Computing and Bounding Equilibrium Concentrations in Athermic Chemical Systems

Hamidreza Akef ✉ ⓘ
The University of Texas at Austin, TX, USA

Minki Hhan ✉ ⌂ ⓘ
The University of Texas at Austin, TX, USA

David Soloveichik ✉ ⌂ ⓘ
The University of Texas at Austin, TX, USA

──── **Abstract** ────

Computing equilibrium concentrations of molecular complexes is generally analytically intractable and requires numerical approaches. In this work we focus on the polymer-monomer level, where indivisible molecules (monomers) combine to form complexes (polymers). Rather than employing free-energy parameters for each polymer, we focus on the athermic setting where all interactions preserve enthalpy. This setting aligns with the strongly bonded (domain-based) regime in DNA nanotechnology when strands can bind in different ways, but always with maximum overall bonding – and is consistent with the saturated configurations in the Thermodynamic Binding Networks (TBNs) model. Within this context, we develop an iterative algorithm for assigning polymer concentrations to satisfy detailed-balance, where on-target (desired) polymers are in high concentrations and off-target (undesired) polymers are in low. Even if not directly executed, our algorithm provides effective insights into upper bounds on concentration of off-target polymers, connecting combinatorial arguments about discrete configurations such as those in the TBN model to real-valued concentrations. We conclude with an application of our method to decreasing leak in DNA logic and signal propagation. Our results offer a new framework for design and verification of equilibrium concentrations when configurations are distinguished by entropic forces.

1 Introduction

In general, chemical equilibria of complex chemical systems are not analytically solvable and numerical tools are required for analysis. Such tools include NUPACK [17] for thermodynamic analysis of nucleic-acid systems, as well as more abstract platforms that support domain-level abstraction and free energy specification [12] including via rule-based modeling [11], and software for computing steady-state concentrations of chemical reaction networks [7, 10]. However, engineers often want a deeper understanding of the equilibrium than what analytically opaque numerical calculations can provide. Moreover, we often seek to understand infinite classes of designs such as logic circuits constructed from gate modules or parameterized constructions. For example, there is a growing body of work on leak in DNA-based systems,

exhibiting a family of schemes parameterized by a "redundancy parameter" meant to decrease leak arbitrarily at the cost of additional system components [13, 5, 15, 1, 16, 14]. It is proven that producing off-target species necessarily decreases the overall number of separate complexes – the change controlled by the redundancy parameter – and thus incurs an entropic penalty. However, due to system complexity, the relationship between this rigorously proven thermodynamic unfavorability and the actual concentrations of off-target species often remains implicit.

While properties of equilibria of abstract coupled chemical reactions have been extensively explored in the chemical reaction network theory literature (e.g., [6], Chapter 14, for detailed-balance equilibria), and full explicit-parameter schemes (e.g., [9]) can characterize the entire space of equilibria, these analytical approaches have severe limitations. For large systems the resulting explicit formulas are unwieldy and offer no effective guidance on how to choose the parameters so that off-target concentrations remain below a desired bound.

In this paper we are interested in the following problem: Our chemical species are complexes (termed polymers) made up of indivisible units called monomers. We assume there are finitely many possible polymers, and among these we are given a set of on-target species that we want to have desired equilibrium concentrations, and all other off-target species to have some sufficiently low equilibrium concentrations. Our task is to determine a consistent (detailed-balance) equilibrium, and therefore the concentrations of the monomers that would lead to this equilibrium. Note that our interest in setting equilibrium concentrations rather than solving for them (based on initial concentrations or total monomer concentrations, for example) is misaligned with most computational approaches. For instance, it can be seen as the reverse problem to NUPACK 3's `concentrations` tool which takes the concentrations of the monomers (strands) and returns the corresponding equilibrium concentrations of the polymers (complexes).

While having the flexibility to set monomer concentrations may appear to simplify the problem in the same way that finding *some* detailed-balance equilibrium is easier than computing the one consistent with input monomer concentrations, the complexity comes from simultaneously ensuring that off-target polymers remain in low concentration. Consider the natural approach of taking logarithms of all concentrations, thereby converting the detailed-balance equations (balancing each reaction) into a linear system amenable to standard linear solvers. When we fix the concentrations of on-target polymers, the remaining (off-target) concentrations are typically under-determined. The linear system then describes an unbounded affine subspace, and there is no obvious way to extract upper bounds on off-target species' concentrations without leaving the linear framework. Our solution is an iterative algorithm that assigns concentrations to off-target polymers in decreasing order of concentration, ensuring that each off-target species remains below a desired threshold concentration when possible. Importantly, terminating the algorithm at any iteration still provides valid upper bounds for all remaining off-target polymers.

In the most general formulation of the monomer-polymer equilibrium concentration problem, the polymer free energies can be assigned arbitrarily incorporating binding strength, geometric constraints, etc. In this work, we focus on the simpler *athermic* case rather than tackling the problem in its full generality. Our allowed polymers are such that all possible reactions between them are enthalpy-neutral. This model is consistent with systems of strong fully-complementary DNA domains in which domain-level bonds can only switch binding partners but not de-hybridize. Thermodynamic Binding Network (TBN) [5, 2] saturated configurations capture this condition, but our setting is more general without a built-in notion of domains (binding sites).

The main result of this paper is Algorithm 1 and Theorem 5.4 showing how starting with desired concentrations of on-target polymers (already in detailed balance), we can set the concentrations of off-target polymers to satisfy detailed balance and thus thermodynamic equilibrium. In Section 4 we explain the apparent difficulties in balancing reactions which our approach needed to overcome.

If, rather than computing exact concentrations of off-target polymers, it is sufficient to bound them, then we refer the reader to Section 6. In Section 7 we apply our framework to the analysis of systems in the TBN model specifically, connecting the combinatorial notions of stability and entropy loss in the TBN model to equilibrium concentrations. In Section 8, we show applications of our method to the analysis of a simple TBN AND gate, as well as a parameterized family of signal propagation systems (translator cascades) from prior work. For the translator cascade, we argue that tuning concentrations of on-target polymers according to our framework is essential for leak to decrease exponentially with the redundancy parameter. We conclude with a discussion of future work (Section 9), including a formulation of new combinatorial conditions in the TBN model to make our framework easily applicable.

2   Model

Let $\mathbb{N}$ denote the set of nonnegative integers. Given a finite set $\mathcal{A}$, we define $\mathbb{N}^{\mathcal{A}}$ as the set of functions $f : \mathcal{A} \to \mathbb{N}$.

A multiset $\mathcal{M}$ over the finite set $\mathcal{A}$ is described by its *counting function* $f_{\mathcal{M}} \in \mathbb{N}^{\mathcal{A}}$, where for each element $a \in \mathcal{A}$, the value $f_{\mathcal{M}}(a)$ indicates how many times a appears in $\mathcal{M}$. We often write $\mathcal{M} \in \mathbb{N}^{\mathcal{A}}$ to mean that $\mathcal{M}$ is a multiset over $\mathcal{A}$, and we denote the count of $a \in \mathcal{A}$ in $\mathcal{M}$ by $\mathcal{M}[a]$. The notation $a \in \mathcal{M}$ means that $\mathcal{M}[a] \geq 1$. The *cardinality* of a multiset $\mathcal{M}$, denoted $|\mathcal{M}|$, is the total number of elements in the multiset $|\mathcal{M}| = \sum_{a \in \mathcal{A}} \mathcal{M}[a]$. For two multisets $\mathcal{M}$ and $\mathcal{M}'$ over $\mathcal{A}$, we define their union $\mathcal{M} + \mathcal{M}'$ as the multiset whose counting function is the pointwise sum $(\mathcal{M} + \mathcal{M}')(a) = \mathcal{M}[a] + \mathcal{M}'[a]$ for all $a \in \mathcal{A}$. For example, let $\mathcal{M} = \{a, a, b, c\}$. Then, $\mathcal{M}[a] = 2$, $\mathcal{M}[b] = 1$, $\mathcal{M}[c] = 1$, and $\mathcal{M}[d] = 0$ for all $d \notin \{a, b, c\}$. Also, the cardinality of $\mathcal{M}$ is $|\mathcal{M}| = 2 + 1 + 1 = 4$. Finally, note that $\mathcal{M}$ could also be written as the union $\{a, b\} + \{a, c\}$, for example.

The linear combination of multisets with nonnegative integers is defined analogously: For multisets $\mathcal{M}_1, ..., \mathcal{M}_n$ and $a_1, ..., a_n \in \mathbb{N}$, $\sum_{i=1}^{n} a_i \cdot \mathcal{M}_i$ corresponds to the counting function $\sum_{i=1}^{n} a_i \cdot \mathcal{M}_i[a]$. Let $\mathcal{M}_1, \mathcal{M}_2 \in \mathbb{N}^{\mathcal{S}}$ be two multisets over the same set $\mathcal{S}$. The difference $\mathcal{M}_1 - \mathcal{M}_2$ is defined as the multiset $\mathcal{M} \in \mathbb{N}^{\mathcal{S}}$ such that for every $a \in \mathcal{S}$, $(\mathcal{M}_1 - \mathcal{M}_2)[a] = \mathcal{M}_1[a] - \mathcal{M}_2[a]$ provided that $\mathcal{M}_1[a] \geq \mathcal{M}_2[a]$ for all $a \in \mathcal{S}$.

We also define the intersection of a multiset with a *set*. Given a multiset $\mathcal{M}$ over $\mathcal{A}$ and a subset $\mathcal{S} \subset \mathcal{A}$, the intersection $\mathcal{M} \cap \mathcal{S}$ is a *multiset* over $\mathcal{A}$ defined by $(\mathcal{M} \cap \mathcal{S})[a] = \mathcal{M}[a]$ if $a \in \mathcal{S}$, otherwise $(\mathcal{M} \cap \mathcal{S})[a] = 0$. For instance, $\{a, a, b, c\} \cap \{a, c\} = \{a, a, c\}$.

The main object of this paper is an abstract model of monomers and polymers, motivated by systems in DNA nanotechnology. This model captures how simple indivisible molecules (monomers) combine to form complexes (polymers) under specific physical and chemical constraints.

▶ **Definition 2.1.** *Let $\mathbf{\Psi^0}$ be a finite set of* monomers, *and $\mathbf{\Psi} \subseteq \mathbb{N}^{\mathbf{\Psi^0}}$ be a finite set of* polymers *over these monomers, where each polymer $P \in \mathbf{\Psi}$ is a multiset of monomers.*

Let $\mathbf{x^0} \in (0, 1)^{\mathbf{\Psi^0}}$ represent the vector of concentrations for all monomers, and let $\mathbf{x} \in (0, 1)^{\mathbf{\Psi}}$ represent the vector of concentrations for all polymers (also called configuration). The relationship between monomer and polymer concentrations is governed by mass conservation.

Specifically, we require

$$\mathbf{x^0} = \mathbf{A} \cdot \mathbf{x} \tag{1}$$

where $\mathbf{A} \in \mathbb{N}^{|\mathbf{\Psi^0}| \times |\mathbf{\Psi}|}$ is a matrix such that each entry A_{ij} specifies the number of monomers of type i in polymer j.

For example, the polymer $P = \{m_1, m_1, m_2, m_3\}$ contains two copies of m_1, one of m_2, and one of m_3. Note that we will be interested in cases where the set of polymers of interest $\mathbf{\Psi}$ is a finite (proper) subset of all possible polymers over $\mathbf{\Psi^0}$.

In DNA nanotechnology the monomers are typically DNA strands with different sequences. Polymers are analogous to a multistranded DNA structure composed of multiple DNA strands. We use the term polymer rather than "complex" in order to better emphasize their composition from monomers and to be consistent with the TBN literature.

To model the equilibrium behavior of such systems, we use the free energy formulation in the notation of Dirks et al. [4], where the equilibrium concentrations are obtained by minimizing the following free energy function (corresponding to the pseudo-Helmholtz free energy used throughout chemical reaction network theory literature [8, 6]):

$$\mathbf{g}(\mathbf{x}) = \sum_{P \in \mathbf{\Psi}} x_P (\log x_P - \log \Omega_P - 1) \tag{2}$$

where x_P denotes the concentration of polymer P, and Ω_P is its partition function corresponding to the exponential of the polymer's negative free energy. The minimization is subject to the mass conservation constraint given in Equation (1).

In this work, we focus on *athermic* systems where all interactions are equally favored enthalpically. Thus we assume that $\sum_{P \in \mathbf{\Psi}} x_P \cdot \log \Omega_P$ is constant for all configurations $\mathbf{x}$ satisfying mass conservation (Equation (1)), yielding a simplified cost function that is entirely entropic:

$$g(\mathbf{x}) = \sum_{P \in \mathbf{\Psi}} x_P (\log x_P - 1). \tag{3}$$

This function serves as the objective of our optimization problem. Its minimizer under the constraint of Equation (1) corresponds to the equilibrium concentration of polymers.

Understanding the equilibrium requires us to formalize how polymers can transform into one another. We do so by defining reconfigurations and the reactions they induce.

▶ **Definition 2.2.** *Two multisets of polymers $\mathcal{M}_1, \mathcal{M}_2 \in \mathbb{N}^{\mathbf{\Psi}}$ are reconfigurations of each other, written $\mathcal{M}_1 \cong \mathcal{M}_2$, if for every monomer $m \in \mathbf{\Psi^0}$, the total count of m is the same in $\mathcal{M}_1$ as in $\mathcal{M}_2$; i.e., $\sum_{P \in \mathbf{\Psi}} \mathcal{M}_1[P] \cdot P[m] = \sum_{P \in \mathbf{\Psi}} \mathcal{M}_2[P] \cdot P[m]$. Whenever $\mathcal{M}_1 \cong \mathcal{M}_2$, we also define reaction $\mathcal{M}_1 \to \mathcal{M}_2$.*

We occasionally use Greek alphabets to denote the reactions. Note that while $\cong$ is a binary relation, a reaction $\alpha : \mathcal{M}_1 \to \mathcal{M}_2$ is an ordered pair representing a directed transformation between the multisets. Our arguments will refer to *reactants* (left-hand side) and *products* (right-hand side) of particular reactions.[1]

An important property of the minimum of Equations (2) and (3) is that it satisfies detailed balance over reactions [8, 6]. For us (Equation (3)), for any single reaction $\alpha : \mathcal{M}_1 \to \mathcal{M}_2$, the equilibrium concentrations satisfy:

$$\prod_{P \in \mathcal{M}_1} x_P^{\mathcal{M}_1[P]} = \prod_{P \in \mathcal{M}_2} x_P^{\mathcal{M}_2[P]}.$$

---

[1] All reactions are of course reversible, but we treat each direction as a separate reaction.

We say that a reaction α is balanced when this equality holds. This notion of balance will be central to our characterization of equilibrium concentrations formed by on-target polymers, which we explore next.

The fact that balance leads to the minimum of the free energy is well-established, and we present its sketch in Section B for the completeness of the paper.

3 Characterizing On-target Polymers

We now define the set $\mathcal{S}$ of on-target polymers – intuitively, these are the high-concentration polymers whose equilibrium concentration we set as input to our algorithm.

▶ **Definition 3.1.** *Let $\mathcal{S} \subseteq \boldsymbol{\Psi}$ be a set of polymers, and let $\mu : \mathcal{S} \to (0, 1]$ be a function assigning a* concentration exponent *to each polymer in $\mathcal{S}$. We say that $\mathcal{S}$ is* on-target *with concentration exponents μ if:*
1. *For every polymer $P \in \boldsymbol{\Psi}$, there are multisets of polymers $\mathcal{M} \in \mathbb{N}^{\mathcal{S}}$ and $\mathcal{M}' \in \mathbb{N}^{\boldsymbol{\Psi}}$ where $\mathcal{M}'$ contains P such that $\mathcal{M} \cong \mathcal{M}'$.*
2. *For any two multisets $\mathcal{M}_1, \mathcal{M}_2 \in \mathbb{N}^{\mathcal{S}}$, if $\mathcal{M}_1 \cong \mathcal{M}_2$, then their concentration exponents are equal $\mu(\mathcal{M}_1) = \mu(\mathcal{M}_2)$. Here, the concentration exponent of a multiset $\mathcal{M}$ is defined as $\mu(\mathcal{M}) = \sum_{P \in \mathcal{M}} \mathcal{M}[P] \cdot \mu(P)$.*

While presented in terms of restricting $\mathcal{S}$, another (perhaps more apropos) interpretation of the first condition is that we are restricting $\boldsymbol{\Psi}$ to be only those polymers that can be obtained via a reconfiguration of polymers over $\mathcal{S}$. The interpretation of the second condition is that every reaction among polymers in $\mathcal{S}$ is in detailed balance. More precisely:

▶ Remark 3.2. Let $0 < c < 1$, and suppose every on-target polymer $P \in \mathcal{S}$ has concentration $c^{\mu(P)}$. Then every reaction $\mathcal{M}_1 \to \mathcal{M}_2$, where $\mathcal{M}_1, \mathcal{M}_2 \in \mathbb{N}^{\mathcal{S}}$, satisfies $c^{\mu(\mathcal{M}_1)} = c^{\mu(\mathcal{M}_2)}$ and thus the product of the concentrations of the left-hand side polymers is equal to the product of the concentrations of the right-hand side polymers.[2][3]

If $\mathcal{S} = \boldsymbol{\Psi}$ then we are done, being guaranteed that all reactions are in detailed balance. However, the case of interest is where we do not know the concentration exponents of all polymers – making it the goal of this paper to compute them or bound them to ensure that the equilibrium concentration of off-target polymers is small.

We now define canonical reactions, which are the reactions with reactants from $\mathcal{S}$. These reactions can generate polymers outside of $\mathcal{S}$ from polymers in $\mathcal{S}$. Each canonical reaction has key quantities we call imbalance and novelty which will play a key role in our algorithm.

▶ **Definition 3.3.** *For an on-target set $\mathcal{S}$ with μ, a reaction $\alpha : \mathcal{M}_1 \to \mathcal{M}_2$ is called* canonical *if all its reactants are over $\mathcal{S}$ (i.e., $\mathcal{M}_1 \in \mathbb{N}^{\mathcal{S}}$). Then the* imbalance *of α is defined as $k(\alpha) := \mu(\mathcal{M}_1) - \mu(\hat{\mathcal{M}_2})$, where $\hat{\mathcal{M}_2} := \mathcal{M}_2 \cap \mathcal{S}$. The* novelty *of α is defined by $l(\alpha) := |\mathcal{M}_2 - \hat{\mathcal{M}_2}|$.*

Given the definition of multiset subtraction, the novelty $l(\alpha)$ is always non-negative.

Intuitively, the imbalance of the reaction represents how far from detailed balance it is *if we consider only the polymers whose concentrations are known.* The novelty is the number of off-target (new) polymers produced by the same reaction.

---

[2] Note that we use symbol μ for concentration exponents because of the direct parallel to the standard notion of chemical potential, which is expressed with logarithmic terms of concentration: chemical potential $\mu = \mu^\circ + RT \ln(x)$, where x is the mole fraction.

[3] Our base concentration c is smaller than one because the units of concentration are mole fractions – the ratio of polymers to solvent molecules, and the derivation of the energy function Equation (2) only holds in the regime of less polymer than solvent [4].

Intuitively, a large $k(\alpha)$ means a larger bias of the reaction toward its reactants, i.e., elements of $\mathcal{S}$, prior to assigning concentration exponents to off-target polymers. This is desired since it implies less pressure to make off-target polymers and gives us more room to maneuver in assigning concentrations to them.

Intuitively, a larger novelty $l(\alpha)$ means that the canonical reaction is more entropically favored to its products. Note that the term $\sum_{P \in \Psi} x_P \log x_P$ in Equation (3) is minimized when the concentration is more "spread out" over the different species (like Shannon entropy), and thus there is an entropic force to generate new polymers. A large $l(\alpha)$ is thus undesired. As justified in the subsequent results, the ratio $k(\alpha)/l(\alpha)$ captures the effective tradeoff between imbalance and novelty.

Note that summing canonical reactions gives another canonical reaction. While on the one hand this leads to infinitely many canonical reactions, on the other hand, this additive property allows us to analyze the set of all canonical reactions using a Hilbert basis of *finite size*. We explain how to employ the Hilbert basis in our algorithm to avoid the infinity of canonical reactions in Section A.

The notion of stability of $\mathcal{S}$, defined below, captures the idea that on-target polymers are in higher concentration compared to off-target polymers. Recall that concentration exponents μ are ≤ 1 for polymers in $\mathcal{S}$ (Definition 3.1). As shown later by Theorem 5.4, the following definition ensures that all concentration exponents of polymers outside of $\mathcal{S}$ are greater than 1 by our algorithm. Since the base concentration $c < 1$, this implies that off-target polymers are in smaller concentration.

▶ **Definition 3.4.** *The on-target set $\mathcal{S}$ with μ is called* stable *if, for every canonical reaction α with $l(\alpha) \neq 0$, the ratio $k(\alpha)/l(\alpha) > 1$.*

While our formalism allows different concentration exponents for different polymers in $\mathcal{S}$, nearly all insight can be obtained from the simple case of uniform on-target concentration exponents:

▶ **Definition 3.5.** *The on-target set $\mathcal{S}$ with μ is called* uniform *if every polymer $P \in \mathcal{S}$ has concentration exponent $\mu(P) = 1$.*

4 Why Balancing is Nontrivial

Our main goal is to ensure that in equilibrium the concentration of on-target polymers is much higher than the others. This section presents two examples illustrating why this is nontrivial.

Consider a uniform set of on-target polymers $\mathcal{S} = \{A, B, C, D\}$, aiming at the equilibrium concentration $x_A = x_B = x_C = x_D = c$. For two off-target polymers P_1 and P_2, suppose that we figured out a canonical reaction

$$\alpha : 2A + B + C \rightarrow P_1 + P_2.$$

This reaction is in detailed balance if and only if $x_{P_1} \cdot x_{P_2} = x_A^2 \cdot x_B \cdot x_C = c^4$. At this point, it is unclear how to balance x_{P_1} and x_{P_2} without inspecting the other reactions: For example, the reaction α is balanced by assigning $x_{P_1} = x_{P_2} = c^2$, but another canonical reaction $A + 2B + 2D \rightarrow 2P_1$ would not be if it exists.

The following example exhibits a different potential problem. While Remark 3.2 shows that reactions entirely within $\mathcal{S}$ are balanced, it is not clear that given our choice of concentration exponents μ for $\mathcal{S}$, the reactions involving polymers outside of $\mathcal{S}$ could be balanced at all. Suppose there are two reactions

$$\beta : A + B \rightarrow P_3 \text{ and } \gamma : B + C + D \rightarrow 2P_3.$$

These reactions intuitively say that the off-target polymer P_3 is non-interacting to other polymers in these reactions so that the above problem of balancing the concentrations over off-target polymers is absent.[4] However, balancing the reaction β suggests the concentration $x_{P_3} = c^2$ at equilibrium, but the reaction γ gives $x_{P_3} = c^{1.5}$. In other words, it is unclear if the concentrations of each of the polymers in $\mathcal{S}$ can be the same or even close to each other.

Despite these difficulties, our main result guarantees that the configuration on on-target polymers we demand is in equilibrium without the above issue, and can be *extended* to the configuration over all polymers at equilibrium, and ensures that the off-target polymers remain in very low concentration at equilibrium. Intuitively, we prove that there exists a canonical reaction for which we can assign concentrations to product polymers without creating conflicts with other reactions, and we provide a method to find this reaction. Moreover, as we will see later, these assignments occur in order of decreasing concentration allowing us to restrict the concentration of off-target species.

5 Main Result: Concentration of Off-target Polymers

For the remainder, we fix $\mathbf{\Psi^0}$, $\mathbf{\Psi}$, and a particular set of on-target polymers $\mathcal{S}$ with concentration exponents μ.

To systematically assign concentration exponents to all polymers, we will organize them into hierarchical groups called levels. The process begins with a designated set of polymers $\mathcal{S}$, the on-target polymers, which are assigned initial concentration exponents via $\mu : \mathcal{S} \to \mathbb{R}^+$.

All other polymers, called off-target polymers, will be partitioned into level sets $\mathcal{S}_1, \mathcal{S}_2, \ldots$, constructed inductively based on how these polymers appear in certain canonical reactions. At each level i, we will compute a scalar value μ_i that serves as the concentration exponent for all polymers newly added at that level.

In this way, we gradually extend the initial function μ to a global function $\bar{\mu} : \mathbf{\Psi} \to \mathbb{R}^+$ that assigns a concentration exponent to every polymer in the system. This inductive process and the precise definitions of μ_i, $\mathcal{S}_i$, and $\bar{\mu}$ will be described in detail below.

▶ **Definition 5.1.** *Given $\mathcal{S}_0, ..., \mathcal{S}_{i-1}$, assume $\bar{\mu}(P)$ is assigned for any $P \in \bigcup_{j=0}^{i-1} \mathcal{S}_j$. For a canonical reaction $\alpha : \mathcal{M}_1 \to \mathcal{M}_2$, we define $\hat{\mathcal{M}}_2 := \mathcal{M}_2 \cap \left(\bigcup_{j=0}^{i-1} \mathcal{S}_j \right)$. The ith-level imbalance of α is defined as $k_i(\alpha) := \bar{\mu}(\mathcal{M}_1) - \bar{\mu}(\hat{\mathcal{M}}_2)$, and the ith-level novelty by $l_i(\alpha) := |\mathcal{M}_2| - |\hat{\mathcal{M}}_2|$.*

While $\hat{\mathcal{M}}_2$ technically depends on the level index i, we suppress this dependency in the notation for simplicity; it will be clear from the context that it is updated at each level. Additionally, by the definition of a canonical reaction, we have $\bar{\mu}(\mathcal{M}_1) = \mu(\mathcal{M}_1)$ in the expression above.

▶ **Definition 5.2.** *Let $\mu_i = \min_\alpha \{k_i(\alpha)/l_i(\alpha)\}$ where the minimum is taken over all canonical reactions α with $l_i(\alpha) \neq 0$.[5] The canonical reactions achieving the minimum are termed i-levelizing canonical reactions. The ith level set $\mathcal{S}_i$ is defined as the set of all polymers P appearing in any i-levelizing reactions not already in $\bigcup_{j=0}^{i-1} \mathcal{S}_j$. Every polymer $P \in \mathcal{S}_i$ is assigned concentration exponent $\bar{\mu}(P) := \mu_i$.*

At the heart of this inductive construction is the requirement that each i-levelizing reaction maintains balance with respect to the assigned concentration exponents. That is, for every reaction $\alpha : \mathcal{M}_1 \to \mathcal{M}_2$ used to define level $\mathcal{S}_i$, the assigned exponents ensure the

---

[4] Looking ahead, these non-interacting polymers are indeed easier to assign concentrations to as shown in Section 5.1.

[5] The minimum can actually be taken over a finite subset of canonical reactions using Hilbert basis. We refer the reader to Section A for more detail.

reaction remains detailed balanced, $c^{\bar{\mu}(\mathcal{M}_1)} = c^{\bar{\mu}(\mathcal{M}_2)}$. This holds because each polymer in $\mathcal{M}_2$ that already appeared in previous levels contributes its already-defined exponent, while newly introduced polymers in $\mathcal{S}_i$ all receive the same value μ_i. Decomposing $\mathcal{M}_2$ into previously levelized polymers $\hat{\mathcal{M}}_2$ and new polymers in $\mathcal{S}_i$, we obtain:

$$\bar{\mu}(\mathcal{M}_1) = \bar{\mu}(\hat{\mathcal{M}}_2) + k_i(\alpha) = \bar{\mu}(\hat{\mathcal{M}}_2) + \mu_i \cdot l_i(\alpha) = \bar{\mu}(\hat{\mathcal{M}}_2) + \mu_i \cdot (|\mathcal{M}_2| - |\hat{\mathcal{M}}_2|) = \bar{\mu}(\mathcal{M}_2),$$

ensuring that the total concentration exponent on both sides of the reaction is equal. This confirms that the assignment of μ_i for polymers in level $\mathcal{S}_i$ is consistent with the balance requirements dictated by the underlying reactions.

The full procedure for determining the level sets $\mathcal{S}_i$ and the corresponding concentration exponents μ_i is formally specified in Definitions 5.1 and 5.2 and carried out through the step-by-step process described in Algorithm 1. This algorithm inductively builds the extended concentration exponent function $\bar{\mu}$ by identifying polymers that can be levelized at each step and assigning them an appropriate exponent value based on their participation in canonical reactions.

While the algorithm itself does not explicitly state a stopping condition, its termination is ensured by the structure of the level construction process. Each time a new level $\mathcal{S}_i$ is defined, it includes at least one polymer not present in any previous level. Since the set of all polymers $\mathbf{\Psi}$ is finite by assumption, only finitely many levels can be introduced. As a result, the inductive process must terminate after assigning a level and concentration exponent to each polymer, thus completing the definition of the extended function $\bar{\mu}$.

For us, canonical reactions represent the simplest meaningful interactions and serve as building blocks for more complex behavior. Their central role for our results is motivated by the following lemma:

▶ **Lemma 5.3.** *Let* $\mathbf{x^0} \in (0,1)^{\mathbf{\Psi^0}}$ *be a vector of concentrations of the monomers. If all canonical reactions are balanced at configuration* $\mathbf{x} \in (0,1)^{\mathbf{\Psi}}$, *then the cost function* $g(\mathbf{x})$ *is minimum subject to* $\mathbf{A} \cdot \mathbf{x} = \mathbf{x^0}$.

Proof. We will show that any arbitrary reaction can be decomposed into a canonical reaction. Consequently, if detailed balance holds for all canonical reactions, it follows that detailed balance holds for all reactions. Per Section B, the cost function g reaches its minimum – under the constraint $\mathbf{A} \cdot \mathbf{x} = \mathbf{x^0}$ – when all reactions are balanced. This confirms the statement of the lemma.

Consider an arbitrary non-canonical reaction $\alpha : \mathcal{M}_1 + P \rightarrow \mathcal{M}_2$ with $P \notin \mathcal{S}$ where $\mathcal{M}_1 + P$ denotes the union of two multisets $\mathcal{M}_1$ and $\{P\}$. According to the first condition of Definition 3.1, there exists a canonical reaction $\beta : \mathcal{M}_1' \rightarrow \mathcal{M}_2' + P$ that produces P. Now, apply β and α sequentially on the reactants $\mathcal{M}_1 + \mathcal{M}_1'$. This gives rise to a new reaction:

$$\gamma : \mathcal{M}_1 + \mathcal{M}_1' \rightarrow \mathcal{M}_1 + (\mathcal{M}_2' + P) = (\mathcal{M}_1 + P) + \mathcal{M}_2' \rightarrow \mathcal{M}_2 + \mathcal{M}_2'.$$

Therefore, the reaction α, when catalyzed by $\mathcal{M}_2'$ (i.e., adding $\mathcal{M}_2'$ to reactants and products) and using the inverse of the canonical reaction β, can be replaced by the reaction γ, which involves fewer reactant polymers outside of $\mathcal{S}$ than α did originally:

$$(\mathcal{M}_1 + P) + \mathcal{M}_2' = \mathcal{M}_1 + (\mathcal{M}_2' + P) \rightarrow \mathcal{M}_1 + \mathcal{M}_1' \rightarrow \mathcal{M}_2 + \mathcal{M}_2'.$$

By repeating this procedure, α decomposes into a canonical reaction. This concludes the proof. ◀

With this groundwork in place, we now state our main theorem, which characterizes the equilibrium distribution of the entire polymer system, both on-target and off-target, in terms of their assigned levels.

▶ **Theorem 5.4.** *Let $\mathcal{S}$ be the stable set of on-target polymers with concentration exponents $\mu : \mathcal{S} \to (0, 1]$. For the extended concentration exponent $\bar{\mu} : \Psi \to \mathbb{R}^+$ generated by Algorithm 1 and for any $0 < c < 1$, there are monomer concentrations $\mathbf{x^0} \in (0, 1)^{\Psi^0}$ such that the configuration $\mathbf{x} \in (0, 1)^{\Psi}$ with each polymer $P \in \Psi$ at concentration $c^{\bar{\mu}(P)}$ is the minimum of $g(\mathbf{x})$ subject to $\mathbf{A} \cdot \mathbf{x} = \mathbf{x^0}$ (i.e., Equations (1) and (3)). Furthermore, $\bar{\mu}(P) \geq \mu_1 > 1$ for all $P \notin \mathcal{S}$.*

In this configuration, every off-target polymer has strictly lower concentration than any on-target polymer. This is because concentrations scale exponentially with the extended exponent $\bar{\mu}(P)$, and off-target polymers are assigned strictly larger exponents than the shared minimal value μ of the on-target set $\mathcal{S}$. As a result, for $0 < c < 1$, off-target concentrations are exponentially suppressed.

■ **Algorithm 1** Calculating level sets and concentration exponents. The repeat loop will terminate because there are finitely many polymers. The Hilbert basis implementation avoiding the infinite set Λ is discussed in Section A.

1: **Input:** A set of level-0 polymers $\mathcal{S}_0 = \mathcal{S}$ with $\mu(\cdot)$ and all canonical reactions Λ
2: **Output:** Sets of level-i polymers $\mathcal{S}_i$ and concentration exponents $\bar{\mu}(\cdot)$
3: Set $i \leftarrow 0$, $\mathcal{S}_j \leftarrow \emptyset$ for all $j \in \mathbb{N}$
4: **repeat**
5:    $i \leftarrow i + 1$
6:    **for** each canonical reaction $\alpha : \mathcal{M}_1 \to \mathcal{M}_2 \in \Lambda$ **do**
7:       $\hat{\mathcal{M}_2} \leftarrow$ the multiset of all level $\leq (i-1)$ polymers in $\mathcal{M}_2$.
8:       $k_i(\alpha) \leftarrow \bar{\mu}(\mathcal{M}_1) - \bar{\mu}(\hat{\mathcal{M}_2})$ and $l_i(\alpha) \leftarrow |\mathcal{M}_2| - |\hat{\mathcal{M}_2}|$
9:    Compute $\mu_i = \min_\alpha \{k_i(\alpha)/l_i(\alpha)\}$ where min is taken over $\alpha \in \Lambda$ with $l_i(\alpha) \neq 0$.
10:    **for** each $\alpha : \mathcal{M}_1 \to \mathcal{M}_2 \in \Lambda$ such that $\mu_i = k_i(\alpha)/l_i(\alpha)$ **do**
11:       Append α to the set of i-levelizing reactions
12:       Append all polymers P in $\mathcal{M}_2$ that are not in $\hat{\mathcal{M}_2}$ to $\mathcal{S}_i$ and assigns $\bar{\mu}(P) = \mu_i$
13: **until** All polymers are included in $\cup_{j=0}^i \mathcal{S}_j$
14: **return** The sets $\mathcal{S}_i$ and the concentration exponents $\bar{\mu}(\cdot)$

Proof of Theorem 5.4. We use a proof by contradiction to show that each canonical reaction must be balanced in the configuration with $[P] = c^{\bar{\mu}(P)}$, which suffices to conclude the proof thanks to Lemma 5.3.

Suppose that there exists a canonical reaction $\alpha : \mathcal{M}_1 \to \mathcal{M}_2$. We consider two cases $\bar{\mu}(\mathcal{M}_1) < \bar{\mu}(\mathcal{M}_2)$ or $\bar{\mu}(\mathcal{M}_1) > \bar{\mu}(\mathcal{M}_2)$. Recall all polymers are levelized.

Case 1. $\bar{\mu}(\mathcal{M}_1) < \bar{\mu}(\mathcal{M}_2)$. Let t be the top level among the polymers in $\mathcal{M}_2$. Note that $\bar{\mu}(P) = \mu_t$ for each $P \in \mathcal{S}_t$. Since μ_t is defined as the minimum of $k_t(\alpha')/l_t(\alpha')$ over all canonical reactions $\alpha' \in \Lambda$ with $l_t(\alpha') \neq 0$, we must have:

$$\mu_t \leq \frac{k_t(\alpha)}{l_t(\alpha)} = \frac{\bar{\mu}(\mathcal{M}_1) - \bar{\mu}(\hat{\mathcal{M}_2})}{l_t(\alpha)} < \frac{\bar{\mu}(\mathcal{M}_2) - \bar{\mu}(\hat{\mathcal{M}_2})}{l_t(\alpha)} = \mu_t$$

which is a contradiction.

Case 2. $\bar{\mu}(\mathcal{M}_1) > \bar{\mu}(\mathcal{M}_2)$. Let β_P be the levelizing reaction including P as product for each polymer $P \in \mathcal{M}_2$. Consider a canonical reaction $\sum_P \mathcal{M}_2[P] \cdot \beta_P : \mathcal{M}_1' \to \mathcal{M}_2' + \mathcal{M}_2$ obtained by summing $\mathcal{M}_2[P]$ copies of β_P which includes $\mathcal{M}_2$ as products.[6] By applying $\mathcal{M}_2 \to \mathcal{M}_1$ to this reaction, we have a canonical reaction $\beta : \mathcal{M}_1' \to \mathcal{M}_2' + \mathcal{M}_2 \to \mathcal{M}_2' + \mathcal{M}_1$ such that

$$\bar{\mu}(\mathcal{M}_1') = \bar{\mu}(\mathcal{M}_2') + \bar{\mu}(\mathcal{M}_2) < \bar{\mu}(\mathcal{M}_2') + \bar{\mu}(\mathcal{M}_1).$$

Therefore this case is reduced to **Case 1**, leading once again to a contradiction.

The "Furthermore" part is proven in Corollary 6.4.   ◄

5.1   Non-Interacting Off-target Polymers

Our main theorem implies that if the on-target polymers $\mathcal{S}$ and the corresponding concentration exponents μ are chosen appropriately, the equilibrium *exists* with the extended concentration exponents $\bar{\mu}$ consistent with μ. In this section, we show that $\bar{\mu}$ exponents for *non-interacting* off-target polymers P can be assigned in a particularly easy way. Non-interacting off-target polymers are those that can be independently made by a canonical reaction:

▶ **Corollary 5.5.** *Let $\mathcal{S}$ be a stable set of on-target polymers. Suppose that an off-target polymer $P \notin \mathcal{S}$ is a product of a canonical reaction $\alpha_P : \mathcal{M}_P \to \mathcal{M}_P' + P$ for some multisets $\mathcal{M}_P$ and $\mathcal{M}_P'$ from $\mathcal{S}$.[7] Then $\bar{\mu}(P) = \mu(\mathcal{M}_P) - \mu(\mathcal{M}_P')$.*

Proof. By Theorem 5.4, the reaction α_P is in detailed balance, $c^{\bar{\mu}(\mathcal{M}_P)} = c^{\bar{\mu}(\mathcal{M}_P'+P)}$, and we take the logarithm of both sides.   ◄

The same idea extends further. At any point of the execution of the algorithm, suppose that the extended concentration exponents of the set of polymers $\bar{\mathcal{S}}$ have been already assigned. For a non-interacting polymer P outside of $\bar{\mathcal{S}}$ with the reaction $\alpha_P : \mathcal{M}_P \to \mathcal{M}_P' + P$ for multisets $\mathcal{M}_P$ and $\mathcal{M}_P'$ from $\bar{\mathcal{S}}$, we can assign $\bar{\mu}(P) = \bar{\mu}(\mathcal{M}_P) - \bar{\mu}(\mathcal{M}_P')$. This assignment is valid by the same reason to the corollary.

▶ **Example 5.6.** Consider the set of monomers and polymers given by $\mathbf{\Psi^0} = \{a, b, c\}$ and $\mathbf{\Psi} = \{A, B, C, X, Y, Z\}$; where $A = \{a, a\}$, $B = \{a, b\}$, $C = \{c\}$, $X = \{a, a, a, b\}$, $Y = \{b, b, c, c\}$, and $Z = \{b, b, c, c, c\}$. Consider the uniform on-target set $\mathcal{S} = \{A, B, C\}$. Instead of inspecting all canonical reactions, we can focus on three canonical reactions

$$\alpha : A + B \to X, \quad \beta : 3B + 2C \to X + Y, \quad \gamma : A + 2B + 3C \to X + Z.$$

Here X is the non-interacting off-target polymer in α, thus $\bar{\mu}(X) = 2$. After assigning $\bar{\mu}(X)$, Y and Z become non-interacting off-target polymers in β and γ, respectively, so that we derive $\bar{\mu}(Y) = 3$ and $\bar{\mu}(Z) = 4$.

6   Framework for Upper-Bounding Off-target Polymers

Often we are only interested in an upper bound on the concentration of an off-target undesired polymer. In this case, rather than generating its exact equilibrium concentration via Algorithm 1, we can more efficiently compute an upper bound by narrowing our focus to reactions that directly produce P instead of examining the full set of canonical reactions. This leads to a simpler surrogate quantity that approximates $\bar{\mu}(P)$ from below.

---

[6] Here we use the linear combination of reactions in a standard way; for example, for $\alpha : \mathcal{N}_1 \to \mathcal{N}_2$ and $\beta : \mathcal{N}_1' \to \mathcal{N}_2'$, the reaction $2 \cdot \alpha + \beta$ denotes the reaction $2\mathcal{N}_1 + \mathcal{N}_1' \to 2\mathcal{N}_2 + \mathcal{N}_2'$.

[7] Note that there may be other reactions involving P with other off-target polymers.

▶ **Definition 6.1.** *For $P \notin \cup_{j=0}^{i-1} \mathcal{S}_j$, let $\widetilde{\mu}_i(P) = \min_\alpha \{k_i(\alpha)/l_i(\alpha)\}$ where the minimum is taken over all canonical reactions α that include P as a product.*

Since P has not yet been levelized by construction, any such reaction α producing P, obviously involves at least one unlevelized polymer, ensuring $l_i(\alpha) \neq 0$.

Intuitively, $\widetilde{\mu}_i(P)$ captures a conservative estimate of the exponent with which polymer P can grow in concentration at level i, based only on reactions that actually produce P. Since it is defined as a minimum over a subset of possible reactions, it is easier to compute than the full concentration exponent $\bar{\mu}(P)$, yet is still meaningful for analysis:

▶ **Theorem 6.2.** *For any polymer $P \notin \cup_{j=0}^{i-1} \mathcal{S}_j$, $\bar{\mu}(P) \geq \widetilde{\mu}_i(P) \geq \mu_i$.*

▶ **Lemma 6.3.** *For a canonical reaction α with $l_{i+1}(\alpha) \neq 0$, it holds that $k_i(\alpha)/l_i(\alpha) \leq k_{i+1}(\alpha)/l_{i+1}(\alpha)$.*

Proof. Note that $\mu_i \leq k_i(\alpha)/l_i(\alpha)$ by definition of μ_i. Let n be the count of level-i product polymers in α, which must be smaller than $l_{i+1}(\alpha)$. Then we have

$$\frac{k_{i+1}(\alpha)}{l_{i+1}(\alpha)} = \frac{k_i(\alpha) - n\mu_i}{l_i(\alpha) - n} = \frac{k_i(\alpha) - n\mu_i}{l_i(\alpha) - n} \geq \frac{k_i(\alpha) - n(k_i(\alpha)/l_i(\alpha))}{l_i(\alpha) - n} = \frac{k_i(\alpha)}{l_i(\alpha)}. \qquad ◀$$

Proof of Theorem 6.2. Since all polymers will eventually be levelized, there exists a j-levelizing canonical reaction α that includes P as a product for some $j \geq i$. Then, $\widetilde{\mu}_i(P) \leq k_i(\alpha)/l_i(\alpha) \leq k_j(\alpha)/l_j(\alpha) = \bar{\mu}(P)$ where the first inequality follows from Definition 6.1, and the second follows from Lemma 6.3.

Furthermore, the value $\widetilde{\mu}_i(P)$ is computed by taking the minimum of $k_i(\alpha)/l_i(\alpha)$ only over canonical reactions that produce P (and have $l_i(\alpha) \neq 0$). In contrast, μ_i is defined as the minimum over all canonical reactions, regardless of whether they produce P or not. In other words, the set of reactions considered when computing $\widetilde{\mu}_i(P)$ is a subset of those considered for μ_i, making the search space for $\widetilde{\mu}_i(P)$ more restricted. Since a minimum over a smaller set cannot be smaller than the minimum over a larger set, we conclude that $\widetilde{\mu}_i(P) \geq \mu_i$. ◀

To summarize, $\widetilde{\mu}_i(P)$ serves as a computationally efficient lower bound on $\bar{\mu}(P)$. When used alongside Theorem 5.4, this upper bounds the equilibrium concentration of P.

The usefulness of Theorem 6.2 extends beyond individual estimates. It contributes to structural insights about level assignments during the iterations of our algorithm. Specifically, after levelizing up through S_{i-1}, we know that any unlevelized polymer $P \notin \cup_{j=0}^{i-1} \mathcal{S}_j$ satisfies $\bar{\mu}(P) \geq \mu_i > \mu_{i-1}$. This inequality ensures that polymers awaiting level assignment must exhibit smaller concentrations. In particular, because the first level imbalance k_1 and novelty l_1 are precisely identical to k and l defined in Definition 3.3, we derive $\bar{\mu}(P) \geq \mu_i \geq \min_\alpha \{k(\alpha)/l(\alpha)\} > 1$ and the following corollary.

▶ **Corollary 6.4.** *$\bar{\mu}(P) > 1$ for all $P \notin \mathcal{S}$.*

7 Concentrations of Polymers in the TBN Model

While the TBN model is combinatorial in nature, quantifying over discrete (saturated) configurations, at the end we are most often interested in determining real-valued concentrations, which are accessible to (bulk) wet-lab experiment. The framework developed in this paper helps to bridge this gap between combinatorics of discrete configurations and concentrations.[8]

---

[8] Of course, much of the heavy lifting in bridging this gap is done by the derivation [4] of the cost function $g(\mathbf{x})$, but our work expands on it beyond numerical simulation.

Consider a saturated (i.e., maximally bonded) configuration $\mathcal{M}$ in the TBN model. Quantified over all saturated reconfigurations $\mathcal{M}' \cong \mathcal{M}$, the key quantity of interest in the TBN model is the "entropy" of $\mathcal{M}'$, defined as the number of polymers in $\mathcal{M}'$ (i.e., $|\mathcal{M}'|$). The TBN model defines the "stable" configurations to be those that have maximum entropy among all saturated configurations.

Corresponding to the TBN literature [5], a multiset of polymers $\mathcal{M}$ is called *TBN-stable* if any reaction $\mathcal{M} \to \mathcal{M}'$ has $|\mathcal{M}| \geq |\mathcal{M}'|$. We say that a set $\mathcal{S}$ of polymers is *TBN-stability closed* if every multiset $\mathcal{M} \in \mathbb{N}^{\mathcal{S}}$ is TBN-stable and further any reaction $\mathcal{M} \to \mathcal{M}'$ where $\mathcal{M}'$ contains a polymer outside of $\mathcal{S}$ (i.e., $P \in \mathcal{M}'$ for some $P \notin \mathcal{S}$) satisfies $|\mathcal{M}| > |\mathcal{M}'|$. In other words, producing a polymer outside of a TBN-stability closed set $\mathcal{S}$ costs entropy. The following lemma connects our notion of stability of on-target polymers (Definition 3.4) to TBN-stability.

▶ **Lemma 7.1.** *If $\mathcal{S}$ is TBN-stability closed and satisfies condition (1) of Definition 3.1, then $\mathcal{S}$ is on-target with $\mu(P) = 1$ for all $P \in \mathcal{S}$ (i.e., uniform) and stable.*

Proof. We first check $\mathcal{S}$ is on-target with μ. For a reaction $\alpha : \mathcal{M}_1 \to \mathcal{M}_2$ where $\mathcal{M}_1$ and $\mathcal{M}_2$ are multisets over $\mathcal{S}$, $\mu(\mathcal{M}_1) = |\mathcal{M}_1| \geq |\mathcal{M}_2| = \mu(\mathcal{M}_2)$. Applying the same argument to $\alpha' : \mathcal{M}_2 \to \mathcal{M}_1$ yields $\mu(\mathcal{M}_1) \leq \mu(\mathcal{M}_2)$. Combining the two gives $\mu(\mathcal{M}_1) = \mu(\mathcal{M}_2)$, proving $\mathcal{S}$ with μ is an on-target set.

Now we will prove that uniform on-target $\mathcal{S}$ is stable. Consider a canonical reaction $\beta : \mathcal{M}'_1 \to \mathcal{M}'_2$ with $l(\beta) > 0$. By definition, we have $k(\beta) = \mu(\mathcal{M}'_1) - \mu(\hat{\mathcal{M}'_2}) = |\mathcal{M}'_1| - |\hat{\mathcal{M}'_2}|$ and $l(\beta) = |\mathcal{M}'_2| - |\hat{\mathcal{M}'_2}|$. Therefore, the condition for $\mathcal{S}$ being stable, namely $k(\beta)/l(\beta) > 1$, is implied by $|\mathcal{M}'_1| > |\mathcal{M}'_2|$.   ◀

Thus, the polymers in the TBN-stability closed set $\mathcal{S}$ represent the intended "high-concentration" polymers, while everything outside of $\mathcal{S}$ is considered undesired (off-target).

Let canonical reaction α be $\mathcal{M} \to \mathcal{M}'$; we define $|\mathcal{M}| - |\mathcal{M}'|$ to be the *entropy loss* $e(\alpha)$ of the reaction α. Recall that during the first iteration of our algorithm, novelty $l(\alpha)$ is the number of off-target polymers generated in canonical reaction α. The imbalance $k(\alpha)$ in the first iteration can be understood in terms of the entropy loss of the reaction: the decrease in the number of polymers of a reaction is exactly $e(\alpha) = k(\alpha) - l(\alpha)$. Thus to have an upper bound on the concentration of off-target polymers via Theorem 6.2, it is sufficient to find the smallest ratio $e(\alpha)/l(\alpha)$ of any reaction:

▶ **Corollary 7.2.** *Let set $\mathcal{S} \subseteq \mathbf{\Psi}$ be a TBN-stability closed set of polymers. Let $\mu_1 = \min_\alpha \{e(\alpha)/l(\alpha)\} + 1$ minimized over all reactions $\alpha : \mathcal{M} \to \mathcal{M}'$ where $\mathcal{M} \in \mathbb{N}^{\mathcal{S}}$, $\mathcal{M}' \in \mathbb{N}^{\mathbf{\Psi}}$ and $l(\alpha) > 0$. For any $0 < c < 1$, there are monomer concentrations $\mathbf{x^0} \in (0,1)^{\mathbf{\Psi^0}}$ and configuration $\mathbf{x} \in (0,1)^{\mathbf{\Psi}}$ that minimizes $g(\mathbf{x})$ subject to $\mathbf{A} \cdot \mathbf{x} = \mathbf{x^0}$ where $\mathbf{x}$ satisfies: each polymer $P \in \mathcal{S}$ has concentration exactly c, and each polymer $P \in \mathbf{\Psi} \setminus \mathcal{S}$ has concentration not more than c^{μ_1}.*

In other words, if $\mathcal{S}$ is a TBN-stability closed set of polymers then we can assign concentration c to each of them (uniform assignment). Then we consider the "worst" way to generate any polymers outside of $\mathcal{S}$ (i.e., off-target): the way that has the smallest entropy loss and the largest novelty. The ratio of the entropy loss to novelty gives us an upper bound on the concentration exponent μ_1 of any off-target polymer, bounding its concentration by c^{μ_1}. The smallest entropy loss to generate off-target polymers is already the key item of interest in the TBN literature. Thus the above corollary helps to bootstrap concentration bounding arguments via existing arguments based on entropy loss.

8   Example Applications

In this section we show several applications of our mathematical tools in the analysis of existing systems of interest in the DNA molecular programming literature. We base our arguments on previous results proving the entropy loss (quantity $e(\alpha)$ of Corollary 7.2) of the systems in producing off-target (leak) species.

We note that while existing literature succeeds in characterizing entropy loss via TBN-like combinatorial arguments, more work is needed to develop similarly rigorous combinatorial arguments on novelty (quantity $l(\alpha)$); see also the discussion in Section 9. In the examples to follow, we claim to have identified the worst-case canonical reactions – i.e., having the least $e(\alpha)/l(\alpha)$ or $k(\alpha)/l(\alpha)$ ratio – without proof.

We first consider the TBN AND gate introduced in prior work [5, 2] and recently experimentally realized [14]. Figure 1 (left) shows the desired functionality of the AND gate in which inputs A and B cooperate to produce C. We are interested in bounding the concentration of C (leak) that can be produced in the absence of inputs A and B. Phrased in our terminology, combinatorial TBN arguments have shown that $\mathcal{S} = \{X, Y, Z\}$ is TBN-stability closed, as well as $\{X, Y, Z, A\}$ and $\{X, Y, Z, B\}$ where one of the two inputs is present. This implies that any canonical reaction α producing C has entropy loss $e(\alpha) \geq 1$. However, such arguments do not directly connect entropy loss to leak concentrations, justifying the need for a tool like our Corollary 7.2.

To apply our framework to the TBN AND gate, we use Corollary 7.2. In the absence of both inputs, take $\mathcal{S} = \{X, Y, Z\}$. We claim that the worst-case canonical reaction α producing C is $X + Y + Z \to G_1 + C$ shown in Figure 1 (middle). This reaction has entropy loss $e(\alpha) = 1$ and novelty $l(\alpha) = 2$ since G_1 and C are outside of $\mathcal{S}$. Thus for any base concentration $0 < c < 1$, there is an equilibrium with $[X] = [Y] = [Z] = c$ where the leak concentration is $[C] \leq c^{1.5}$.

Similarly, consider the case of having input B but no input A. We take $\mathcal{S} = \{X, Y, Z, B\}$ and claim[9] that the worst-case canonical reaction α producing C is $X + Y + Z + B \to G_2 + G_3 + C$ as shown in Figure 1 (right). This reaction has entropy loss $e(\alpha) = 1$ and novelty $l(\alpha) = 3$ since G_2, G_3, and C are outside of $\mathcal{S}$, yielding the equilibrium leak concentration of $[C] \leq c^{1.33}$. Our analysis thus provides concrete polynomial upper bounds on leak, consistent with the qualitative expectation of the TBN model that leak should become comparatively negligible in the limit of decreasing c. The bound is smaller when both inputs are absent than when B is present.

As the next example, we consider the "leakless" DNA strand displacement system previously theoretically and experimentally studied [15]. That work focuses on a family of "translator" modules that convert an input strand to an output strand of independent sequence. The family is parameterized by the redundancy parameter N defined as the number of bound domains in each fuel polymer F_i, such that the number of domains in each signal X_i is $N + 1$. Leak is expected to decrease with decreasing overall concentration (as for the AND gate), as well as with increasing redundancy N. The decrease of leak with increasing N was confirmed by experiment, at least for small N. The reactions of the translator with $N = 3$ are shown in Figure 2.

---

[9] While we do not provide a formal proof that this reaction is worst-case, our confidence is based on the observation that we cannot increase the novelty further without increasing the entropy loss in proportion. For example, we can combine two copies of reaction α to yield reaction α' but then $e(\alpha') = 2e(\alpha)$ and $l(\alpha') = 2l(\alpha)$, maintaining the same ratio.
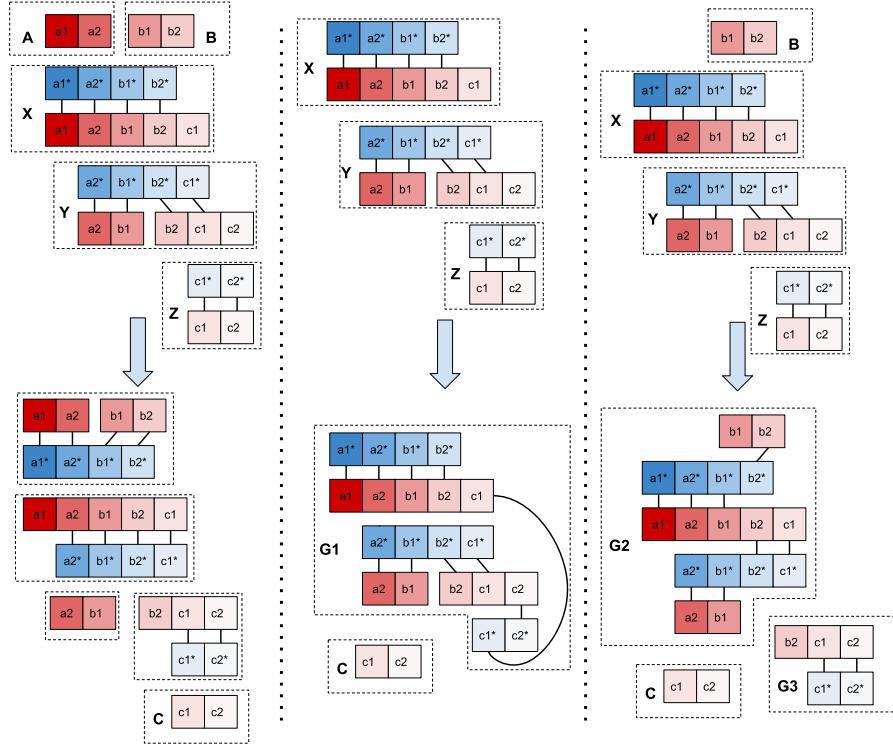
Figure 1 A TBN module implementing an AND gate. The presence of polymers A and B represents input 1, while their absence represents input 0. (Left) The intended reaction pathway generates the output polymer C when both A and B are present. (Middle) When both A and B are absent, the worst-case canonical reaction is $\alpha : X + Y + Z \to G_1 + C$ producing an erroneous output C (leak). The entropy loss and novelty of the reaction are $e(\alpha) = 1$ and $l(\alpha) = 2$, resulting in the concentration upper bound of C of $c^{1.5}$ via Corollary 7.2. (Right) If only one input is present (B), then the worst-case canonical reaction is $\beta : B + X + Y + Z \to G_2 + G_3 + C$ with $e(\beta) = 1$ and $l(\beta) = 3$, yielding the upper bound concentration of C of $c^{1.33}$.

To apply our framework, we consider the system without toeholds, driven solely by entropy; with long domains alone, we are in the enthalpy neutral (athermic) regime. Prior work focused on the case where the system was prepared in a state with only fuel polymers (F_i), all at equal concentration, and zero initial concentration of waste polymers (W_i). Let on-target set be $\mathcal{S} = \{F_i, W_i \mid i = 1, \dots, n\}$ with uniform $\mu = 1$. Rephrasing in our terminology, ref. [13] proved that any canonical reaction $\alpha : \mathcal{M} \to \mathcal{M}'$ where $X_i \in \mathcal{M}'$ has entropy loss $e(\alpha) = N - 1$, and this fact was used to argue that by increasing N we can make leak arbitrarily small. We now show that considering novelty in addition to entropy loss makes this argument problematic, and suggest an alternative parameter setting to arbitrarily decrease leak.

Consider a cascade of two translators. Importantly, the number of fuels for a single translator module is $N + 1$; i.e., X_i and X_j overlap in sequence for $i < j < i + N + 1$ but are completely sequence-independent for $j = i + N + 1$. We claim that the worst-case leak pathway α is where the first translator leaks resulting in the triggering of the second translator. This pathway generates $l(\alpha) = N + 3$ new off-target polymers: 1 for the leaked upstream translator (all fuels coming together), $N + 1$ for the triggered fuels of the second translator, and 1 for the output. Thus $e(\alpha)/l(\alpha) = (N - 1)/(N + 3)$. By Corollary 7.2, the
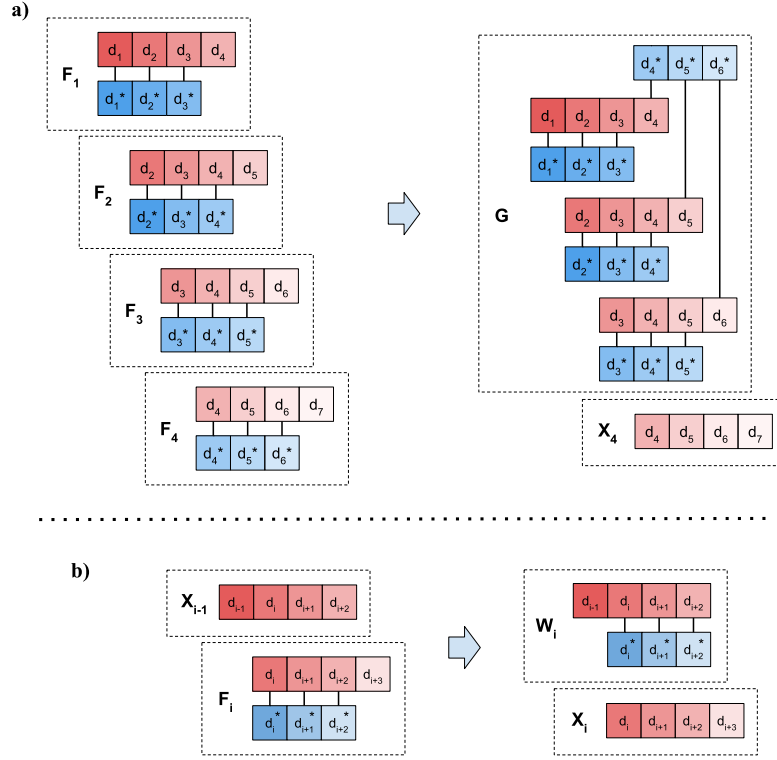
Figure 2 Translator cascade with redundancy parameter $N = 3$. Polymer X_0 serves as input and polymer X_n as output signal. The leak pathway with $N + 1$ fuels coming together to generate a signal X_{N+1} in the absence of input is described in (a). Part (b) describes the intended reaction for each i; iterating them for $i = 1, ..., N + 1$ produces X_{N+1} given input X_0. Note that if several translators are composed, then X_{N+1} is the input to the downstream translator and once the leak signal is generated via the pathway shown in (a) it can propagate via intended reactions to the output X_n of the last translator.

concentration of the leak product in the uniform setting is bounded above by $c^{4/3}$ for $N = 3$. The bound tightens to c^2 with increasing N; however, it does not get arbitrarily small. This suggests that we do not decrease equilibrium leak concentration arbitrarily by increasing "redundancy" N because while the entropy loss increases, the novelty increases as well.

To arbitrarily decrease leak, we propose to use positive initial concentrations of the waste polymers W_i. The fuel polymers are denoted $F_1, ..., F_n$, and the waste polymers produced after translator triggering are $W_1, ..., W_n$. Let X_0 represent the input and X_n represent the output. We have $n = 2(N + 1)$ for two translators composed together. This cascade of length n proceeds through the following reactions described in Figure 2:

$$X_{i-1} + F_i \rightarrow X_i + W_i$$
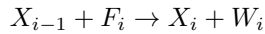
for $i = 1, ..., n$.

We first consider the triggered system (with-input) showing at least a constant fraction of the signal is propagated to the end as X_n. We then focus on the case without input and bound the leak. Note that this analysis utilizes Theorem 5.4 directly rather than Corollary 7.2 because we will have different concentration exponents μ for F_i and W_i in $\mathcal{S}$.

For the triggered (with-input) system we define our on-target set as $\mathcal{S} = \{F_i, W_i, X_0, X_i \mid i = 1, \ldots, n\}$. We assign concentrations to the on-target polymers as follows: All fuel concentrations are equal to $2c$, and all waste concentrations are equal to c. These concentrations correspond to concentration exponents $\mu(F_i) = 1 + \log_c 2$ and $\mu(W_i) = 1$. Let the concentration of the output in the final layer X_n be y and assign balancing concentrations of the other X_i. Since $[F_i]/[W_i] = 2$, we have $\sum_{i=0}^{n}[X_i] < 2y$, meaning that more than half of the total signal (X_i) is at the output layer.

Now, we investigate the system without input. Let $\mathcal{S} = \{F_i, W_i \mid i = 1, \ldots, n\}$. For the situation to properly correspond to the with-input case, we need to ensure that all monomer concentrations are the same between the two cases, except removing the monomer corresponding to input X_0. Rather than thinking about specific monomers, we start in the with-input case and conceptually run reactions $X_i + W_i \to X_{i-1} + F_i$ to completely push all X_i to X_0, and then remove X_0 from the system.[10] Since the total amount of all X_i is less than $2y$ in the with-input case, this results in: $[F_i] < 2c + 2y$ and $[W_i] > c - 2y$. These correspond to $\mu(F_i) > \log_c(2c + 2y)$ and $\mu(W_i) < \log_c(c - 2y)$.

Recall that redundancy N results in entropy penalty $N - 1$. We claim that the reaction with the smallest imbalance-novelty ratio (i.e., worst-case) is reaction β:

$$F_1 + \cdots + F_n \;\to\; G + W_{N+2} + \cdots + W_n + X_n,$$
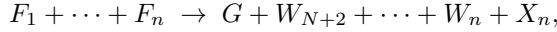
where G is the "large polymer" formed after all $F_1, \ldots, F_N$ displace the top strand from F_{N+1}, and X_n is the leak output. The imbalance of this reaction is:

$$k(\beta) = \sum_{i=1}^{n} \mu(F_i) - \sum_{i=N+2}^{n} \mu(W_i) > n \log_c(2c + 2y) - \frac{n}{2} \log_c(c - 2y). \tag{4}$$

We can ensure that $k(\beta)$ is at least a constant fraction of n for small enough c. For example, if we let $y \le c/4$, then $k(\beta) \ge n/4$ for any $c \le 0.0064$. The novelty is independent of n: $l(\beta) = 2$ since G and X_n are not in $\mathcal{S}$. Therefore, the imbalance-novelty ratio $k(\beta)/l(\beta)$ of the worst case reaction is at least $n/8$, which increases linearly with N (recall $n = 2(N+1)$ for two translators composed together). Applying Theorem 5.4 leads to leak concentration of at most $c^{n/8} = c^{1/4 + N/4}$. This upper bound[11] implies smaller-than c concentration of leak for $N \ge 4$, with the leak exponentially decreasing for larger N.

To summarize, by increasing redundancy N in the appropriate regime, we maintain the property that a constant fraction of the input is converted to output in the with-input case, while arbitrarily (exponentially in N) decreasing leak in the without-input case.

9 Discussion

Our results suggest a few important directions for future work. Given the central role of worst-case canonical reactions – i.e., canonical reactions with the lowest imbalance-novelty ratio (for Algorithm 1 and Theorem 5.4) or entropy loss-novelty ratio (Corollary 7.2) – it is important to develop formal techniques to prove that a given canonical reaction is indeed worst-case overall, at least or for a particular off-target polymer (for Section 6). Note that while combinatorial techniques in prior work in the TBN model have focused on proving

---

[10] More precisely, this ensures that the concentrations of monomers making up on-target polymers are the same between the with-input and without-input case (other than X_0).

[11] Note that this upper bound is loose because of inequalities such as Equation (4).

entropy loss, more work is needed to study the ratio directly, making our framework more easily applicable. While we believe the canonical reactions highlighted in Section 8 are indeed worst-case, the argument is informal.

Another promising avenue of research is to establish a more direct link between a polymer's monomer composition and its equilibrium concentration. Our current framework is effectively reaction-centric, inferring concentrations based on how polymers transform into one another. An alternative approach could be to derive concentration bounds directly from the structural properties of the off-target polymers, such as their size (monomer count) or degree of overlap with one another (multiset difference). Nonetheless, we hope that a variety of structure-based results could be proven based on a reduction to our canonical reaction framework.

Finally, this work has focused exclusively on the athermic case, where all molecular interactions are enthalpically neutral. While this is a reasonable and useful abstraction for systems with strong, saturated bonds, many real-world molecular systems, including many popular in DNA molecular programming, involve a range of binding strengths and enthalpic effects (e.g., from toehold binding). Extending our algorithmic framework to incorporate user-specified ΔG's for each polymer could significantly broaden its applicability, and although this would complicate our algorithm, we do not anticipate any insurmountable difficulties.

References

1 Keenan Breik, Cameron Chalk, David Doty, David Haley, and David Soloveichik. Programming substrate-independent kinetic barriers with thermodynamic binding networks. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, 18(1):283–295, 2021. `doi:10.1109/TCBB.2019.2959310`.

2 Keenan Breik, Chris Thachuk, Marijn Heule, and David Soloveichik. Computing properties of stable configurations of thermodynamic binding networks. *Theor. Comput. Sci.*, 785:17–29, 2019. `doi:10.1016/j.tcs.2018.10.027`.

3 W. Bruns, B. Ichim, C. Söger, and U. von der Ohe. Normaliz. Algorithms for rational cones and affine monoids. Available at `https://www.normaliz.uni-osnabrueck.de`.

4 Robert M Dirks, Justin S Bois, Joseph M Schaeffer, Erik Winfree, and Niles A Pierce. Thermodynamic analysis of interacting nucleic acid strands. *SIAM review*, 49(1):65–88, 2007. `doi:10.1137/060651100`.

5 David Doty, Trent A. Rogers, David Soloveichik, Chris Thachuk, and Damien Woods. Thermodynamic binding networks. In *DNA Computing and Molecular Programming, 23rd International Conference, DNA 23*, pages 249–266, 2017. `doi:10.1007/978-3-319-66799-7_16`.

6 Martin Feinberg. *Foundations of chemical reaction network theory.* Springer, 2019.

7 Stefan Hoops, Sven Sahle, Ralph Gauges, Christine Lee, Jürgen Pahle, Natalia Simus, Mudita Singhal, Liang Xu, Pedro Mendes, and Ursula Kummer. COPASI—a complex pathway simulator. *Bioinformatics*, 22(24):3067–3074, 2006. `doi:10.1093/BIOINFORMATICS/BTL485`.

8 Fritz Horn and Roy Jackson. General mass action kinetics. *Archive for rational mechanics and analysis*, 47:81–116, 1972.

9 Matthew D Johnston, Stefan Müller, and Casian Pantea. A deficiency-based approach to parametrizing positive equilibria of biochemical reaction systems. *Bulletin of Mathematical Biology*, 81:1143–1172, 2019.

10 A. M. M. Leal. Reaktoro: An open-source unified framework for modeling chemically reactive systems, 2015. URL: `https://reaktoro.org`.

11 John AP Sekar, Justin S Hogg, and James R Faeder. Energy-based modeling in BioNetGen. In *2016 IEEE international conference on bioinformatics and biomedicine (BIBM)*, pages 1460–1467. IEEE, 2016.

12 Elie Soloveichik, Leo Orshansky, Cameron Chalk, and Boya Wang. Concentrat.io, 2025. Accessed 25 June 2025. URL: `https://concentrat.io/`.

13   Chris Thachuk, Erik Winfree, and David Soloveichik. Leakless DNA strand displacement systems. In Andrew Phillips and Peng Yin, editors, *DNA Computing and Molecular Programming*, pages 133–153, Cham, 2015. Springer International Publishing. `doi:10.1007/978-3-319-21999-8_9`.

14   Boya Wang, Cameron Chalk, David Doty, and David Soloveichik. Molecular computation at equilibrium via programmable entropy. *bioRxiv*, 2025. `doi:10.1101/2024.09.13.612990`.

15   Boya Wang, Chris Thachuk, Andrew D. Ellington, Erik Winfree, and David Soloveichik. Effective design principles for leakless strand displacement systems. *Proc. Natl. Acad. Sci. USA*, 115(52):E12182–E12191, 2018. `doi:10.1073/pnas.1806859115`.

16   Boya Wang, Chris Thachuk, and David Soloveichik. Speed and correctness guarantees for programmable enthalpy-neutral DNA reactions. *ACS Synthetic Biology*, 12(4):993–1006, 2023.

17   Joseph N Zadeh, Conrad D Steenberg, Justin S Bois, Brian R Wolfe, Marshall B Pierce, Asif R Khan, Robert M Dirks, and Niles A Pierce. NUPACK: analysis and design of nucleic acid systems. *Journal of computational chemistry*, 32(1):170–173, 2011. `doi:10.1002/JCC.21596`.

A   Hilbert Basis Implementation

This section recalls some results on Hilbert bases and explains how to use the Hilbert basis in Algorithm 1. While this section makes the main theorem mathematically rigorous, most readers can safely skip this section.

The main contents of this section are (1) the termination of Algorithm 1 and (2) the use of minimum (versus infimum) in Definition 5.2. We address both concerns by showing that, although there are infinitely many canonical reactions, we can always restrict our attention to a finite subset of them in our analysis and algorithm.

For integral vectors $\mathbf{v}_1, ..., \mathbf{v}_m \in \mathbb{Z}^n$, the set $C = \{\sum_{i=1}^m b_i \mathbf{v}_i : b_1, ..., b_m \geq 0\}$ is called a (rational polyhedral) cone, which is also known to be described by a system of inequalities $C = \{\mathbf{v} : \mathbf{B} \cdot \mathbf{v} \leq 0\}$ for some matrix $\mathbf{B} \in \mathbb{Z}^{l \times n}$. It is known that the set $C \cap \mathbb{Z}^n$ has a *finite* subset $\mathcal{H}(C) = \{\mathbf{h}_1, ..., \mathbf{h}_t\}$, called *Hilbert basis* of C, that generates $C \cap \mathbb{Z}^n$ with non-negative integer coefficients, that is, for any $\mathbf{v} \in C \cap \mathbb{Z}^n$ there are $a_1, ..., a_t \in \mathbb{N}$ such that $\mathbf{v} = \sum_{i=1}^t a_i \mathbf{h}_i$.

The set of canonical reactions Λ can be precisely described in terms of a rational polyhedral cone. For a canonical reaction $\alpha : \mathcal{M}_1 \to \mathcal{M}_2$, we define a vector

$$\mathbf{v}_\alpha = (\mathcal{M}_1[P] - \mathcal{M}_2[P])_{P \in \mathbf{\Psi}} \in \mathbb{Z}^{|\mathbf{\Psi}|}$$

capturing the stoichiometric change of concentrations due to reaction α. Note that $\mathcal{M}_1[P'] - \mathcal{M}_2[P'] = -\mathcal{M}_2[P'] \leq 0$ for $P' \notin \mathcal{S}$, thus $\mathbf{v}_\alpha \cdot \mathbf{e}_{P'} \leq 0$ for $\mathbf{e}_{P'} = (\delta_{PP'})_{P \in \mathbf{\Psi}}$ for the delta function $\delta_{ij} = 1$ iff $i = j$. Definition 2.2 ensures that this vector must satisfy the condition $\mathbf{A} \cdot \mathbf{v}_\alpha = 0$. Combining these, the cone

$$C^{\mathcal{S}} = \{\mathbf{v}_\alpha \in \mathbb{R}^{|\mathbf{\Psi}|} : \mathbf{A} \cdot \mathbf{v}_\alpha \geq 0 \text{ and } \mathbf{A} \cdot \mathbf{v}_\alpha \leq 0 \text{ and } \mathbf{v}_\alpha \cdot \mathbf{e}_P \leq 0 \text{ for all } P \notin \mathcal{S}\}$$

characterizes the canonical relations: $C^{\mathcal{S}} \cap \mathbb{Z}^{|\mathbf{\Psi}|}$ is the set of vectors $\mathbf{v}_\alpha$. Therefore, there exists a finite set of canonical reactions H that corresponds to the Hilbert basis $\mathcal{H}(C^{\mathcal{S}})$.

The following lemma implies that for the purposes of our algorithm and analysis, we can focus on the Hilbert basis H of canonical reactions.

▶ **Lemma A.1.** *Let $\alpha : \mathcal{M}_1 \to \mathcal{M}_2$ and $\beta : \mathcal{N}_1 \to \mathcal{N}_2$ be canonical reactions with $l_i(\alpha), l_i(\beta) \neq 0$. Then, for any $a, b \in \mathbb{N}$, it holds that*

$$\frac{k_i(a \cdot \alpha + b \cdot \beta)}{l_i(a \cdot \alpha + b \cdot \beta)} \geq \min\left(\frac{k_i(\alpha)}{l_i(\alpha)}, \frac{k_i(\beta)}{l_i(\beta)}\right). \tag{5}$$

The equality holds only when $a = 0$, $b = 0$, or $k_i(\alpha)/l_i(\alpha) = k_i(\beta)/l_i(\beta)$.

Proof. It is not hard to see that k_i and l_i are linear, i.e., $k_i(a \cdot \alpha + b \cdot \beta) = a \cdot k_i(\alpha) + b \cdot k_i(\beta)$ and $l_i(a \cdot \alpha + b \cdot \beta) = a \cdot l_i(\alpha) + b \cdot l_i(\beta)$. Then, Equation (5) is identical to the mediant inequality, which states that for $p, q, r, s \geq 0$ with $q, s \neq 0$ it holds that $\min(p/q, r/s) \leq (p+r)/(q+s)$. ◄

Consider a canonical reaction $\alpha = a_1 \cdot \eta_1 + ... + a_t \cdot \eta_t$ for $a_1, ..., a_t > 0$ and $\eta_1, ..., \eta_t \in H$. If α is i-levelizing, then the equality $k_i(\alpha)/l_i(\alpha) = \mu_i = \min_{j=1,...,t}(k_i(\eta_j)/l_i(\eta_j))$ must hold, and the equality condition above ensures that $\mu_i = k_i(\eta_j)/l_i(\eta_j)$ for all $j = 1, ..., t$. In other words, the set of i-levelizing reactions is

$$\left\{ \sum_{j=1}^{t} a_j \cdot \eta_j : a_1, ..., a_t \in \mathbb{N} \right\} \text{ where } \{\eta_1, ..., \eta_t\} \text{ is the set of } i\text{-levelizing reactions in } H.$$

This allows us to inspect the minimum over the finite set H instead of Λ in Definition 5.2. Similarly, as the Hilbert basis can be computed in finite time and implemented in [3], we can run Algorithm 1 with guaranteed termination by computing the minimum over H.

B   Detailed Balance and Equilibrium

Recall that $\mathbf{A} \in \mathbb{N}^{|\mathbf{\Psi^0}| \times |\mathbf{\Psi}|}$ is the matrix such that each entry A_{ij} specifies the number of monomers of type i in polymer j, and that $\mathbf{A} \cdot \mathbf{x} = \mathbf{x^0}$ (Equation (1)) captures the mass-conservation constraint.

▶ **Theorem B.1.** *Let $\mathbf{x^0} \in (0,1)^{\mathbf{\Psi^0}}$ be a fixed vector of monomer concentrations. If all reactions are balanced at the configuration $\mathbf{x} \in (0,1)^{\mathbf{\Psi}}$ of polymer concentrations, then the cost function $g(\mathbf{x})$ is minimum subject to $\mathbf{A} \cdot \mathbf{x} = \mathbf{x^0}$.*

Proof. (Sketch) The function g is strictly convex since its Hessian H is positive definite (specifically diagonal with $H_{jj} = 1/x_j > 0$). Strict convexity of g implies that the local minimum of g becomes the unique (global) minimum.

We associate a vector $\mathbf{v}_\alpha \in \mathbb{Z}^{|\mathbf{\Psi}|}$ with every reaction α, capturing the net stoichiometric effect of reaction α.[12] It is straightforward to show that the function g along with the direction of $\mathbf{v}_\alpha$ has zero derivative at $\mathbf{x}$ if and only if the reaction α is balanced at $\mathbf{x}$. More explicitly, for $\alpha : \mathcal{M}_1 \to \mathcal{M}_2$, the P-th entry of the vector $\mathbf{v}_\alpha$ is $(\mathcal{M}_1[P] - \mathcal{M}_2[P])$. The directional derivative $D_{\mathbf{v}_\alpha} g(\mathbf{x}) = \sum_P (\mathcal{M}_1[P] - \mathcal{M}_2[P]) \log x_P = 0$ implies that $\prod_{P \in \mathcal{M}_1} x_P^{\mathcal{M}_1[P]} = \prod_{P \in \mathcal{M}_2} x_P^{\mathcal{M}_2[P]}$ holds, which means the reaction α is balanced.

The set $\{\mathbf{x} : \mathbf{A} \cdot \mathbf{x} = 0\}$ is spanned by the vectors $\mathbf{v}_\alpha$ for all reactions by Definition 2.2. Therefore, if all reactions are balanced at $\mathbf{x}$, then any directional derivative at $\mathbf{x}$ vanishes and $\mathbf{x}$ is a critical point, which is the unique minimum. ◄

---

[12] This is the same vector $\mathbf{v}_\alpha$ as defined in Section A. For example, $(1, -1, 0, ...)$ corresponds to $X_1 \to X_2$ for $\Psi = \{X_1, X_2, ...\}$.