Fundamental Costs in the Production and Destruction of Persistent Polymer Copies

Thomas E. Ouldridge[1,*] and Pieter Rein ten Wolde[2]

[1]*Department of Bioengineering, Imperial College London, London SW7 2AZ, United Kingdom*
[2]*FOM Institute AMOLF, Science Park 104, 1098 XE Amsterdam, The Netherlands*

Producing a polymer copy of a polymer template is central to biology, and effective copies must persist after template separation. We show that this separation has three fundamental thermodynamic effects. First, polymer-template interactions do not contribute to overall reaction thermodynamics and hence cannot drive the process. Second, the equilibrium state of the copied polymer is template independent and so additional work is required to provide specificity. Finally, the mixing of copies from distinct templates makes correlations between template and copy sequences unexploitable, combining with copying inaccuracy to reduce the free energy stored in a polymer ensemble. These basic principles set limits on the underlying costs and resource requirements, and suggest design principles, for autonomous copying and replication in biological and synthetic systems.

Polymer copying is ubiquitous in living cells, occurring during replication, transcription, and translation. These processes yield two physically separated, sequence-related polymers from a single input [1]. Previous work has addressed the growth of a copy attached to a template [2–6], but these processes of templated self-assembly or templated polymerization do not directly produce *persistent* copies that are physically separated from their templates. Notably, while templated self-assembly has been realized in autonomous artificial systems [7–10], subsequent separation of copies without external manipulation has not. Similarly, a tendency to remain template bound has inhibited the generalization to polymers [11] of autocatalytic dimerization [12–14]. These difficulties emphasize that producing persistent copies involves more than just templated self-assembly.

We consider the fundamental thermodynamics of producing persistent copies, identifying the minimal work input through nonequilibrium free-energy changes. Eventual separation implies that, unlike in templated self-assembly, copy-template interactions cannot reduce the work required to produce a persistent copy. Moreover, a more accurate copy, which is more similar to its template, has a higher free energy and requires more work to create it. Different persistent copies produced from distinct templates can mix, however, rendering copy-template sequence correlations unexploitable and reducing the minimal work required for copying. Our analysis provides fundamental bounds on the efficiency of cellular recylcing networks and on the resource requirements for natural and artificial copying systems, while suggesting design principles for (autonomous) copying systems.

We consider a polymer template of N monomers, with m different monomer types of class A, which might be deoxyribonucleotides with $m = 4$. We label the whole polymer A, with a sequence vector $\mathbf{a}$ [Fig. 1(a)]. We then grow a polymer B from monomers of class B of m different types, with a sequence $\mathbf{b}$ that is a copy of $\mathbf{a}$. After the protocol, B is physically separated from A, as illustrated in Fig. 1(b). The sequences $\mathbf{a}$ and $\mathbf{b}$, and whether or not the two polymers are bound, together define a biochemical macrostate $\mathbf{y}$ of system Y. For a fixed sequence $\mathbf{a}$, the set of possible macrostates is then $\mathcal{B} = (\emptyset, \{1^*\}, \{2\}, \{2^*\}\ldots)$, where $\emptyset$ indicates no B polymer is present and no B monomers are bound to A, $\{n^*\}$ includes macrostates of all possible sequences of B of length n when bound to A, and $\{n\}$ includes all sequences of B of length n when unbound.

For our simple protocols we can work at the macrostate level. The work required to convert Y from a macrostate distribution $\phi(\mathbf{y})$ to $\phi'(\mathbf{y})$ is bounded by the nonequilibrium free energy difference [15,16]: $\langle w_{\phi \to \phi'} \rangle \geq \mathscr{F}[\phi'(\mathbf{y})] - \mathscr{F}[\phi(\mathbf{y})]$, with the equality holding for a reversible process, and $\mathscr{F}[\phi(\mathbf{y})] = \mathcal{U}[\phi(\mathbf{y})] - T\mathcal{S}[\phi(\mathbf{y})]$. Here, $\mathcal{U}[\phi(\mathbf{y})] = \sum_{\mathbf{y}} \phi(\mathbf{y}) U(\mathbf{y})$, and $\mathcal{S}[\phi(\mathbf{y})] = \sum_{\mathbf{y}} \phi(\mathbf{y}) S(\mathbf{y}) - k_B \sum_{\mathbf{y}} \phi(\mathbf{y}) \ln \phi(\mathbf{y})$, are the average energy
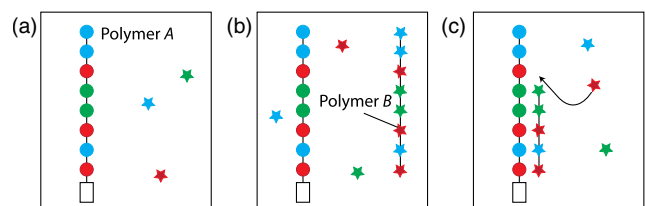

FIG. 1. Persistent copying of a polymer. (a) The initial state, with a polymer of class A monomers. The final state, with A unchanged and a second polymer of class B monomers. The copying protocol induces a sequence of B that is a copy of A, but no direct interactions are present in the final state. (c) Possible intermediate state, in which direct binding allows the sequence of A to influence the sequence of B as it grows.

and entropy, respectively. The average chemical free energy $\mathcal{F}[\phi(\mathbf{y})] = \sum_{\mathbf{y}} \phi(\mathbf{y}) F(\mathbf{y}) = \sum_{\mathbf{y}} \phi(\mathbf{y}) [U(\mathbf{y}) - TS(\mathbf{y})]$ incorporates the chemical energy and entropy of implicit microscopic degrees of freedom; the additional term $\mathcal{H}(Y) = \mathcal{H}[\phi(\mathbf{y})] = -k_B \sum_{\mathbf{y}} \phi(\mathbf{y}) \ln \phi(\mathbf{y})$ is the Shannon entropy of the macrostate distribution.

For our protocols, A is initially drawn from a sequence distribution $\phi(\mathbf{a}) = p(\mathbf{a})$, and B is absent (state $\emptyset$). At the end of the protocol, A is unchanged, but a persistent copy B is created with a sequence drawn from $\phi(\mathbf{b}|\mathbf{a}) = p_f(\mathbf{b}|\mathbf{a})$. The absence of A-B interactions in the initial and final states implies that the chemical free energy is a sum over separate contributions from A and B: $\mathcal{F}[\phi(\mathbf{a}, \mathbf{b})] = \mathcal{F}_A[\phi(\mathbf{a})] + \mathcal{F}_B[\phi(\mathbf{b})]$, with $\phi(\mathbf{b}) = \sum_{\mathbf{a}} \phi(\mathbf{a}) \phi(\mathbf{b}|\mathbf{a})$. However, the details of copying will generate sequence correlations [Fig. 1(c)], so the sequence entropy is not additive: $\mathcal{H}(A, B) = \mathcal{H}(A) + \mathcal{H}(B|A) = \mathcal{H}(A) + \mathcal{H}(B) - k_B \mathcal{I}(A; B)$ [16]. Here the conditional entropy $\mathcal{H}(B|A) = -k_B \sum_{\mathbf{a},\mathbf{b}} \phi(\mathbf{a}) \phi(\mathbf{b}|\mathbf{a}) \ln \phi(\mathbf{b}|\mathbf{a})$ is the average sequence entropy of B given A, and the mutual information $\mathcal{I}(A; B) = \sum_{\mathbf{a},\mathbf{b}} \phi(\mathbf{a}) \phi(\mathbf{b}|\mathbf{a}) \ln \phi(\mathbf{b}|\mathbf{a})/\phi(\mathbf{b}) \geq 0$ is the reduction in $\mathcal{H}(B)$ given knowledge of A. Since $\phi(\mathbf{a}) = p(\mathbf{a})$ is unchanged by the protocol, and $\mathcal{H}[p_0(\mathbf{b})] = 0$ for the initial B-distribution $p_0(\mathbf{b})$, the reversible work is

$$\langle w^f \rangle_{\text{rev}} = \mathcal{F}_B[p_f(\mathbf{b})] - \mathcal{F}_B[p_0(\mathbf{b})] - T\mathcal{H}_f(B)$$
$$+ k_B T \mathcal{I}_f(A; B), \tag{1}$$

with $\mathcal{H}_f(B) = \mathcal{H}[p_f(\mathbf{b})]$ and $\mathcal{I}_f(A; B) = \mathcal{I}[p_f(\mathbf{b}|\mathbf{a}), p(\mathbf{a})]$. Setting $\mathcal{F}_B[p_0(\mathbf{b})] = 0$ would be a valid normalization.

Previous studies on templated self-assembly have shown that favorable A-B interactions reduce the work required to assemble a polymer B on a template A [2,4–6]. Moreover, the presence of these interactions influences the equilibrium state of the B polymer, not only reducing the minimal work to grow a specific (desired) sequence, but also by providing a thermodynamic bias towards that sequence [4–6]. By contrast, the absence of A-B interactions after copy separation implies that the final free energy in a persistent copy process depends solely on interactions within B, and not with A. Thus both the reversible work $\langle w^f \rangle_{\text{rev}}$ and the equilibrium distribution $p_{\bar{N}}^{\text{eq}}(\mathbf{b})$ that minimizes $\mathcal{F}_B[\phi(\mathbf{b})] - T\mathcal{H}(B)$ for an average length $\bar{N}$ are A independent. Transitory binding during copying can neither reduce the overall work of copying, nor the relative cost of accurate versus inaccurate copying. Indeed, a protocol producing a template-specific $\phi(\mathbf{b}|\mathbf{a}) = p_f(\mathbf{b}|\mathbf{a})$ always requires more work than one yielding a template-independent equilibrium distribution with the same average length $\bar{N}_f$, $\phi(\mathbf{b}|\mathbf{a}) = p_{\bar{N}_f}^{\text{eq}}(\mathbf{b})$:

$$\langle w^f \rangle_{\text{rev}} - \langle w_{\bar{N}_f}^{\text{eq}} \rangle_{\text{rev}} = \mathcal{F}_B[p_f(\mathbf{b})] - \mathcal{F}_B[p_{\bar{N}_f}^{\text{eq}}(\mathbf{b})] + T[\mathcal{H}_{\bar{N}_f}^{\text{eq}}(B)$$
$$- \mathcal{H}_f(B)] + k_B T \mathcal{I}_f(A; B) \geq 0. \tag{2}$$

Here, we have used $\mathcal{I}_{\bar{N}_f}^{\text{eq}}(A; B) = 0$ for independent A and B. The inequality follows from $\langle w^f \rangle_{\text{rev}} = \mathcal{F}_B[p_f(\mathbf{b})] - \mathcal{F}_B[p_0(\mathbf{b})] - T\mathcal{H}_f(B) + k_B T \mathcal{I}_f(A; B) \geq \mathcal{F}_B[p_f(\mathbf{b})] - \mathcal{F}_B[p_0(\mathbf{b})] - T\mathcal{H}_f(B) \geq \mathcal{F}_B[p_{\bar{N}_f}^{\text{eq}}(\mathbf{b})] - \mathcal{F}_B[p_0(\mathbf{b})] - T\mathcal{H}_{\bar{N}}^{\text{eq}}(B) = \langle w_{\bar{N}_f}^{\text{eq}} \rangle_{\text{rev}}$. The lowest-cost output is template independent, with sequences drawn from $p_{\bar{N}}^{\text{eq}}(\mathbf{b})$. Template-specific persistent copies *necessarily* require more work because specific copies necessarily have higher free energy, unlike in templated self-assembly.

Neither $\langle w_{\bar{N}_f}^{\text{eq}} \rangle_{\text{rev}}$, nor $\langle w^f \rangle_{\text{rev}} - \langle w_{\bar{N}_f}^{\text{eq}} \rangle_{\text{rev}}$, are dissipated, but stored in the final free energy. Three terms contribute to $\langle w^f \rangle_{\text{rev}} - \langle w_{\bar{N}_f}^{\text{eq}} \rangle_{\text{rev}}$: a difference in chemical bonds within B, $\mathcal{F}_B[p_f(\mathbf{b})] - \mathcal{F}_B[p_{\bar{N}_f}^{\text{eq}}(\mathbf{b})]$; a difference in sequence entropy $\mathcal{H}_{\bar{N}_f}^{\text{eq}}(B) - \mathcal{H}_f(B)$; and $k_B T \mathcal{I}_f(A; B)$, reflecting the free energy stored in correlations [16–19], since noninteracting A and B are statistically independent in equilibrium. The first two terms can be individually positive or negative, but the third, and the sum, are necessarily non-negative. Combining the final two terms gives a single copying accuracy contribution, $T[\mathcal{H}_{\bar{N}}^{\text{eq}}(B) - \mathcal{H}_f(B)] + k_B T \mathcal{I}_f(A; B) = T[\mathcal{H}_{\bar{N}}^{\text{eq}}(B) - \mathcal{H}_f(B|A)]$. Perfect copying, with $\mathcal{H}_f(B|A) = 0$, has a large cost.

Despite not being dissipated, the minimal work required for accurate copying has implications for optimal replication. England used the total entropy increase as the replication cost, bounding it by the logarithm of the ratio of the replicator's birth and death rates [20]. Since this ratio can approach unity at an arbitrary net replication rate, there is no apparent minimal cost per replication. However, replication accuracy is absent in this analysis. Yet, replicators must make persistent copies, and our analysis shows that copy accuracy bounds the chemical work or resources required. Even if replication is reversible, generating zero total entropy, these resources cannot be recovered by the parent without reversing the copy and hence destroying the offspring. Thus, increased accuracy necessarily requires more resources that could be used elsewhere, such as to produce more offspring.

We illustrate a reversible copying protocol in Fig. 2. We nucleate B from a seed, to which an external force can be applied, and we manipulate the chemical potential of B-type monomers via a series of buffers [18]. To produce a single copy, B must only grow or shrink from its tip when in contact with A, and cannot grow beyond length $N = |\mathbf{a}|$; a catalyst could facilitate the desired reactions while keeping all others slow. We also assume that the ith monomer in A can only interact with the ith monomer in B. Though idealized, the system is thermodynamically valid since all reactions have a microscopic reverse.

Given $F(\mathbf{b})$, $\langle w^f \rangle_{\text{rev}}$ is calculable. Let the binding free energy of seeds be ΔF_s, and assume that adding a monomer x to an isolated B changes the chemical free energy of B by
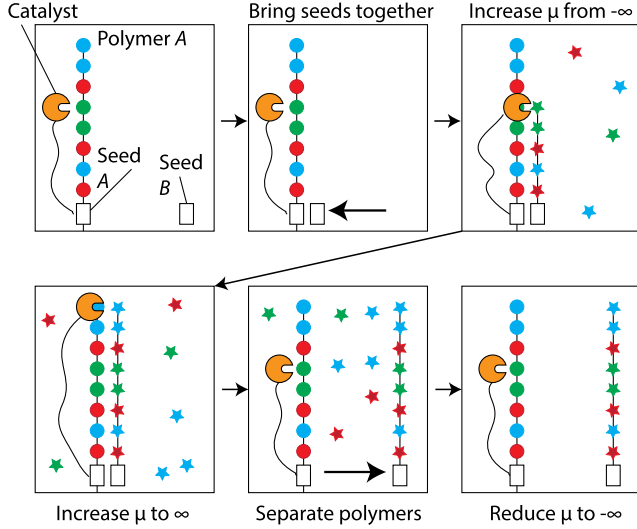
FIG. 2. A reversible protocol for persistent copying. Initially, seed B is separate from A, with B monomers present at a low chemical potential, $\mu \to -\infty$. The external force brings seed B into contact with A quasistatically, extracting work. The chemical potential μ of monomers is slowly raised, causing B to grow. Eventually, $\mu \to \infty$ and $|\mathbf{b}| = |\mathbf{a}|$. At this point, the external force separates the two polymers quasistatically, doing work against the binding free energy. Finally, the chemical potential of monomers is returned to its initial value.

ΔF_x. When in contact with A, ΔF_x is modified by ΔF_c for correct matches, and ΔF_{nc} otherwise. Mechanical work $\langle w_{\text{seed}} \rangle = \Delta F_s + C$ is extracted on bringing the seeds into contact (C reflects initial dilution). Chemical work is done during polymer growth, as the chemical potential of monomers is raised:

$$\langle w_{\text{pol}}(\mathbf{a}) \rangle = -k_B T \ln$$

$$\left(\sum_{\mathbf{b},|\mathbf{b}|=n} \prod_{x=1}^{N} e^{\frac{-\Delta F_{b_x}}{k_B T}} \left((1 - \delta_{a_x b_x}) e^{\frac{-\Delta F_{nc}}{k_B T}} + \delta_{a_x b_x} e^{\frac{-\Delta F_c}{k_B T}} \right) \right), \quad (3)$$

as shown in Sec. 1 of Ref. [21]. Separation requires mechanical work $\langle w_{\text{sep}}(\mathbf{a}) \rangle = -\Delta F_s - C - \Delta F_{AB}(\mathbf{a})$. Here, $\Delta F_{AB}(\mathbf{a})$ is the average contribution to the chemical free energy of polymerization from the A-B interaction,

$$\Delta F_{AB}(\mathbf{a}) = \sum_{\mathbf{b}} p_f(\mathbf{b}|\mathbf{a}) \sum_{x=1}^{|\mathbf{b}|} [(1 - \delta_{a_x b_x}) F_{nc} + \delta_{a_x b_x} F_c]. \quad (4)$$

The double-edged role of attractive interactions between A and B (negative F_c and F_{nc}) is evident. They reduce $\langle w_{\text{pol}}(\mathbf{a}) \rangle$, but provide a corresponding increase in $\langle w_{\text{sep}}(\mathbf{a}) \rangle$. Summing $\langle w_{\text{pol}}(\mathbf{a}) \rangle$, $\langle w_{\text{sep}}(\mathbf{a}) \rangle$, and $\langle w_{\text{seed}} \rangle$, and averaging over $p(\mathbf{a})$ (Sec. 2 of Ref. [21]), yields

$$\langle w^f \rangle = \sum_{\mathbf{b}} p_f(\mathbf{b}) \sum_{n=1}^{|\mathbf{b}|} \Delta F_{b_n} - T \mathcal{H}_f(B) + k_B T \mathcal{I}_f(A; B),$$

$$(5)$$

in which the dependence on F_c and F_{nc} has canceled. The first term is $\mathcal{F}_B[p_f(\mathbf{b})] - \mathcal{F}_B[p_0(\mathbf{b})]$, the change in average chemical free energy. Thus Eq. (5) matches Eq. (1), confirming that the protocol is reversible. Indeed, reversing the protocol recovers $\langle w^f \rangle$ and restores the initial state. A finite growth rate or nonequilibrium proofreading during the polymerization stage, as considered in Refs. [2–6], would lead to an increase in work over the minimum required by the output distribution $p_f(\mathbf{b}|\mathbf{a})$, $\langle w^f \rangle > \langle w^f \rangle_{\text{rev}}$.

Cells produce different persistent RNA and protein molecules from multiple distinct templates, and these copies subsequently mix. Motivated by this observation we now consider an ideal mixture of M persistent copies of a given set of templates. The copy macrostate is now specified by the numbers of each sequence present $\{M_{\mathbf{b}}\}$, with a distribution $\phi(\{M_{\mathbf{b}}\})$. The copies have free energy

$$\mathcal{F}[\phi(\{M_{\mathbf{b}}\})] = -k_B T \sum_{\{M_{\mathbf{b}}\}} \phi(\{M_{\mathbf{b}}\}) \ln \prod_{\mathbf{b}} \frac{Z_{\mathbf{b}}^{M_{\mathbf{b}}}}{M_{\mathbf{b}}!}$$

$$+ k_B T \sum_{\{M_{\mathbf{b}}\}} \phi(\{M_{\mathbf{b}}\}) \ln \phi(\{M_{\mathbf{b}}\}). \quad (6)$$

The first term is the average chemical free energy $\mathcal{F}_B[\phi(\{M_{\mathbf{b}}\})] = \sum_{\{M_{\mathbf{b}}\}} \phi(\{M_{\mathbf{b}}\}) F(\{M_{\mathbf{b}}\})$, and the second the macrostate entropy $-k_B T \mathcal{H}[\phi(\{M_{\mathbf{b}}\})]$. Here, $F(\{M_{\mathbf{b}}\}) = -k_B T \ln \prod_{\mathbf{b}} Z_{\mathbf{b}}^{M_{\mathbf{b}}}/M_{\mathbf{b}}!$ is the standard expression for dilute solutes with $-k_B T \ln Z_{\mathbf{b}}$ the chemical free energy of an isolated polymer [22]. For the simple model considered previously, $Z_{\mathbf{b}} = Z_0 \prod_{x=1}^{|\mathbf{b}|} e^{-\Delta F_{b_x}/k_B T}$, with $-k_B T \ln Z_0$ the free energy of an isolated seed.

To compare with our previous result, let each copied template be drawn from $p(\mathbf{a})$ (for an alternative, see Sec. 3 of Ref. [21]), giving $p_f(\mathbf{b}) = \sum_{\mathbf{a}} p(\mathbf{a}) p_f(\mathbf{b}|\mathbf{a})$. In this case, $\phi(\{M_{\mathbf{b}}\}) = M! \prod_{\mathbf{b}} p_f(\mathbf{b})^{M_{\mathbf{b}}}/M_{\mathbf{b}}!$. Substituting into Eq. (6) and using $\sum_{\{M_{\mathbf{b}}\}} \phi(\{M_{\mathbf{b}}\}) M_{\mathbf{b}} = \langle M_{\mathbf{b}} \rangle = M p_f(\mathbf{b})$, we obtain

$$\mathcal{F}[\phi(\{M_{\mathbf{b}}\})] = -k_B T M \sum_{\mathbf{b}} p_f(\mathbf{b}) \ln Z_{\mathbf{b}}$$

$$+ k_B T M \sum_{\mathbf{b}} p_f(\mathbf{b}) \ln p_f(\mathbf{b}) + k_B T \ln M!. $$

$$(7)$$

The first term is the average chemical free energy of M isolated copies, $M \mathcal{F}_B[p_f(\mathbf{b})]$, and the second the entropy $-T M \mathcal{H}_f(B)$. The third term is independent of the copying details. As before, $\mathcal{F}$ (and hence required work) is template independent, and is minimal for $p_f(\mathbf{b}) = p_{\bar{N}_f}^{\text{eq}}(\mathbf{b})$. Thus for many copies, $(\langle W^f \rangle_{\text{rev}} - \langle W_{\bar{N}_f}^{\text{eq}} \rangle_{\text{rev}})/M = \mathcal{F}_B[p_f(\mathbf{b})] - \mathcal{F}_B[p_{\bar{N}_f}^{\text{eq}}(\mathbf{b})] + T \mathcal{H}_{\bar{N}_f}^{\text{eq}}(B) - T \mathcal{H}_f(B)$. Absent is the $k_B T \mathcal{I}_f(A; B) \geq 0$ copy-template correlation term that is present in the single copy case [Eq. (2)]. Only the template-

averaged distribution $p_f(\mathbf{b})$ matters, and differences between copies of distinct templates are irrelevant.

Correlations do not contribute to $\mathcal{F}$ in the multicopy case due to mixing. When pairs of correlated noninteracting molecules are identifiable, as when copy-template pairs are isolated, the correlations are exploitable [19]. Once mixed, however, templates cannot be matched to copies *a priori*, and correlations cannot be leveraged. The stored free energy is no higher than if each template gave a nonspecific distribution $p_f(\mathbf{b}|\mathbf{a}) = p_f(\mathbf{b})$. If all templates have the same sequence, mixing copies has no effect, and the free energy is unchanged. Indeed, $\mathcal{I}_f(A;B) = 0$ in this case, since $\mathcal{H}(A) = 0$, and, hence, $(\langle W^f \rangle_{\mathrm{rev}} - \langle W^{\mathrm{eq}}_{\bar{N}_f} \rangle_{\mathrm{rev}})/M = \langle w^f \rangle_{\mathrm{rev}} - \langle w^{\mathrm{eq}}_{\bar{N}_f} \rangle_{\mathrm{rev}}$.

To reach the lower bound $\langle W^f \rangle_{\mathrm{rev}}$ on the work to produce a mixed ensemble $\phi(\{M_{\mathbf{b}}\}) = M! \prod_{\mathbf{b}} p_f(\mathbf{b})^{M_{\mathbf{b}}}/M_{\mathbf{b}}!$, a process must exploit the free energy released upon mixing —we outline such a protocol in Sec. 4 of Ref. [21]. If, instead, mixing simply occurred irreversibly after reversible copying, the entropy of the Universe would increase by the excess work $(\langle W^f \rangle - \Delta\mathcal{F}[\phi(\{M_{\mathbf{b}}\})])/T = k_B \mathcal{I}(A;B)$.

Cells recycle RNA and proteins via irreversible nonspecific depolymerization pathways [23], rather than by measuring sequences and depolymerizing with an appropriate template. In such *cyclic* operations, unlike replication, total entropy generation measures recycling inefficiency and is the natural metric for cost. The entropy generated in depolymerization sets a lower bound on the cost of the entire cycle. Bennett claimed that template-free depolymerization would generate at least $kT \ln m$ of entropy per monomer depolymerized, with m the number of distinct monomer types; other authors have found similar results [23–27]. However, these analyses consider a single initial sequence and hence underestimate the initial polymer entropy by assuming it is zero [24–27]. In reality the sequence entropy depends on the distribution of initial sequences, with a broader distribution implying a greater initial entropy.

For concreteness, consider the earlier model with M polymers, each with a sequence drawn from $p_z(\mathbf{b})$, giving a distribution of macrostates $\phi(\{M_{\mathbf{b}}\}) = M! \prod_{\mathbf{b}} p_z(\mathbf{b})^{M_{\mathbf{b}}}/M_{\mathbf{b}}!$. To depolymerize nonspecifically, we set $\mu = \bar{\mu}_z$ such that the average equilibrium length equals the average initial length $\bar{N}_z$ of polymers, and introduce catalysts that allow growth or shrinking. With this choice of μ there is no change in $\bar{N}$ when the catalysts are first introduced, and hence no chemical work since the net number of monomers transferred from the buffer is zero. Nonetheless, the distribution relaxes irreversibly to the equilibrium $\phi(\{M_{\mathbf{b}}\}) = M! \prod_{\mathbf{b}} p^{\mathrm{eq}}_{\bar{N}_z}(\mathbf{b})^{M_{\mathbf{b}}}/M_{\mathbf{b}}!$, generating entropy

$$
\begin{aligned}
T\Delta\sigma_{\mathrm{relax}} = &-k_B T M \sum_{\mathbf{b}} (p_z(\mathbf{b}) - p^{\mathrm{eq}}_{\bar{N}_z}(\mathbf{b})) \ln Z_{\mathbf{b}} \\
&+ k_B T M \mathcal{H}[p^{\mathrm{eq}}_{\bar{N}_z}(\mathbf{b})] - k_B T M \mathcal{H}[p_z(\mathbf{b})], \quad (8)
\end{aligned}
$$

using Eq. (7). Any other choice of initial μ would generate more entropy through unbalanced growth or shrinking. On taking $\mu \to -\infty$, the polymers shrink reversibly to zero, meaning that $T\Delta\sigma_{\mathrm{relax}} = T\Delta\sigma$ is the total increase in the entropy of the Universe during depolymerization.

We verify this dissipation for a specific model in Sec. 5 of Ref. [21]. For the special case in which ΔF_x is x independent, $\ln Z_{\mathbf{b}} \propto |\mathbf{b}|$ and thus as $\bar{N} \to \infty$, $T\Delta\sigma = k_B T M \bar{N}_z \ln m - k_B T M \mathcal{H}[p_z(\mathbf{b})]$, generalizing Bennett's result [23] to a distribution of input polymers. Thus the minimal entropy generation of nonspecific recycling depends on the details of the preceding production of persistent copies (Fig. 3). Nonspecific depolymerization is cheap if the polymers are drawn from a broad distribution due to inaccurate copying and/or a broad distribution of templates. For the biological case of high accuracy and a limited number of templates, the effect of nonzero $\mathcal{H}[p_z(\mathbf{b})]$ is small compared to $\bar{N}_z \ln m$.

Our analysis uses free-energy calculations, and the resulting bounds can only be reached by quasistatic operations. Our optimal protocol is nonautonomous, involving external manipulation. Nonetheless, it provides insight into autonomous copying in natural and synthetic systems. First, our results allow a meaningful definition of the efficiency of polymer copying, by comparing the work done to $\langle w^f \rangle_{\mathrm{rev}}$. Our analysis and its bounds provide a framework for the thermodynamics of producing persistent polymer copies, like the Carnot cycle does for heat engines. Recently, we have shown the relevance of a similar bound for the autonomous, finite-speed copying of a receptor by a biochemical network [18].
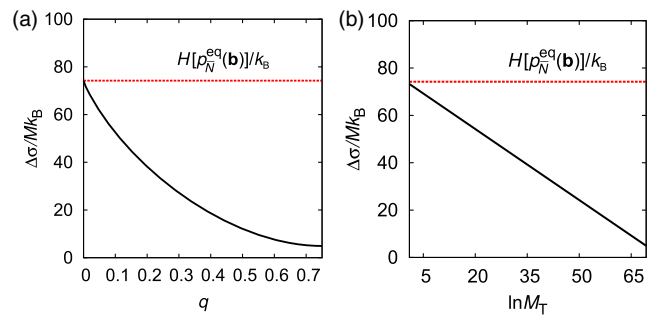

FIG. 3. Inaccurate copying and the presence of copies from multiple templates reduce the minimal entropy generation during nonspecific depolymerization. We plot $\Delta\sigma = \mathcal{H}[p^{\mathrm{eq}}_{\bar{N}_z}(\mathbf{b})] - \mathcal{H}[p_z(b)]$, the entropy generated by the nonspecific depolymerization protocol discussed in the text when all monomers are equally stable within B. We consider an ensemble of polymers all within initial length $\bar{N} = 50$ and with four distinct monomers ($m = 4$). (a) All copies produced from a single template, with an error rate of $q \leq (m-1)/m$ per monomer. (b) Copies produced with 100% accuracy and equal probability from M_T distinct templates, with $1 \leq M_T \leq 4^{\bar{N}}$. Neither graph reaches zero because the initial ensembles always contain a single polymer length.

Second, our results reveal fundamental differences between the optimal designs of copying networks and superficially similar self-assembling systems. Autonomous templated self-assembly can occur accurately and reversibly due to the equilibrium thermodynamic bias provided by favorable interactions between the matching monomers [4–6]. Indeed, quasireversible conditions are generally seen as optimal for self-assembly [28,29]. We show, however, that the minimal work to make persistent copies does not depend on template-copy interactions [Eq. (2)], which means that no equilibrium bias towards correct copying is possible. The fact that template-copy interactions are absent in the final state implies that these interactions can only provide specificity if they selectively stabilize the intermediate states of the copy process. For an autonomous and continuously operating system, this means that the template must act as a catalyst, providing specificity via kinetic discrimination (we discuss nonautonomous systems in Sec. S6 of Ref. [21]). Kinetic discrimination, however, requires that the system is driven out of thermodynamic equilibrium; we therefore predict that autonomous networks producing persistent copies *must* be nonspecific in the reversible limit, as seen for templated self-assembly when discrimination is based on kinetics rather than thermodynamics [2,5]. Dissipation in natural copying systems is therefore not only necessary to provide enhanced accuracy through proofreading [2,6,30], but to provide any accuracy at all. Synthetic copying networks should therefore be designed fundamentally differently from near-equilibrium self-assembling systems.

Finally, by highlighting the double-edged role of template-copy interactions, which enhance accurate polymerization but inhibit dissociation, our work draws attention to the differences between the distinct mechanisms that cells employ for persistent copying. Nature has two approaches. Viewing DNA replication at the level of the single strands, a copy is grown in contact with its template, and the cost of its separation is paid for after the copy is made in full (to enable the next replication). By contrast, in transcription and translation, the copy is only attached to the template by a handful of monomers at any one time; as new monomers join, older ones detach from the template. The importance of template-copy separation in terms of function and underlying thermodynamics suggests that the unique characteristics of these two distinct mechanisms warrant further consideration.

*t.ouldridge@imperial.ac.uk

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 4th ed. (Garland Science, New York, 2002).
[2] C. H. Bennett, BioSystems **11**, 85 (1979).
[3] F. Cady and H. Qian, Phys. Biol. **6**, 036011 (2009).
[4] D. Andrieux and P. Gaspard, Proc. Natl. Acad. Sci. U.S.A. **105**, 9516 (2008).
[5] P. Sartori and S. Pigolotti, Phys. Rev. Lett. **110**, 188101 (2013).
[6] P. Sartori and S. Pigolotti, Phys. Rev. X **5**, 041039 (2015).
[7] A. Luther, R. Brandsch, and G. von Kiedrowski, Nature (London) **396**, 245 (1998).
[8] J. Kim, J. Lee, S. Hamada, S. Murata, and S. H. Park, Nat. Nanotechnol. **10**, 528 (2015).
[9] J. W. Sadownik, E. Mattia, P. Nowak, and S. Otto, Nat. Chem. **8**, 264 (2016).
[10] R. Schulman, B. Yurke, and E. Winfree, Proc. Natl. Acad. Sci. U.S.A. **109**, 6405 (2012).
[11] L. E. Orgel, Nature (London) **358**, 203 (1992).
[12] D. Sievers and G. von Kiedrowski, Nature (London) **369**, 221 (1994).
[13] A. Vidonne and D. Philp, Eur. J. Org. Chem. **2009**, 593 (2009).
[14] T. A. Lincoln and G. F. Joyce, Science **323**, 1229 (2009).
[15] M. Esposito and C. Van den Broeck, Europhys. Lett. **95**, 40004 (2011).
[16] J. M. Parrondo, J. M. Horrowitz, and T. Sagawa, Nat. Phys. **11**, 131 (2015).
[17] J. M. Horowitz, T. Sagawa, and J. M. R. Parrondo, Phys. Rev. Lett. **111**, 010602 (2013).
[18] T. E. Ouldridge, C. C. Govern, and P. R. Wolde, Phys. Rev. X **7**, 021004 (2017).
[19] T. McGrath, N. S. Jones, P. R. ten Wolde, and T. E. Ouldridge, Phys. Rev. Lett. **118**, 028101 (2017).
[20] J. L. England, J. Chem. Phys. **139**, 121923 (2013).
[21] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.118.158103 for additional derivations.
[22] K. Huang, *Statistical Mechanics*, 2nd ed. (John Wiley & Sons, New York, 1987).
[23] C. H. Bennett, Int. J. Theor. Phys. **21**, 905 (1982).
[24] D. Andrieux and P. Gaspard, Europhys. Lett. **103**, 30004 (2013).
[25] P. Gaspard and D. Andrieux, J. Chem. Phys. **141**, 044908 (2014).
[26] P. Gaspard, Eur. Phys. J. Spec. Top. **224**, 825 (2015).
[27] P. Gaspard, J. Stat. Phys. **164**, 17 (2016).
[28] A. W. Wilber, J. P. K. Doye, A. A. Louis, E. G. Noya, M. A. Miller, and P. Wong, J. Chem. Phys. **127**, 085106 (2007).
[29] A. Reinhardt and D. Frenkel, Phys. Rev. Lett. **112**, 238103 (2014).
[30] J. J. Hopfield, Proc. Natl. Acad. Sci. USA **71**, 4135 (1974).