates of γ and δ are robust to changes of $[dNTP]$, which mainly affects the driving ϵ . For Pol γ however the value $\gamma = 0$ has, as in previous models [10], been assumed rather than inferred from measures. We were unable to find measures for off-rates allowing for a direct estimation of this parameter. However, our estimate leads to a polymerization velocity compatible with previous studies [10] and experimental evidences [21]. Note that only a very large value $\gamma > \delta = 11$ would invalidate the hypothesis that Pol γ operates in the kinetic region. Furthermore, an energetic mechanism (depending on

the value of γ) would yield a polymerization velocity too low to match the observed value of 37 bps [21].

In vivo DNA copying achieves lower error rates than those described so far. This is done by the usage of error-correcting schemes such as proofreading. In the next section we discuss how the strategies so far presented are combined in two proofreading examples.

2.4 Combining copying strategies in proofreading schemes

So far, we have shown that the kinetic and energetic discrimination mechanisms cannot be combined to further reduce the errors in the same reaction, as expressed in Eq. 4. However, the two mechanisms can be combined in more complex multi-step copying schemes. In this section, we explore this possibility in two simple examples of copying schemes involving a proofreading step.

In kinetic proofreading, an initially copied base can be removed via an alternative “proofreading” pathway, typically assisted by an exonuclease, see Fig. 1c and 1d. Such erasing pathway is also characterized by a discrimination which, a priori, can be energetic γ_p or kinetic δ_p , a distinct time-scale $1/\omega_p$, and a (backward) driving ϵ_p . In an effective proofreading scheme, the minimal copying error of Eq. 4 is reduced by another discrimination factor, which a priori can also be energetic $f(\gamma_p)$ or kinetic $f(\delta_p)$. We discuss two different schemes of proofreading. In the first, represented in Fig. 1c, there are no intermediate states between the pre-copied and copied states, so that the proofreading pathway carried by the exonuclease and characterized by $\tilde{k}_{\pm}^{r/w}$ has the same structure as the copying pathway, apart from the backward driving. A restricted version of this model allowing only for kinetic discrimination in copying and proofreading was originally studied in [1]. Here, we allow both pathways to have both types of discriminations. In the second scheme, sketched in Fig. 1d, the first copying step leads to an intermediate state. The second step takes this intermediate state into its final form without discrimination between right or wrong bases, and is only characterized by a chemical driving ϵ^* and a time-scale $1/\omega^*$ which define the rates $\tilde{k}_{\pm}^{r/w}$. We take the proofreading reaction to be the same in the two cases. The scheme with an intermediate step has the same structure of Hopfield’s proofreading model [2], where he focused on a regime where the copying step was implemented through energetic discrimination. Details on the parametrization are in *Materials and Methods*.

In both proofreading models, the entropy production rate $\dot{S}$ and the velocity v can be analytically expressed with a procedure similar to that used for the copolymerization model (see *SI*). A main difference is that fixing the discriminations δ , δ_p , γ , γ_p and the error rate η does not fully specify $\dot{S}$ and v due to the presence of the additional free parameters presented above (the proofreading timescales and chemical drivings). With a numerical algorithm described in the *SI*, we minimize for each error rate η the entropy production per step ΔS over the free parameters, thus obtaining the curves of minimum dissipation vs error shown in Fig. 4.

By a systematic analysis of parameter space it is easy to reach the conclusion that the proofreading step always uses kinetic discrimination. As shown in detail in *SI*, there are no regimes in any of the two models where it is possible to lower the copying error by performing energetic discrimination in the proofreading step. That is, the standard copying error can never be reduced by a factor $f(\gamma_p)$. It is however possible to obtain error reduction with a factor $f(\delta_p)$, corresponding to kinetic discrimination in the proofreading step. As kinetic discrimination is a very dissipative process this justifies the well known fact that a proofreading scheme requires an additional energy input.

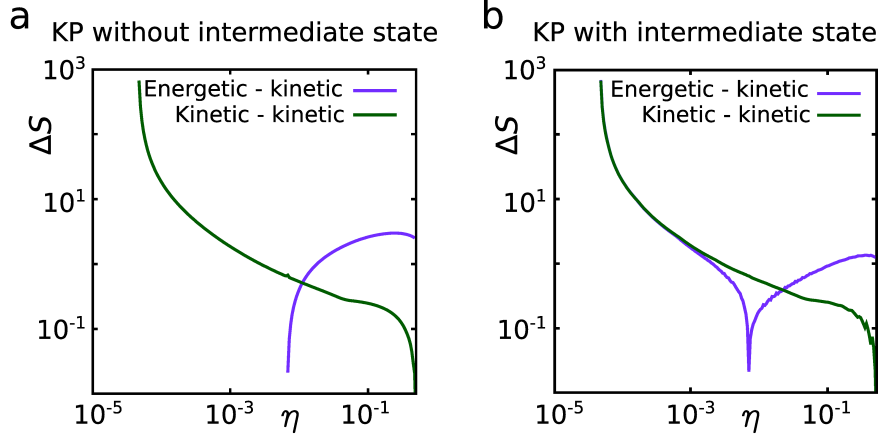


Fig. 4: a) Minimum dissipation per step in the proofreading model without an intermediate step, Fig. 1c. For all curves, we chose $\gamma_p = 0$ and $\delta_p = 5$. In the case of energetic discrimination of the copying step and kinetic discrimination of the proofreading step (energetic-kinetic), the other parameters are $\gamma = 5$, $\delta = 0$. In the kinetic-kinetic case, the other parameters are $\gamma = 0$ and $\delta = 5$. We see how the energetic-kinetic scheme does not reduce the error beyond the standard value $f(\gamma)$, while the kinetic-kinetic scheme reduces it beyond $f(\delta)$ down to $f(\delta)f(\delta_p)$. b) Minimum dissipation per step in the proofreading model with an intermediate step, model d in Fig. 1. Parameters are the same as in panel a), but in this case both schemes reduce the error by $f(\delta_p)$. In the kinetic-kinetic curve down to $f(\delta)f(\delta_p)$ and in the energetic-kinetic down to $f(\gamma)f(\delta_p)$.

An intuitive explanation stems from the fact that proofreading implies a backward flux on the polymerization process, i.e. proofreading is a mechanism to undo wrong copies. This means that the copying and proofreading pathways together constitute loops carrying net probability currents (see Fig. 2c and 2d), which implies that their dynamics must be dissipative. Such result is also consistent with Landauer's principle [6] as, unlike copying, the proofreading step essentially involves erasure of information.

We have shown that proofreading is a dissipative process and only operates through kinetic discrimination. But the question remains on whether it can be freely combined with adiabatic (energetic) and dissipative (kinetic) copying. By numerically exploring the parameter space, we could show that while proofreading is always compatible with dissipative copying, it is only compatible with adiabatic copying when an intermediate state is present. In other words, the critical error of the energetic-kinetic curve in Fig. 4a, is simply $f(\gamma)$, and so kinetic proofreading is not effective. The same curve in the presence of an intermediate state (Fig. 4b) presents a minimum error equal to $f(\gamma)f(\delta_p) \approx f(\gamma + \delta_p)$. The necessity of an intermediate state for the energetic-kinetic scheme can be intuitively understood as follows. Since effective proofreading is always kinetic, it must be performed very fast. On the other hand, energetic copying requires

a long relaxation time. As a consequence, both reactions can not be combined in parallel unless an additional step from which the proofreading step flows is added. The combination of kinetic proofreading with an adiabatic copying step has the clear advantage of a lower dissipation as compared to its combination with a dissipative copying step (see Fig. 4b, green vs. blue lines).

3 Discussion

In this paper, we analyzed stochastic models for biological copying systems. Our main finding is that *each* copying step in these models can be unambiguously classified into one of two radically different classes, that we termed kinetic and energetic discrimination. Kinetic discrimination exploits energy barriers; in this regime, both the velocity and the dissipation per copied monomer diverge when the error approaches its minimum. Conversely, energetic discrimination exploits energy differences between right and wrong copies, and is characterized by a slower and less dissipative copying as the error approaches its minimum value. The existence of an energetic regime in the copolymerization model complements the view in the literature [1, 10, 11] that low copy errors are achieved only in a highly dissipative regime. It also demonstrates how entropy driven growth, a phenomenon studied in the large error limit of the isoenergetic model [1, 13, 10, 14], can indeed be exploited to reliably copy information. In this regime, copolymerization is also compatible with the principles of reversible computing, which state that a copy can be performed adiabatically since no information erasure is involved [7].

The analysis of measured kinetic rates of two DNA polymerases, T7 and Pol γ , shows that the first operates in the energetic regime, while the second in the kinetic one. This shows how both mechanisms can be used by real systems. The copolymerization model we used can be readily generalized to account for different kinetic rates of different nucleotides, as in [10]. Such generalization would not alter qualitatively our result, as the same analysis can be performed separately for each nucleotide species. It is known that very low errors can be achieved in multi-step reactions [2, 24], the prime example being kinetic proofreading. Our study of two-steps and three-steps proofreading reactions combining both regimes constitutes a proof of principle that the two regimes defined in this paper are present and can help analyzing the parameter space of more complex copying reaction schemes.

We conclude by mentioning that, while we focused here on biological copying processes, the conceptual framework we presented can be applied to a wider range of biological problems. Indeed, most of the fundamental thermodynamical features of biological copying are actually related to a discrimination process operating at a finite temperature. These features can thus be extended to cases in which discrimination is not used to copy information, but rather to perform some other biological task. A well studied example is the discrimination of antigens by T-cell receptors, a system which has been shown to use a proofreading mechanism [23]. Other cases of interest come from neural dynamics. For example, it has been shown that discrimination of a binary input signal can be explained by a model consisting of a bi-stable stochastic neural circuit [25]. Finally, a recent advance has revealed a trade-off between adaptation error in bacterial chemotaxis and dissipation [27], similar to the behavior of copying in the kinetic regime. In general, at the sub-cellular level, thermal fluctuations dominate and impose constraints on biological tasks. While the thermodynamics of bio-mechanical systems such as molecular motors is well understood [26], the role of fluctuations in

biological information processing presents still many open questions. Our work shows that the emerging trade-offs may be complex, and depend on the region in parameter space where the system operates.

4 Materials and Methods: details of proofreading schemes

In the proofreading scheme without intermediate states, there are two pathways. The standard copying pathway carried by the polymerase is characterized by the rates in Eq. 1. The proofreading pathway assisted by the exonuclease is determined by additional four rates parametrized as:

$$\tilde{k}_{-}^r = \omega_p e^{\epsilon_p - \delta_p} ; \tilde{k}_{+}^r = \omega_p e^{-\delta_p} ; \tilde{k}_{-}^w = \omega_p e^{\epsilon_p} ; \tilde{k}_{+}^w = \omega_p e^{-\gamma_p}. \quad (7)$$

Notice that, at variance with the rates in Eq. (1), here the definition is such that $\epsilon_p > 0$ corresponds to a backward driving. The parameters δ_p , γ_p and ω_p have the same meaning as in Eq. 1.

In the proofreading scheme which incorporates an intermediate state, there are three sets of rates. The first set are the copying rates to the intermediate state, these are those of the standard copying scheme given in Eq. 1. The second set corresponds to the copying from the intermediate to the final state. These rates have a forward driving ϵ^* and a time scale $1/\omega^*$ while, in analogy with Hopfield's original model [2], they exhibit no discrimination:

$$\bar{k}_{+}^r = \omega^* e^{\epsilon^*} ; \bar{k}_{-}^r = \omega^* ; \bar{k}_{+}^w = \omega^* e^{\epsilon^*} ; \bar{k}_{-}^w = \omega^*. \quad (8)$$

The third set of parameters are the four proofreading rates from the final state to the initial state, and are the same as in the proofreading scheme above, Eq. (27).

5 Acknowledgements

This work was partially supported by a Max Planck Society scholarship (to P.S.) and a Ramon y Cajal Grant (to S.P.). We are grateful to A. Bernacchia, J. Garcia-Ojalvo, N. Mitarai, L. Granger, M. A. Muñoz and Y. Tu for a critical reading of the manuscript.

References

- [1] H.S. Zaher and R. Green, *Fidelity at the Molecular Level: Lessons from Protein Synthesis*, Cell 136:746–762 (2009).
- [2] K.A. Johnson, *Conformational coupling in DNA polymerase fidelity*, Annu. Rev. biochem. 62:685–713 (1993).
- [3] J.J. Hopfield, *Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity*, Proc. Natl. Acad. Sci. USA 71:4135–4139 (1974).
- [4] J. Ninio, *Kinetic amplification of enzyme discrimination*, Biochimie. 57 (5): 587–95 (1975).
- [5] J.W. Shaevitz, E.A. Abbondanzieri, R. Landick, S.M. Block, *Backtracking by single RNA polymerase molecules observed at near-base-pair resolution*, Nature 426(6967):684–687 (2003).

- [6] R. Landauer, Irreversibility and Heat Generation in the Computing Process. IBM J. Res. Dev. 5, 183-191 (1961).
- [7] C.H. Bennett, *The Thermodynamics of Computation - a Review*, International Journal of Theoretical Physics, Vol. 21, 12 (1982).
- [8] M. Johansson, M. Lovmar and M. Ehrenberg, *Rate and accuracy of bacterial protein synthesis revisited*, Curr. Opin. Microbiol. 11:141-147 (2008).
- [9] C.H. Bennett, *Dissipation-error trade-off in proofreading*, Biosystems 11:85-91 (1979).
- [10] D. Andrieux and P. Gaspard, *Nonequilibrium generation of information in copolymerization processes*, Proc. Natl. Acad. Sci. USA 105:9561-9521 (2008).
- [11] M. Esposito, K. Lindenberg and C. Van den Broeck, *Extracting chemical energy by growing disorder: efficiency at maximum power*, J. Stat. Mech., doi:10.1088/1742-5468/2010/01/P01008 (2010).
- [12] R.C. Thompson and A.M.R. M. Karim, *The accuracy of protein biosynthesis is limited by its speed: High fidelity selection by ribosomes of aminoacyl-tRNA ternary complexes containing GTP[γ S]*, Proc. Natl. Acad. Sci. USA 79:4922-4926 (1982).
- [13] C.H. Bennett and M. Donkor, *Thermodynamics of error correction: speed-error-dissipation trade-off in copying*, Information Theory Workshop, 2008.
- [14] C. Jarzynski, *The thermodynamics of writing a random polymer*, Proc. Natl. Acad. Sci., 105:9451-9452 (2008).
- [15] A. Murugan, D. A. Huse, and S. Leibler, *Speed, dissipation, and error in kinetic proofreading*, Proc. Natl. Acad. Sci. USA 109(30), 12034-9 (2012).
- [16] R. Kawai, J.M.R. Parrondo and C. Van den Broeck, *Dissipation: The Phase-Space Perspective*, Phys. Rev. Lett. 98, 080602 (2007).
- [17] T. Sagawa and M. Ueda, *Minimal Energy Cost for Thermodynamic Information Processing: Measurement and Information Erasure*, Phys. Rev. Lett. 102, 250602 (2009).
- [18] L. Granger and H. Kantz, *Thermodynamic cost of measurements*, Phys. Rev. E 84, 061110 (2011).
- [19] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*, Wiley, 1985.
- [20] Y.C. Tsai and K.A. Johnson *A New Paradigm for DNA Polymerase Specificity*, Biochemistry 45:9675-9687 (2006)
- [21] H.R. Leei and K.A. Johnson *Fidelity of the Human Mitochondrial DNA Polymerase*, The Journal of Biological Chemistry, Vol. 281:36236-36240 (2006)
- [22] M.V. Rodnina and W. Wintermeyer, *Ribosome fidelity: tRNA discrimination, proofreading and induced fit*, Trends in Biochemical Sciences 26:124-130 (2001).
- [23] T.W. McKeithan, *Kinetic proofreading in T-cell receptor signal transduction*, Proc. Natl. Acad. Sci. 92:5042-5046 (1995).
- [24] R.R. Freter and M.A. Savageau, *Proofreading systems of multiple stages for improved accuracy of biological discrimination*, Jour. Theo. Biol. 85(1), 99-123 (1980).

-
- [25] X. J. Wang, *Probabilistic Decision Making by Slow Reverberation in Cortical Circuits*, Neuron 36:955-968 (2002).
 - [26] A. Parmeggiani, F. Juelicher, A. Ajdari, and J. Prost, *Energy transduction of isothermal ratchets: Generic aspects and specific examples close to and far from equilibrium*, Phys. Rev. E 60, 2127-2140 (1999).
 - [27] G. Lan, P. Sartori, S. Neumann, V. Sourjik and Y. Tu, *The energy-speed-accuracy trade-off in sensory adaptation*, Nat. Phys. 8, 422-428 (2012).

6 Supplementary Information

This document contains additional details and derivation of the results presented in the paper “Energetic vs. kinetic discrimination in biological copying”. The document is organized as follows. Section 1 presents a full solution of the single bit copying model (model A in Fig. 1 of the main text), and a demonstration that the error is always a monotonic function of time. Section 2 details results on the co-polymerization model (model B in the main text). Section 3 illustrates the mapping between Hopfield’s and Bennett’s copying schemes. Finally, section 4 and 5 present details on the proofreading models (models C and D in the main text, respectively).

6.1 Stochastic copying of a single bit

We wish to demonstrate that, for any choice of the rates, the solution of the system of differential equations:

$$\begin{aligned}\dot{r} &= k_+^r(1 - r - w) - k_-^r r \\ \dot{w} &= k_+^w(1 - r - w) - k_-^w w.\end{aligned}\tag{9}$$

with initial condition $r(t=0) = w(t=0) = 0$ leads to a time dependent error

$$\eta(t) = \frac{w(t)}{r(t) + w(t)}.\tag{10}$$

being a monotone function of time for all $t > 0$. In particular, $\eta(t)$ either strictly increasing, strictly decreasing or constant depending on the choice of the parameters.

The solution of the system of equations (9) can be obtained with standard methods. First of all, the steady state solution is:

$$\begin{aligned}r_{eq} &= \frac{1}{1 + \frac{k_-^r}{k_+^r} + \frac{k_-^r k_+^w}{k_+^r k_-^w}} \\ w_{eq} &= \frac{1}{1 + \frac{k_-^w}{k_+^w} + \frac{k_-^w k_+^r}{k_+^w k_-^r}}.\end{aligned}\tag{11}$$

Upon defining $\delta r = r - r_{eq}$ and $\delta w = w - w_{eq}$ the time-dependent distances from the steady state, a lengthy but straightforward calculation leads to

$$\begin{aligned}\delta r(t) &= \frac{N_- \left\{ -r_{eq} + \frac{w_{eq}}{2k_+^w} [q + \sqrt{q^2 + 4k_+^r k_+^w}] \right\} e^{\lambda_+ t}}{4 + \frac{q^2}{k_+^r k_+^w}} \\ &+ \frac{N_+ \left\{ -r_{eq} + \frac{w_{eq}}{2k_+^w} [q - \sqrt{q^2 + 4k_+^r k_+^w}] \right\} e^{\lambda_- t}}{4 + \frac{q^2}{k_+^r k_+^w}} \\ \delta w(t) &= \frac{-N_- (q + \sqrt{q^2 + 4k_+^r k_+^w}) \left\{ -r_{eq} + \frac{w_{eq}}{2k_+^w} [q + \sqrt{q^2 + 4k_+^r k_+^w}] \right\} e^{\lambda_+ t}}{2k_+^r \left(4 + \frac{q^2}{k_+^r k_+^w} \right)} \\ &- \frac{N_+ (q - \sqrt{q^2 + 4k_+^r k_+^w}) \left\{ -r_{eq} + \frac{w_{eq}}{2k_+^w} [q - \sqrt{q^2 + 4k_+^r k_+^w}] \right\} e^{\lambda_- t}}{2k_+^r \left(4 + \frac{q^2}{k_+^r k_+^w} \right)}\end{aligned}\tag{12}$$

where we defined the eigenvalues

$$\lambda_{\pm} = \frac{-\Sigma \pm \sqrt{\Sigma^2 - 4c}}{2} \quad (13)$$

with $\Sigma = k_+^r + k_-^r + k_+^w + k_-^w$ and $c = (k_+^r + k_-^r)(k_+^w + k_-^w) - k_+^r k_+^w$, and the quantities

$$\begin{aligned} q &= (k_+^r + k_-^r - k_+^w - k_-^w) \\ \mathcal{N}_{\pm} &= 2 + \frac{q}{2k_+^r k_+^w} \left(q \pm \sqrt{q^2 + 4k_+^r k_+^w} \right). \end{aligned} \quad (14)$$

We now study the sign of the derivative of the error $\eta(t)$. Clearly, its sign is the same as the sign of the derivative of the function

$$f(t) = \frac{w(t)}{r(t)} = \frac{w_{eq} + \delta w(t)}{r_{eq} + \delta r(t)}. \quad (15)$$

The derivative of $f(t)$, before any simplification, reads:

$$\begin{aligned} f' &= D^{-2} \left[2k_w^+(k_w^- - k_r^-)(k_w^- k_r^+ + k_r^-(k_w^- + k_w^+))e^{-(3\Sigma + \sqrt{Q})t/2} \right] \\ &\quad \left\{ -2Q^2 e^{(\Sigma + \sqrt{Q})t/2} + e^{\Sigma t} [k_-^{r^2} + k_-^{w^2} + k_-^r(-2k_-^w + 2k_+^r - 2k_w^p + \sqrt{Q}) \right. \\ &\quad + (k_+^r + k_+^w)(k_+^r + k_+^w + \sqrt{Q}) + k_-^w(-2k_+^r + 2k_+^w + \sqrt{Q})] \\ &\quad + e^{(\Sigma + \sqrt{Q})t} [k_-^{r^2} + k_-^{w^2} - k_-^w(-2k_+^w + 2k_+^r + \sqrt{Q}) \\ &\quad \left. - (k_+^r + k_+^w)(-k_+^r - k_+^w + \sqrt{Q}) + k_-^r(-2k_+^r + 2k_-^w + 2k_+^w + \sqrt{Q}) \right] \} \end{aligned} \quad (16)$$

where Σ and q are defined above, and we have also defined the function of the rates $Q = q^2 + 4k_+^r k_+^w$. The denominator D has a complicated expression that we omit since it is squared, hence it does not contribute to the sign. In the nominator, the term in square brackets clearly has the sign of $k_-^w - k_-^r$. We now move to the study the sign of the term in curly brackets, which can be expressed in terms of hyperbolic functions as

$$\{\} = e^{\frac{1}{2}(\Sigma + \sqrt{Q})t} \left(-2Q + 2e^{\frac{\Sigma}{2}t} \left[Q \cosh(\sqrt{Q}t/2) - \sqrt{Q}\Sigma \sinh(\sqrt{Q}t/2) \right] \right) \quad (17)$$

The prefactor is positive, and we are left with the two terms inside the parenthesis. The first is clearly negative. To determine the sign of the second, particularly the term in brackets, one should note that $\sinh(x) > \cosh(x)$ for all positive x , and that $\Sigma > \sqrt{Q}$, which is shown by expanding the squares. As a consequence, the second term in the brackets is always larger than the first, so that the whole term inside the parenthesis is negative. It follows that the term in the curly brackets is negative, and that the sign of f' is just the sign of $k_-^r - k_-^w$. From this we conclude that the error grows monotonically for all parameter choices such that $k_-^r - k_-^w$.

Finally, the rates are parametrized in the main text by their kinetic and energetic discriminations (δ and γ), the driving ϵ , and an overall time-scale ω . The kinetic discrimination δ appears in the forward rates, so that $k_+^r/k_+^w = e^{\delta}$. The driving ϵ is defined for right bases, so that $k_+^r/k_-^r = e^{\epsilon}$. Finally, the energetic discrimination γ reduces the driving of wrong bases, $k_+^w/k_-^w = e^{\epsilon - \gamma}$. Summarizing, we have:

$$k_+^r = \omega e^{\epsilon + \delta} ; k_-^r = \omega e^{\delta} ; k_+^w = \omega e^{\epsilon} ; k_-^w = \omega e^{\gamma}. \quad (18)$$

The condition $k_-^r - k_-^w$ is then equivalent to $\delta > \gamma$, which we termed the case of *kinetic discrimination*. Conversely, when $k_-^r < k_-^w$ (equivalent to $\delta < \gamma$) the error decreases monotonically (regime of *energetic discrimination*).

6.2 Co-polymerization model

Given the four rates k_+^w , k_-^w , k_+^r , and k_-^r , one can easily write the two rate equations as:

$$\begin{aligned}\partial_t[\&r] &= [\&]k_+^r - [\&r]k_-^r \\ \partial_t[\&w] &= [\&]k_+^w - [\&w]k_-^w\end{aligned}\tag{19}$$

Following Bennett's original approach [1], we consider the steady state in which there are constant fluxes of wrong $\partial_t[\&w] = J_w = \eta v[\&]$ and right $\partial_t[\&r] = J_r = (1 - \eta)v[\&]$ additions of aminoacids into the copied strain. Under these assumptions, one can show [1, 3] that the error is given by:

$$\frac{k_+^w - k_-^w\eta}{k_+^r - k_-^r(1 - \eta)} = \frac{\eta}{1 - \eta}.\tag{20}$$

while the average copying velocity is

$$v = k_+^w - k_-^w\eta + k_+^r - k_-^r(1 - \eta).\tag{21}$$

Using the same parametrization of the previous model, Eq. (18), we write the chemical driving ϵ as a function of the energetic discrimination energy γ , the kinetic discrimination energy δ , and the steady state error η . The expression reads

$$\epsilon = \log \left[\eta(1 - \eta) \frac{1 - e^{\gamma - \delta}}{\eta - (1 - \eta)e^{-\delta}} \right].\tag{22}$$

By means of (22), the average velocity can be expressed as

$$v = \omega \frac{1 - (1 + e^\gamma)\eta}{\eta - (1 - \eta)e^{-\delta}}.\tag{23}$$

We now want to impose that 1) the argument of the logarithm in (22) has to be positive, and 2) the average velocity (23) should also be positive. Assuming of course $0 < \eta < 1$, the first condition is equivalent to $(\delta - \gamma)[\eta - (1 + e^\delta)^{-1}] > 0$, while the second is equivalent to $[(1 - e^\gamma)^{-1} - \eta][\eta - (1 + e^\delta)^{-1}] > 0$. Combining these two conditions leads to Eq. (4) in the main text.

Finally, the dissipation per step is

$$\begin{aligned}\Delta S &= \eta \log \left[\frac{1}{\eta} \right] + (1 - \eta) \log \left[\frac{1}{1 - \eta} \right] + \eta \log \left[\frac{k_+^w}{k_-^w} \right] + (1 - \eta) \log \left[\frac{k_+^r}{k_-^r} \right] \\ &= \left(\eta \log \left[\frac{1}{\eta} \right] + (1 - \eta) \log \left[\frac{1}{1 - \eta} \right] \right) + (1 - \eta)\epsilon + \eta(\epsilon - \gamma),\end{aligned}\tag{24}$$

which can be expressed as a function of η , δ and γ only by using Eq. (22).

6.3 Mapping of Hopfield's original model

In Hopfield's formulation [2], given the template c , by interacting through C and D , either the aminoacid P_C or P_D can be added to an RNA chain. Addition of P_C will be the right addition, and addition of P_D will be considered an error. The rate equation and steady state solution are:

$$c + C \xrightleftharpoons[k_C]{k'_C} [cC] \xrightarrow{v} P_C \quad \text{and} \quad v[Cc] = k'_C[C] - k_C[Cc]\tag{25}$$

and analogously for D . It is assumed that $[C] \sim [D]$, and defined $f_C = [Cc]/[C]$ and $f_D = [Dc]/[D]$ as the fraction of incorporated C and D monomers given a template c . At steady state $f_C = 1 - f_D$, and f_D is the error $\eta = f_D$. Solving the system above we arrive at

$$\frac{\eta}{1 - \eta} = \frac{f_D}{f_C} = \frac{k'_D v + k_C}{k'_C v + k_D} \quad (26)$$

Identifying these rates with those in our model according to Fig. 1, the mapping to Hopfield's model is finished: $k_C = k_-^r$, $k'_C = k_+^r$, $k_D = k_-^w$ and $k'_D = k_+^w$.

To verify the mapping we study two limiting cases. For $\gamma = 0$ (as Bennett assumed in [1]) we have that if $v \rightarrow \infty$, then $\eta/(1 - \eta) \rightarrow k'_D/k'_C = e^{-\delta}$; and if $v \rightarrow 0$, then $\eta/(1 - \eta) \rightarrow k'_D k_C / k'_C k_D = 1$. On the other hand for $\delta = 0$ (as Hopfield assumed in [2]) we have that if $v \rightarrow \infty$, then $\eta/(1 - \eta) \rightarrow k'_D/k'_C = 1$; and if $v \rightarrow 0$, the classical result is obtained $\eta/(1 - \eta) \rightarrow k'_D k_C / k'_C k_D = k_C/k_D = e^{-\gamma}$, in exact agreement with the results obtained above.

6.4 Proofreading model without intermediate state

A minimal model of Kinetic Proofreading (KP) requires at least two different pathways. The first is the copying pathway introduced above, characterized by a driving which tends to make the chain grow. On the other hand, the driving of the second pathway is *backward*, thus undoing copies on average. The copying pathway has a bias towards adding right bases by a faster (kinetic) and more stable (energetic) binding. Conversely, the proofreading pathway has a bias towards removing wrong bases by a faster and less stable unbinding. The combination of both can reduce the minimal error given by the standard copy, by the discrimination factor of the proofreading pathway. The simplest proofreading scheme consists of the copying scheme introduced before, and a parallel reaction which we characterize by four additional proofreading rates $\tilde{k}_\pm^{r/w}$.

6.4.1 Rates parametrization

We choose the same copying rates of the standard copying scheme, see Eq.(18). Further, we introduce proofreading rates which are analogously characterized by a kinetic and energetic proofreading discrimination factors (δ_p and γ_p), a backward driving ϵ_p , and an additional time-scale ω_p . In the case of proofreading, we define the driving in the backward right additions, that is $\tilde{k}_-^r/\tilde{k}_+^r = e^{\epsilon_p}$. The kinetic discrimination is also backwards, and so $\tilde{k}_-^w/\tilde{k}_+^w = e^{\delta_p}$. Finally, the energetic discrimination is reflected in a higher backward driving of wrong bases, such that $\tilde{k}_-^w/\tilde{k}_+^w = e^{\epsilon_p + \gamma_p}$. One can then write the proofreading rates as

$$\tilde{k}_-^r = \omega_p e^{\epsilon_p - \delta_p} ; \quad \tilde{k}_+^r = \omega_p e^{-\delta_p} ; \quad \tilde{k}_-^w = \omega_p e^{\epsilon_p} ; \quad \tilde{k}_+^w = \omega_p e^{-\gamma_p}. \quad (27)$$

The energy levels corresponding to this parametrization of the rates are illustrated in Fig. (5). Notice that the end-states in the proofreading pathway have a difference in energy $\gamma - \gamma_p$. While in some coarse grained models such a behaviour may be justifiable through external agents, typically one would expect this difference not to exist, so that in the main text we always fixed $\gamma_p = \gamma$. Further, we anticipate that numerical results show that the proofreading step is always kinetic. This means that the value of γ_p , as soon as it is positive, will not anyway affect the minimum error achievable by the system.

6.4.2 Solving the model

The kinetic equations in this case are:

$$\begin{aligned}\partial_t[\&zr] &= [\&z](k_+^r + \tilde{k}_+^r) - [\&zr](k_-^r + \tilde{k}_-^r) \\ \partial_t[\&zw] &= [\&z](k_+^w + \tilde{k}_+^w) - [\&zw](k_-^w + \tilde{k}_-^w).\end{aligned}\quad (28)$$

Also in this case, the steady state solution can be obtained by considering the fluxes of right and wrong bases added: $\partial_t[\&zw] = J_w = \eta v[\&z]$ and right $\partial_t[\&zr] = J_r = (1 - \eta)v[\&z]$. The error as a function of the rates is analogous to the one for simple copying:

$$\frac{k_+^w + \tilde{k}_+^w - \eta(k_-^w + \tilde{k}_-^w)}{k_+^r + \tilde{k}_+^r - (1 - \eta)(k_-^r + \tilde{k}_-^r)} = \frac{\eta}{1 - \eta}. \quad (29)$$

The next step is to derive from this expression the driving ϵ as a function of the error, the discriminations, and the two new additional parameters: the proofreading driving ϵ_p and its characteristic time scale ω_p . The result is:

$$\epsilon = \log \left[\frac{1}{1 - \eta(1 + e^\delta)} \left\{ \eta(1 - \eta)(e^\gamma - e^\delta + \omega_p e^{\epsilon_p} - \omega_p e^{\epsilon_p - \delta_p}) - \omega_p e^{-\gamma_p} + \eta \omega_p (e^{-\gamma_p} + e^{-\delta_p}) \right\} \right] \quad (30)$$

The velocity is also analogous to that of the simple copying scheme:

$$v = k_+^w + \tilde{k}_+^w - \eta(k_-^w + \tilde{k}_-^w) + k_+^r + \tilde{k}_+^r - (1 - \eta)(k_-^r + \tilde{k}_-^r). \quad (31)$$

However, for the entropy production rate, one has to consider the transitions corresponding to the two pathways independently:

$$\begin{aligned}\dot{S} &= (k_+^w - \eta k_-^w) \log \left[\frac{k_+^w}{\eta k_-^w} \right] + (k_+^r - (1 - \eta)k_-^r) \log \left[\frac{k_+^r}{(1 - \eta)k_-^r} \right] \\ &+ (\tilde{k}_+^w - \eta \tilde{k}_-^w) \log \left[\frac{\tilde{k}_+^w}{\eta \tilde{k}_-^w} \right] + (\tilde{k}_+^r - (1 - \eta)\tilde{k}_-^r) \log \left[\frac{\tilde{k}_+^r}{(1 - \eta)\tilde{k}_-^r} \right].\end{aligned}\quad (32)$$

Finally, the dissipation per step is simply calculated as $\Delta S = \dot{S}/v$.

6.4.3 Minimization procedure and numerical results

For each given value of the error η and the four parameters γ , γ_p , δ , δ_p , we identified the values of the two remaining free parameters ω_p and ϵ_p corresponding to the minimum dissipation per step. In order to avoid local minima, we adopted a systematic minimization scheme: the two parameters have been varied with a logarithmic step

equal to 1.04, in an interval $10^{-5} < \omega_p, \epsilon_p < 10^9$. In this region, we found the minimum dissipation per step with the constraint of a positive reaction velocity. We also checked a posteriori that no minimum was found at the boundaries of the minimization region.

A systematic simulation study of the 4 possibilities of energetic/kinetic copy coupled to energetic/kinetic proofreading is presented in Fig. (6). The results allows us for reaching the following conclusions:

- The proofreading pathway can reduce the minimum error in the kinetic regime only. This can be seen in the lower panels of Fig. (6), where increasing γ_p does not affect the minimum achievable error. In particular, in the bottom left panel the copying is energetic and the minimum error is given by $f(\gamma) = e^{-\gamma}/(1 + e^{-\gamma}) \approx 0.0067$ for $\gamma = 5$. In the bottom right panel, the copying is kinetic and again the minimum error is given by $f(\delta) \approx 0.0067$ for $\delta = 5$. The minima in the two figures correspond to parameters such as the proofreading reactions has an average forward flux instead of backward, so that the proofreading pathway works as an effective parallel adiabatic (energetic) copying pathway.
- cooperative error reduction only takes place when both pathways are in the kinetic region. In the top right panel, increasing δ_p does not reduce the error. The only case in which the error can be reduced is in the kinetic-kinetic case of the top right panel, where the minimum error is given by $f(\delta)f(\delta_p) \approx f(\delta + \delta_p) \approx 0.0067, 0.0009, 0.0001$ for $\delta_p = 0, 2, 4$ respectively. We remark that this feature is a peculiarity of this model. We will show in the next section how including an intermediate state in the copying pathway allows for error reduction with an energetic copy and a kinetic proofreading.

6.5 Proofreading model with intermediate state

In this section we present more extensive results on model 4 of the main text. This model presents some analogies with the previous one, except that copying occurs via an intermediate state, denoted with a “*”, which is connected with the final state in which the aminoacid is incorporated. This final state has also a proofreading step. The overall reaction scheme is more in the spirit of Hopfield’s original proofreading mechanism.

6.5.1 Parametrization of the rates

The forward copying rates from the unbound to the intermediate state are defined in exactly the same way as the copying rates in the previous models, see Eq. (18). As in Hopfield’s original model, the transition rates from the intermediate state to the final state have no discrimination, but have their own driving ϵ^* and time scale given by ω^* . They obey the relations $\bar{k}_+^w/\bar{k}_+^r = 1$, $\bar{k}_+^w/\bar{k}_-^w = e^{\epsilon^*}$ and $\bar{k}_+^r/\bar{k}_-^r = e^{\epsilon^*}$. It is not hard to show that adding a discrimination below that of the original copying does not reduce the error beyond the critical error. Adding a bigger one simply reduces it to the critical error of this secondary copy, unlike the additive effect of proofreading. The rates can be simply written as:

$$\bar{k}_+^r = \omega^* e^{\epsilon^*} \quad ; \quad \bar{k}_-^r = \omega^* \quad ; \quad \bar{k}_+^w = \omega^* e^{\epsilon^*} \quad ; \quad \bar{k}_-^w = \omega^* \quad (33)$$

The final state is then connected with the initial state by the same proofreading rates defined in the previous section, Eq. (27). The full energy diagram is depicted

in Fig. (7). As before, the energy difference $\gamma - \gamma_p$ is irrelevant as the proofreading step has to be a kinetic step, and so we choose it arbitrarily to be null. Again, this corresponds to the physical requirement that the energy of the chain can not change if no base is added.

KP with no intermediate state

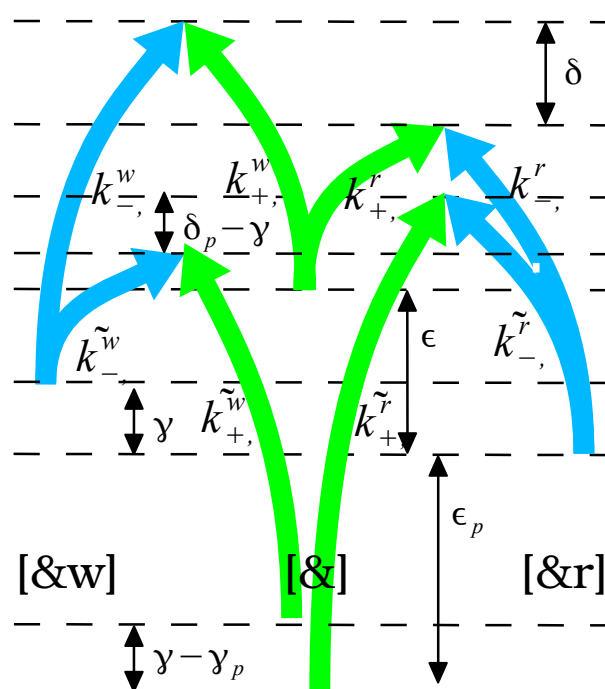


Fig. 5: Energy diagram of the reactions corresponding to the proofreading scheme with no intermediate steps.

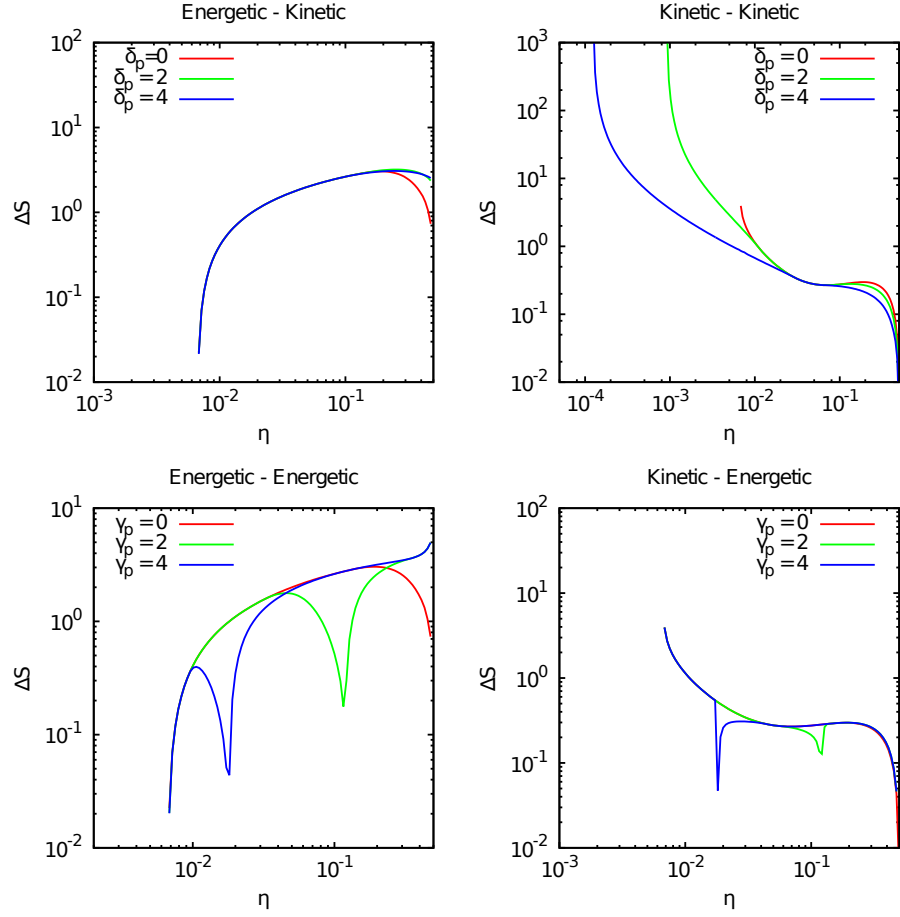


Fig. 6: Study of the four possible combination of energetic/kinetic discrimination and energetic/kinetic proofreading in Bennett's model. In the left panels the copy is energetic; in particular we chose $\delta = 0$ and $\gamma = 5$. Conversely, in the right panels the copy is kinetic with $\delta = 5$ and $\gamma = 0$. In the top panels the proofreading scheme is purely kinetic ($\gamma_p = 0$), while in the bottom panel we fixed $\delta_p = 0$ and varied γ_p .

Proofreading a la Hopfield

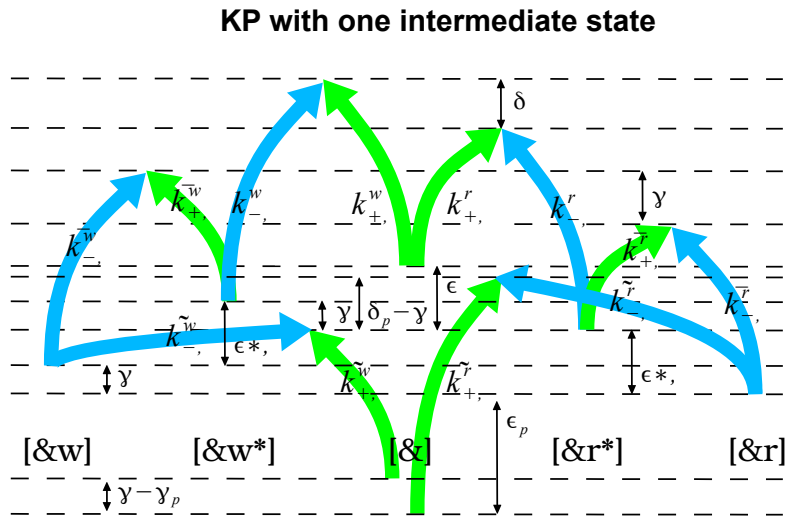


Fig. 7: Energy diagram of the reactions corresponding to the proofreading scheme with an intermediate step.

6.5.2 Solving the model

With the notation introduced in the previous section, it is easy to write the four kinetic equations of this proofreading scheme:

$$\begin{aligned}
\partial_t[R] &= [R^*]\bar{k}_+^r + [\&]\bar{k}_+^r - [R](\bar{k}_-^r + \tilde{k}_-^r) \\
\partial_t[W] &= [W^*]\bar{k}_+^w + [\&]\tilde{k}_+^w - [W](\bar{k}_-^w + \tilde{k}_-^w) \\
\partial_t[R^*] &= [\&]k_+^r + [R]\bar{k}_-^r - [R^*](k_-^r + \bar{k}_+^r) \\
\partial_t[W^*] &= [\&]k_-^w + [W]k_-^{w*} - [W^*](k_-^w + \bar{k}_+^w). \tag{34}
\end{aligned}$$

The easiest way to obtain the solution is by flux balance at the steady state of constant growth velocity v , which corresponds to:

$$\begin{aligned}
 [W]v &= ([\&]\tilde{k}_+^w - [W]\tilde{k}_-^w) + ([W^*]\tilde{k}_+^w - [W]\tilde{k}_-^w) \\
 [\&]k_+^w - [W^*]k_-^w &= [W^*]\tilde{k}_+^w - [W]\tilde{k}_-^w \\
 [R]v &= ([\&]\tilde{k}_+^r - [R]\tilde{k}_-^r) + ([R^*]\tilde{k}_+^r - [R]\tilde{k}_-^r) \\
 [\&]k_+^r - [R^*]k_-^r &= [R^*]\tilde{k}_+^r - [R]\tilde{k}_-^r.
 \end{aligned} \tag{35}$$

As before, we seek equations to determine the error rate and the velocity as a function of the rates. We proceed by dividing each of the equations in section by $\&$ and define $W/\& = \eta$, $R/\& = (1 - \eta)$, $W^*/\& = w^*$, $R^*/\& = r^*$. By means of the 2nd and 4th equations we find an expression for r^* and w^* :

$$\begin{aligned}
 w^* &= \frac{k_+^w + \eta\tilde{k}_-^w}{k_-^w + \tilde{k}_+^w} \\
 r^* &= \frac{k_+^r + (1 - \eta)\tilde{k}_-^r}{k_-^r + \tilde{k}_+^r}.
 \end{aligned} \tag{36}$$

Substituting into the other 2 equations lead to two coupled equations for η and v .

$$\begin{aligned}
 \eta v &= (\tilde{k}_+^w - \eta\tilde{k}_-^w) + \left[\tilde{k}_+^w \frac{k_+^w + \eta\tilde{k}_-^w}{k_-^w + \tilde{k}_+^w} - \eta\tilde{k}_-^w \right] \\
 (1 - \eta)v &= [\tilde{k}_+^r - (1 - \eta)\tilde{k}_-^r] + \left[\tilde{k}_+^r \frac{k_+^r + (1 - \eta)\tilde{k}_-^r}{k_-^r + \tilde{k}_+^r} - (1 - \eta)\tilde{k}_-^r \right].
 \end{aligned} \tag{37}$$

Now we multiply the first equation by $(1 - \eta)$, the second by η and subtract the second from the first to find a closed expression for η :

$$\begin{aligned}
 (1 - \eta)(\tilde{k}_+^w - \eta\tilde{k}_-^w) + (1 - \eta) \left[\tilde{k}_+^w \frac{k_+^w + \eta\tilde{k}_-^w}{k_-^w + \tilde{k}_+^w} - \eta\tilde{k}_-^w \right] \\
 - \eta[\tilde{k}_+^r - (1 - \eta)\tilde{k}_-^r] - \eta \left[\tilde{k}_+^r \frac{k_+^r + (1 - \eta)\tilde{k}_-^r}{k_-^r + \tilde{k}_+^r} - (1 - \eta)\tilde{k}_-^r \right] = 0.
 \end{aligned} \tag{38}$$

Again, this formula can be inverted to obtain the copying driving ϵ as a function of η and the other energy differences:

$$e^\epsilon = \frac{\eta\omega_p e^{-\delta_p} - (1 - \eta)\omega_p e^{-\gamma_p} + \eta(1 - \eta) \left(\omega_p e^{\epsilon_p} (1 - e^{-\delta_p}) + \frac{(\omega^*{}^2 e^{\epsilon^*})(e^\gamma - e^\delta)}{(e^\delta + \omega^* e^{\epsilon^*})(e^\gamma + \omega^* e^{\epsilon^*})} \right)}{\omega^* e^{\epsilon^*} \left(\frac{1 - \eta}{e^\gamma + \omega^* e^{\epsilon^*}} - \frac{\eta e^\delta}{e^\delta + \omega^* e^{\epsilon^*}} \right)}. \tag{39}$$

The velocity is straightforward to calculate from one of the expressions in (37), and is simply:

$$v = \left(\frac{\tilde{k}_+^w}{\eta} - \tilde{k}_-^w \right) + \left[\frac{\tilde{k}_+^w}{\eta} \frac{k_+^w + \eta\tilde{k}_-^w}{k_-^w + \tilde{k}_+^w} - \tilde{k}_-^w \right] \tag{40}$$

Finally, we calculate the entropy production by summing the six contributions of

the local fluxes of the system. This results in the following lengthy expression:

$$\begin{aligned}
&\&\dot{S} = (\&k_+^w - W^* \bar{k}_-^w) \log \left[\frac{\&k_+^w}{W^* \bar{k}_-^w} \right] + (W^* \bar{k}_+^w - W \bar{k}_-^w) \log \left[\frac{W^* \bar{k}_+^w}{W \bar{k}_-^w} \right] \\
&+ (\&\tilde{k}_+^w - W \bar{k}_-^w) \log \left[\frac{\&\tilde{k}_+^w}{W \bar{k}_-^w} \right] + (\&k_+^r - R^* \bar{k}_-^r) \log \left[\frac{\&k_+^r}{R^* \bar{k}_-^r} \right] \\
&+ (R^* \bar{k}_+^r - R \bar{k}_-^r) \log \left[\frac{R^* \bar{k}_+^r}{R \bar{k}_-^r} \right] + (\&\tilde{k}_+^r - R \bar{k}_-^r) \log \left[\frac{\&\tilde{k}_+^r}{R \bar{k}_-^r} \right]. \quad (41)
\end{aligned}$$

Dividing by $\&$ and using the expressions for r^* , w^* and η , we obtain the rate of entropy production:

$$\begin{aligned}
\dot{S} &= \left(k_+^w - \frac{k_+^w + \eta \bar{k}_-^w}{k_-^w + \bar{k}_+^w} k_-^w \right) \log \left[\frac{(k_-^w + \bar{k}_+^w) k_+^w}{(k_+^w + \eta \bar{k}_-^w) k_-^w} \right] \\
&+ \left(\frac{k_+^w + \eta \bar{k}_-^w}{k_-^w + \bar{k}_+^w} k_+^w - \eta \bar{k}_-^w \right) \log \left[\frac{(k_+^w + \eta \bar{k}_-^w) \bar{k}_+^w}{(k_-^w + \bar{k}_+^w) \eta \bar{k}_-^w} \right] \\
&+ (\tilde{k}_+^w - \eta \bar{k}_-^w) \log \left[\frac{\tilde{k}_+^w}{\eta \bar{k}_-^w} \right] \\
&+ \left(k_+^r - \frac{k_+^r + (1-\eta) \bar{k}_-^r}{k_-^r + \bar{k}_+^r} k_-^r \right) \log \left[\frac{(k_-^r + \bar{k}_+^r) k_+^r}{(k_+^r + (1-\eta) \bar{k}_-^r) k_-^r} \right] \\
&+ \left(\frac{k_+^r + (1-\eta) \bar{k}_-^r}{k_-^r + \bar{k}_+^r} k_+^r - (1-\eta) \bar{k}_-^r \right) \log \left[\frac{(k_+^r + (1-\eta) \bar{k}_-^r) \bar{k}_+^r}{(k_-^r + \bar{k}_+^r) (1-\eta) \bar{k}_-^r} \right] \\
&+ (\tilde{k}_+^r - (1-\eta) \bar{k}_-^r) \log \left[\frac{\tilde{k}_+^r}{(1-\eta) \bar{k}_-^r} \right]. \quad (42)
\end{aligned}$$

6.5.3 Minimization procedure and numerical results

In analogy with the previous model, for each value of the parameters δ , δ_p , γ and γ_p and the variable η we found the values of the free parameters corresponding to the minimum dissipation per step. In this case we had to minimize with respect to four free parameters: ω_p , ϵ_p , ω^* and ϵ^* . Given the number of parameters, we implemented a larger logarithmic minimization step, equal to 1.2.

The result of Fig. (8) are consistent to those of the previous model, see Fig. (6). The only important difference is:

- The presence of an additional step in the copying pathway allows for error reduction via an energetic copy - kinetic proofreading scheme. This can be seen in the top left panel of Fig. (8), where the minimum error does depend on δ_p via the usual function $f(\gamma)f(\delta_p) \approx f(\gamma + \delta_p)$. This is at variance with model 3, shown in Fig. (6), where the minimum error in the same case was simply equal to $f(\gamma)$.

References

- [1] C.H. Bennett. Dissipation-error tradeoff in proofreading. *Biosystems* 11:8 5–91 (1979).

-
- [2] J.J. Hopfield. Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity., *Proc. Natl. Acad. Sci. USA* 71:4135–4139 (1974).
 - [3] M. Esposito, K. Lindenberg and C. Van den Broeck. Extracting chemical energy by growing disorder: efficiency at maximum power. *J. Stat. Mech.*, doi:10.1088/1742-5468/2010/01/P01008 (2010).

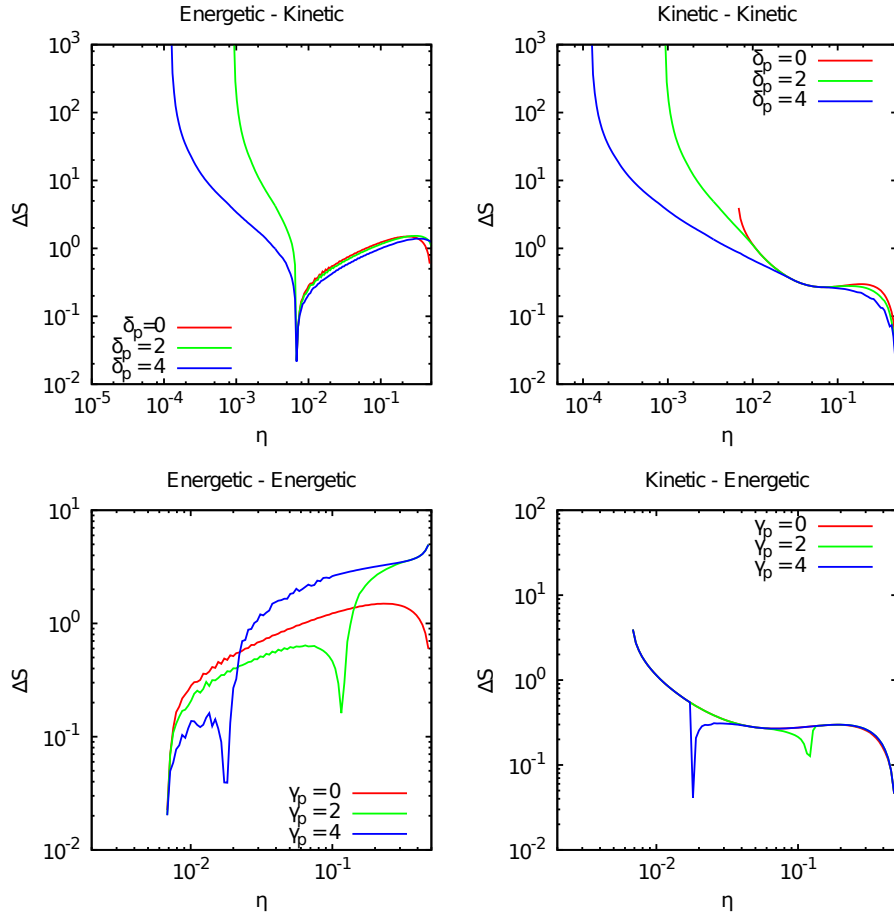


Fig. 8: Study of the four possible combination in the proofreading model with an intermediate state. In the left panels the copy is energetic; in particular we chose $\delta = 0$ and $\gamma = 5$. Conversely, in the right panels the copy is kinetic with $\delta = 5$ and $\gamma = 0$. In the top panels the proofreading scheme is purely kinetic ($\gamma_p = 0$), while in the bottom panel we fixed $\delta_p = 0$ and varied γ_p .