Modularity of Signal Propagation Networks in the Thermodynamic Binding Network Model

And Kaan Ata Yilmaz
Electrical and Computer Engineering, University of Texas at Austin, TX, USA

David Soloveichik
Electrical and Computer Engineering, University of Texas at Austin, TX, USA

Abstract

The Thermodynamic Binding Network (TBN) model provides an equilibrium-based abstraction for molecular systems, but proving correctness of large networks built from individually verified modules remains difficult because modules can interact through shared binding-site types. We develop a method for proving correctness of modular TBNs by reducing a composition of arbitrarily many modules to a fixed coarse-grained TBN, computing the Hilbert basis of the reduced system, and lifting the resulting entropy bounds and polymer characterizations back to the original network. We apply this method to reversible signal-propagation modules: For the previously studied module implementing $A + B \longleftrightarrow C$, we show that stable configurations of an arbitrary network of such modules have the intended local structure, implying that firing an individual module preserves global stability. Under certain restrictions on the reaction set, these networks faithfully simulate a class of reversible chemical reaction networks: every bounded-length CRN execution can be realized by a height-2 path between stable TBN configurations, and any two stable TBN configurations map to CRN states in the same stoichiometric compatibility class. We further show that the same proof strategy extends to a novel module implementing reactions with two reactants and products like $A + B \longleftrightarrow C + D$.

2012 ACM Subject Classification Theory of computation → Models of computation

Keywords and phrases Thermodynamic Binding Networks, Chemical Reaction Networks, Modular Composition, Verification, Simulation

Digital Object Identifier 10.4230/LIPIcs.DNA.32.10

Funding Schmidt Sciences Polymath Award to D.S., NSF SemiSynBio III: GOALI grant 2227578.

1 Introduction

Much like the field of digital electronics, progress in engineering molecular computation has depended on a collection of abstractions and models, ranging from those closer to the underlying hardware like strand displacement cascades [15, 12], to more general ones like algorithmic tile assembly [6] and chemical reaction networks [14, 4]. The Thermodynamic Binding Network (TBN) model [7, 3] was introduced to capture a class of systems designed to obey certain simplifying restrictions: (1) behavior is not programmed through geometric or steric constraints, (2) binding occurs only between complementary site types (aka domains in DNA terminology), (3) configurations maximizing the number of bonds are preferred, with entropy, measured by the number of separate polymers (aka complexes), breaking ties between fully saturated configurations. Recent experimental implementations of simple TBNs reinforce the practical relevance of the model as an engineering abstraction [17].

While the original intent of the TBN model was to capture thermodynamic equilibria, the model has subsequently been extended with a notion of kinetic paths and kinetic barriers [2]. Importantly, this is still quite different from a kinetics-first model such as strand-displacement cascades. Rather than assigning rates and focusing on the most likely trajectories, the TBN model considers all possible rearrangements of the components that are permitted by the

© And Kaan Ata Yilmaz and David Soloveichik;
licensed under Creative Commons License CC-BY 4.0

binding rules. This conservative viewpoint has an important advantage for verification: proving correctness in the TBN model rules out not only failure along the intended pathway, but also all unintended "leak" pathways. In this sense, it helps guard against errors that arise when thermodynamic driving forces conflict with the desired kinetic behavior.

Despite the apparent simplicity of the TBN model, it is surprisingly powerful. For example, there is a certain sense in which the stable configurations of TBNs can simulate the minimum free energy configurations of any other monomer-polymer system [18]. Another indication of the expressive power of the TBN model is that natural analysis problems for it are computationally hard (under reasonable computational complexity assumptions like $P \neq NP$) [3, 10]. Nonetheless, although worst-case analysis may be intractable, engineered TBNs may have certain properties that make their analysis and verification of correctness effective.

In this paper we develop a new method of proving the correctness of TBN systems that have been designed to be modular. Specifically, the monomers on such TBNs can be logically partitioned into "modules" that share a common underlying structure. Most TBNs are modular in this sense, including constructions for logic circuits [7], amplification [11], and signal propagation [17]. Since each individual module contains few monomers, it is easy to check its correctness in isolation using existing computational tools [3, 9, 16, 8]. However, verifying isolated module behavior in no way implies that the combined system behaves as intended. Indeed, modules necessarily share binding site types and the thermodynamically preferred configuration may involve some unintended cross-module hybrid complexes.

The method of analysis we develop is based on restricting the possible form of polymers in the combined system. Specifically, we employ a three-step argument: (1) Create a reduction from the original TBN into another TBN in which all modules collapse into a single one. (2) Compute the Hilbert-basis of the reduced TBN to enumerate all the polymers that could possibly appear in its stable configuration independently of the monomer counts. Importantly, step (2) leaves us with the same set of polymers in the reduced system for infinitely many original TBNs (i.e., arbitrary compositions of the given modules). (3) Eliminate the remaining undesired polymers based on monomer counts and how the modules are composed in the original TBN. We apply our method to two modular TBN signal propagation networks: one theoretically and experimentally studied in prior work [17], and one novel. While we cannot guarantee that other constructions will be amenable to the same analysis, we hope that our method is general enough to handle other systems.

While the emphasis of this work is on developing a general proof technique, of independent interest, our results prove the correctness (in the sense described below) of composition of modules for implementing systems of reversible reactions like $A + B \longleftrightarrow C$. We impose certain restrictions on the modules and how the modules are composed; for example, reactions repeating species such as $A + A \longleftrightarrow B$ are disallowed, as well as pairs of reactions that overlap in two or more species such as $A + B \longleftrightarrow C$ with $C + B \longleftrightarrow D$. Under these restrictions, we prove that all complexes that occur in the overall TBN can be mapped to individual modules (i.e., modules do not interfere with one another). This, in turn, allows us to prove faithful kinetic simulation in the following sense: First, given enough (fuel) monomers, the TBN simulates any number of CRN reaction firings. Specifically, there is a height-2 path as interpreted via TBN kinetics formalized in prior work [2]. Second, for the other direction, we show that stable configurations of the TBN correspond to CRN configurations in the same stoichiometric compatibility class: that is the TBN cannot reach a stable configuration that the CRN's conservation laws forbid.

Note that if the target CRN violates our restrictions, we can add additional reactions to effectively "translate" between species. With enough additional fuel, valid paths in the original CRN have valid images in the new CRN. While it remains as future work to show that any height-2 path in the TBN corresponds to a valid CRN trajectory, our results show that rather simple TBNs can provably simulate reversible chemical reaction networks in the sense described. Further, since reaction $A + B \longrightarrow C$ implements a logic AND gate, and two different reactions outputting the same species implement implicit OR, our results also imply simulation of dual-rail logic circuits. Boolean completeness of TBNs was previously established, but via different constructions [7, 5].

2 TBN Definitions

A *site type* is a formal symbol indexed by an integer, such as a_1. Each site type has a complementary type, denoted a_1^*, with the interpretation that a_1 can bind to a_1^*. We refer to site types according to both their index and whether they carry a star: for example, a_1 is called a *1-unstarred site type*, while a_1^* is called a *1-starred site type*. A *monomer type* is defined as a multiset of site types. For instance, the monomer type $m = \{a_1, a_1, a_1^*, b_2^*, b_2^*\}$ contains two copies of a_1, two copies of b_2^*, and one copy of a_1^*. The *length* of a monomer type m is $|m|$, i.e. the number of binding sites included in m.

A *thermodynamic binding network (TBN)* is a multiset of monomer types. An individual occurrence of a monomer type in a TBN is called a *monomer*, and an individual occurrence of a site type within a monomer is called a *site*. We say a TBN is *star-limiting* if for each site type, there are at least as many sites of the unstarred type as the starred type among all monomers.

A *configuration* of a TBN is a partition of its monomers into submultisets called *polymers*. The number of polymers in a configuration α, counted with multiplicity, is called the *entropy* of α and is denoted by $S(\alpha)$. A subconfiguration of a configuration α is a submultiset of polymers included in α. A site type within a polymer is said to be *exposed* if, among the monomers in that polymer, the number of starred instances of that site type exceeds the number of unstarred instances. A polymer is *self-saturated* if it contains no exposed site types. A self-saturated polymer is *splittable* if it can be split into two self-saturated polymers. For a star-limiting TBN, a configuration is *saturated* if every polymer in the configuration is self-saturated. Intuitively, this means that no polymer contains an unmatched starred site and equivalently, all possible bonds are formed. A configuration α of a TBN T is *stable* if it is saturated and no other saturated configuration of T contains more polymers than α. Equivalently, stable configurations are precisely the entropy-maximizing saturated configurations.

Configurations evolve through sequences of elementary merge and split operations. Specifically, if a configuration α can be transformed into a configuration β by replacing two polymers in α with their union, then α is said to *merge* to β, and correspondingly, β *splits* to α. A *path* is a nonempty sequence of configurations in which each configuration merges or splits to the next. Observe that any two configurations are connected by a path. A path between two saturated configurations is a saturated path if each configuration along the path is also saturated. In the TBN model, bonds are assumed to be strong and maximum bonding is assumed at all times. Since any transition to an unsaturated configuration necessarily involves losing a bond, we consider only saturated paths.

The *height* $h(p)$ of a saturated path p with initial configuration α is the maximum entropy decrease $S(\alpha) - S(\gamma)$ over all configurations γ along p. The height of a path captures the difficulty of a transition by measuring the largest entropy loss incurred along the path. For

this interpretation, we need the initial and final configurations to be stable; for example, a large height need not indicate a difficult transition if the initial configuration is unstable and the whole path is downhill.

3 Reversible Signal Propagation Module

We consider circuits of reversible signal propagation modules introduced in prior work ([7, 3]) and recently realized experimentally [17]. The construction consists of two closely related types of modules, forward and backward. The forward module represents a reaction in the stoichiometric form $A + B \to C$, and the backward module represents the reaction in the reverse direction, i.e., $C \to A + B$.

When a forward or backward module is considered independently from the rest of the network and without input, the module has exactly one stable configuration (forward module in Figure 1(a) and backward module in Figure 2(a)). We call these configurations *untriggered modular configurations* for the forward and backward modules, respectively. Note that the fact that these configurations are the unique stable ones of an isolated module without input can be easily checked via tools such as `StableGen` or `stable_tbn` [3, 9, 16, 8]. Note that the output monomers are not free in either of these configurations.

Parts (b), (c) and (d) of Figure 1 and Figure 2 show all stable configurations of the forward and backward modules in the presence of input monomers, i.e., input A and B for the forward module and input C for the backward module. The only stable configurations where the output monomer(s) is free are shown in Part (d) of each figure. Again, the fact that these are exactly the stable configurations can be verified with the above mentioned software tools.

Importantly, the stable configuration, where the output(s) is released (Part (d) of Figure 1 and 2), contains the untriggered modular configuration of the reverse reaction in addition to free output(s). This is the sense in which these modules are reversible [17]. When the configuration is changed from Part (b) to Part (d), we say that the reaction has fired. Firing a reaction necessitates a height-2 path.

▶ **Definition 1.** *For a given reaction, we define the monomers in the untriggered modular configuration as* module monomers. *For any input or output species X, a monomer of the form $\{x_1, x_2\}$ is called an* input monomer *or* output monomer*, respectively.*

Note that the module monomers of a reaction include its output monomer(s). For a particular reaction R, we denote the set of module monomers by M_R and the set of input monomers by I_R. The definition of input monomers immediately implies a natural correspondence between the monomers of the signal propagation module and the input species of the reaction through $\{x_1, x_2\} \leftrightarrow X$.

As an example, consider the reaction $R : A + B \to C$. The set of *module monomers* is

$$M_R = \{\{a_1^*, a_2^*, b_1^*, b_2^*\}, \{a_1, a_2, b_1, b_2, c_1\}, \{a_2^*, b_1^*, b_2^*, c_1^*\}, \{a_2, b_1\}, \{b_2, c_1, c_2\}, \{c_1^*, c_2^*\}, \{c_1, c_2\}\},$$

that is, the set of monomers appearing in the configuration shown in Figure 1(a). The set of *input monomers* for this reaction is $I_R = \{\{a_1, a_2\}, \{b_1, b_2\}\}$. The output monomer $\{c_1, c_2\}$ can be released if, and only if, the input monomers are present.

Note that the designation of a monomer as being a module monomer (or more specifically an output monomer) or input monomer is with respect to a particular reaction. For example, consider the reactions $R_1 : A + B \to C$ and $R_2 : C \to A + D$. In this case, $I_{R_1} \cap M_{R_2} = \{\{a_1, a_2\}\}$.

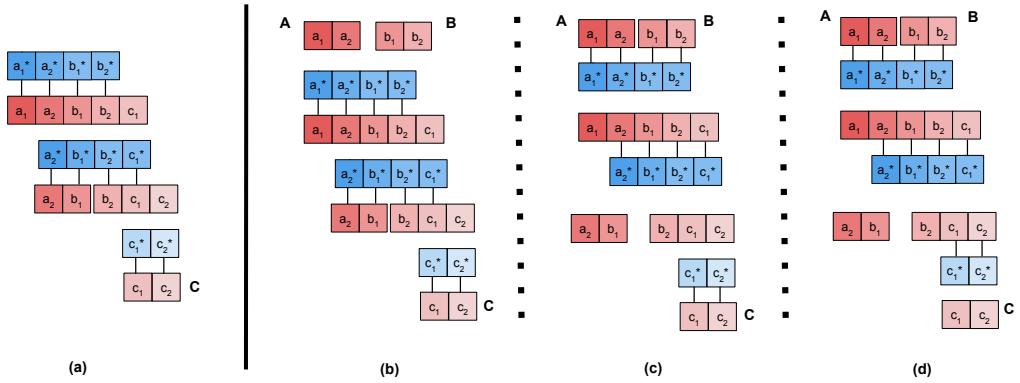

Figure 1 (a) Untriggered modular configuration of the module for the reaction $A + B \longrightarrow C$; (b), (c), (d) three stable configurations when both A and B are present.

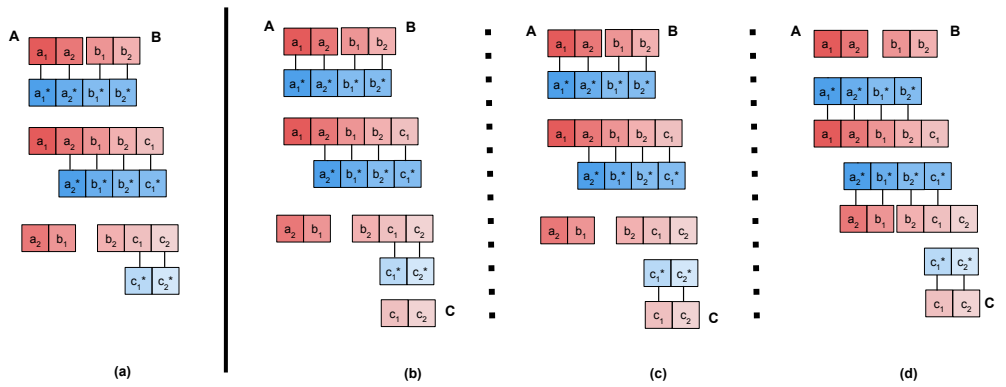

Figure 2 (a) Untriggered modular configuration of the module for $C \longrightarrow A + B$; (b), (c), (d) three stable configurations when C is present.

For a given list of reactions, we formally define the overall TBN network by adding multiplicities to modules and input monomers.

▶ **Definition 2** (Signal Propagation Network). *Let $\mathcal{R} = (R_1, R_2, \ldots, R_n)$ be a list of reactions. Let $I = \bigcup_i I_{R_i}$ denote the set of input monomer types corresponding to the reactions in $\mathcal{R}$. We define the TBN whose set of monomers consists of:*

1. *m_i copies of each module monomer in M_{R_i}, for every $i \in [n]$;*
2. *input monomers with multiplicities specified by the vector $k_I \in \mathbb{N}^{|I|}$.*

We denote this TBN by $\mathbb{SPN}\big(\mathcal{R}, ((m_i, M_i)_{i\in[n]}), k_I\big)$, where $\mathbb{SPN}$ stands for Signal Propagation Network*.*

4 Characterization of the Stable Configurations of a Signal Propagation Network

We determine the entropy of stable configurations of an arbitrary $\mathbb{SPN}$ and fully characterize the stable configurations of $\mathbb{SPN}$s whose underlying reaction sets satisfy certain conditions. Observe that, in an $\mathbb{SPN}$, the multiplicity of each unstarred site is at least as large as the multiplicity of its corresponding starred site. Therefore, $\mathbb{SPN}$s are star-limiting.

Moreover, since each monomer in an $\mathbb{SPN}$ consists either entirely of starred sites (a *star monomer*) or entirely of unstarred sites (an *unstar monomer*), it follows that in a self-saturated polymer, every starred site of a star monomer must form a bond with an unstarred site of an unstar monomer. As any saturated configuration consists of self-saturated polymers, a star monomer cannot be free in such a configuration and the entropy of a saturated configuration is upper bounded by the total number of unstar monomers in the network. A polymer containing d unstar monomers decreases this upper bound by $d-1$. We say that such a polymer performs $d-1$ unstar merge operations. Then, the entropy of a saturated configuration is equal to the value obtained by subtracting the total number of unstar merge operations across all polymers in the configuration from the total number of unstar monomers in the network.

This observation enables us to derive a lower bound on the entropy of any stable configuration. We denote the multiset of unstar monomers contained in the $\mathbb{SPN}$ as $U_{\mathbb{SPN}}$ and that of star monomers as $S_{\mathbb{SPN}}$. The cardinalities of these multisets are denoted as $|U_{\mathbb{SPN}}|$ and $|S_{\mathbb{SPN}}|$, respectively.

▶ **Lemma 3.** *If α is a stable configuration of $\mathbb{SPN}(\mathcal{R}, (m_i, M_i)_{i\in[n]}, k_I)$, then $S(\alpha) \geq |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$.*

Proof. The proof is by constructing a saturated configuration α' that has entropy of $S(\alpha') = |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$. Consider the configuration α' that consists of m_i-many untriggered modular configurations for each reaction R_i, as in Figure 1 and Figure 2 Part (a), and free input monomers, i.e. polymers by themselves. Note that each polymer in an untriggered modular configuration is self-saturated. Hence, α' is saturated.

Furthermore, each untriggered modular configuration has exactly one polymer which consists of two unstar monomers, while all other polymers include exactly one unstar monomer. Each such polymer decreases the maximum possible entropy $|U_{\mathbb{SPN}}|$ by 1. As there are m_i-many of them in α', the result follows. ◀

Next, we show that the quantity $|U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$ provides an upper bound for the entropy of a stable configuration of $\mathbb{SPN}(\mathcal{R}, (m_i, M_i)_{i\in[n]}, k_I)$. To do so, we rely on the notion of a Hilbert basis. Following the characterization of unsplittable polymers in TBNs

established in [9], we refer to the collection of unsplittable polymers as the Hilbert basis of the construction. Intuitively, these polymers serve as the fundamental building blocks of the system, since every self-saturated polymer can be expressed as a nonnegative integer combination of unsplittable polymers. In a stable configuration of any TBN – and therefore in an $\mathbb{SPN}$ – each polymer must be unsplittable. If a polymer were splittable, the entropy of the configuration could be increased by 1, without harming saturation, by splitting it. We focus on configurations consisting solely of unsplittable polymers because they include all stable configurations. Consequently, finding an upper bound on the entropy of every unsplittable configuration also yields an upper bound on the entropy of all stable configurations.

At this point, there are two obstacles to naively computing the Hilbert basis of an $\mathbb{SPN}$: (1) Since the set of reactions of an $\mathbb{SPN}$ can vary arbitrarily from one instance to another, we would need to compute the Hilbert basis separately for each network. This prevents us from deriving general upper bounds in terms of the parameters of an $\mathbb{SPN}$. (2) We would also need to compute the Hilbert bases of large networks with complex binding structures – for example, networks composed of many modules – despite the fact that computing a Hilbert basis is already a computationally hard problem.

Instead, we introduce a novel method based on a reduction map at the monomer level that takes an $\mathbb{SPN}$ and produces a system with a fixed number of monomer types, regardless of the modules included in the original $\mathbb{SPN}$. The induced mapping at the configuration level preserves saturation. We then argue about upper bounds on the entropy of a saturated configuration for an arbitrary $\mathbb{SPN}$ by analyzing this reduced system.

We define a mapping f, which we call the *reduction map*, that takes monomers from the $\mathbb{SPN}$ as input and outputs a monomer in the following way:

1. For $m \in U_{\mathbb{SPN}}$, $f(m) = \{x, x, ..., x\}$ where $|\{x, x, ..., x\}|$ = number of 1-unstarred sites in m.
2. For $m \in S_{\mathbb{SPN}}$, $f(m) = \{x^*, x^*, ..., x^*\}$ where $|\{x^*, x^*, ..., x^*\}|$ = number of 1-starred sites in m.

The binding site x included in the reduced network is different from every other binding site in the $\mathbb{SPN}$. Note that the reduction map outputs the same monomer for two monomers in $\mathbb{SPN}$ as long as the number of 1-unstarred (starred) sites in these monomers is the same, regardless of which reaction's set of module/input monomers these are from.

As an example, for the reaction $R : A + B \to C$, the output monomers are shown below:

- $f(\{a_1, a_2, b_1, b_2, c_1\}) = \{x, x, x\}$
- $f(\{a_1^*, a_2^*, b_1^*, b_2^*\}) = f(\{a_2^*, b_1^*, b_2^*, c_1^*\}) = \{x^*, x^*\}$
- $f(\{a_2, b_1\}) = f(\{b_2, c_1, c_2\}) = f(\{c_1, c_2\}) = f(\{a_1, a_2\}) = f(\{b_1, b_2\}) = \{x\}$
- $f(\{c_1^*, c_2^*\}) = \{x^*\}$

The image of the monomers of an arbitrary $\mathbb{SPN}$ under f consists of exactly four monomers: $\{\{x\}, \{x, x, x\}, \{x^*\}, \{x^*, x^*\}\}$, where the multiplicities of these monomers in the output network are determined by the multiplicities of module monomers and input monomers of the $\mathbb{SPN}$. We refer to these monomers as U1, U3, S1, S2, respectively, based on the number of unstarred or starred sites included in them.

Furthermore, the map f at the monomer level immediately defines mappings for polymers and configurations of the underlying $\mathbb{SPN}$:

1. for a polymer P, $f_{\text{poly}}(P) = \{f(m) | m \in P\}$
2. for a configuration α, $f_{\text{conf}}(\alpha) = \{f_{\text{poly}}(P) | P \in \alpha\}$

Note that f abstracts away binding site types by mapping every 1-unstarred site to a generic unstarred binding site x, and every 1-starred site to a generic starred site x^*, independently of the original binding site type. Since $\mathbb{SPN}$ contains at least as many 1-unstarred sites as 1-starred sites for each binding site type, the reduced network is also star-limiting. Consequently, the notion of self-saturation is preserved in the reduced network.

Moreover, if a polymer in $\mathbb{SPN}$ is self-saturated, then every 1-starred site in the polymer must be bound to the corresponding 1-unstarred site of the same binding site type. Therefore, the image of a self-saturated polymer under f_{poly} is also a self-saturated polymer in the reduced network. It follows that every saturated configuration of $\mathbb{SPN}$ is mapped to a saturated configuration of the reduced network under f_{conf}.

Furthermore, since each monomer in $\mathbb{SPN}$ has an image under f, every polymer has a corresponding image under f_{poly}, i.e. there is no polymer of the $\mathbb{SPN}$ without an image. Consequently, for any configuration α of $\mathbb{SPN}$, we have $S(\alpha) = S(f_{\text{conf}}(\alpha))$. Combining this with the fact that saturation is preserved under f_{conf}, we obtain the following lemma.

▶ **Lemma 4.** *An upper bound on the entropy of all saturated configurations of the reduced network is also an upper bound on the entropy of all saturated configurations of* $\mathbb{SPN}$.

To find an upper bound on the entropy of a saturated configuration of the reduced network, we will first compute the unsplittable polymers of the network, i.e., its Hilbert basis. As the number of monomer types in the reduced network is only four, regardless of the types of reactions in the underlying $\mathbb{SPN}$, this is a fairly easy computation and can even be done by hand in a few minutes. Alternatively, the software tools TBNExplorer2 [13] or `stable_tbn` [8] can compute the Hilbert basis directly from the monomers of the reduced network. The eleven unsplittable polymers of the reduced network are shown in Figure 3.

Next, by arguing about the upper bound on the entropy of an unsplittable configuration of the reduced network corresponding to the parameters of the $\mathbb{SPN}$, we will show that $|U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$ also acts as an upper bound on the entropy of a saturated configuration of an $\mathbb{SPN}$.

▶ **Lemma 5.** *If α is a saturated configuration of an $\mathbb{SPN}(\mathcal{R}, (m_i, M_i)_{i\in[n]}, k_I)$, $S(\alpha) \leq |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$.*

Proof. We show that $\sum_{i=1}^{n} m_i$ serves as an upper bound on the entropy of any saturated configuration of the reduced network. By Lemma 4, this bound also applies to the entropy of any saturated configuration of $\mathbb{SPN}$. As each module corresponding to a reaction has exactly one monomer mapped to a U3 monomer and exactly two monomers that are each mapped to an S2, the number of U3 monomers in the reduced network is equal to $\sum_{i=1}^{n} m_i$ and the number of S2 monomers is equal to $2\sum_{i=1}^{n} m_i$.

Recall that all stable configurations lie in the set of configurations consisting solely of unsplittable polymers. Therefore, it suffices to consider only such configurations. Take an arbitrary saturated configuration α' of the reduced network which consists only of unsplittable polymers by assigning counts $l_1, l_2, ..., l_{11}$ to the 11 unsplittable polymers, $P_1, P_2, ..., P_{11}$ respectively, shown in Figure 3.

The number of S2 monomers is:

$$l_5 + l_7 + l_8 + 2l_9 + 3l_{11} = 2\sum_{i=1}^{n} m_i$$

whereas the number of U3 monomers is:

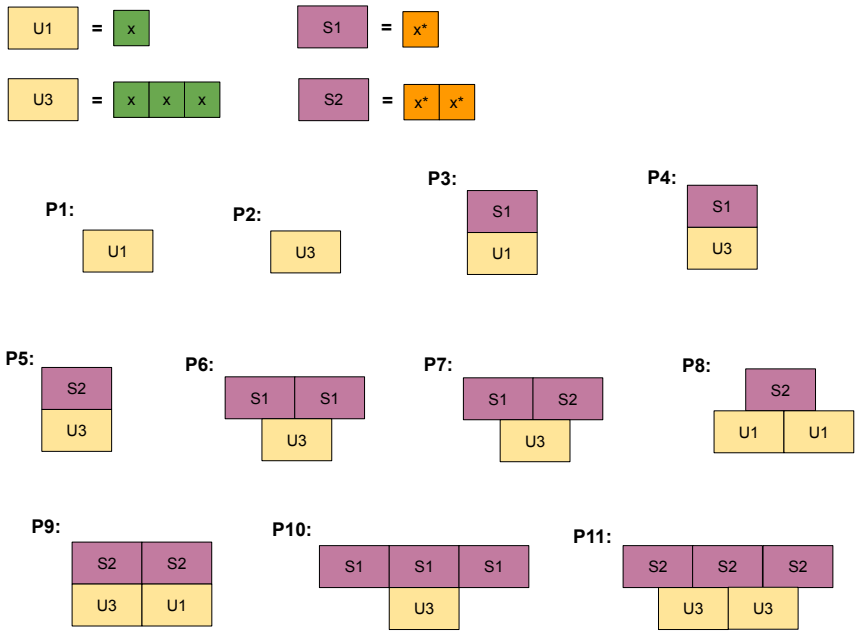

Figure 3 Unsplittable polymers (Hilbert basis) of the reduced network.

$$l_2 + l_4 + l_5 + l_6 + l_7 + l_9 + l_{10} + 2l_{11} = \sum_{i=1}^{n} m_i$$

If we subtract the second equation from the first equation:

$$l_8 + l_9 + l_{11} - l_2 - l_4 - l_6 - l_{10} = \sum_{i=1}^{n} m_i \tag{1}$$

which implies

$$l_8 + l_9 + l_{11} \geq \sum_{i=1}^{n} m_i$$

Note that only polymers P_8, P_9, and P_{11} merge two unstar monomers, each of which reduces the entropy by 1 from the maximum possible value, $|U_{\mathbb{SPN}}|$, upon polymer formation. Hence, $S(\alpha') = |U_{\mathbb{SPN}}| - (l_8 + l_9 + l_{11}) \leq |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$.

If we take an arbitrary saturated configuration α of the $\mathbb{SPN}$, $S(\alpha) = S(f_{\text{conf}}(\alpha)) \leq |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$ as $f_{\text{conf}}(\alpha)$ is a saturated configuration of the reduced network. ◀

▶ **Theorem 6.** *If α is a stable configuration of an $\mathbb{SPN}(\mathcal{R}, (m_i, M_i)_{i\in[n]}, k_I)$, $S(\alpha) = |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$.*

Proof. By Lemma 3, $S(\alpha) \geq |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$ and by Lemma 5, $S(\alpha) \leq |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$ and hence, $S(\alpha) = |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$. ◀

Our technique not only determines the entropy of a stable configuration but also lets us reason about the images of its polymers under f_{poly}.

▶ **Lemma 7.** *If α is a stable configuration of a given $\mathbb{SPN}$, for each polymer $P \in \alpha$, there exists $1 \leq i \leq 11$ s.t. $f_{poly}(P) = P_i$ where P_i is one of the polymers in Figure 3.*

Proof. Suppose α is a stable configuration and there exists $P \in \alpha$ such that $f_{\text{poly}}(P) = P'$ where P' is not one of the polymers in Figure 3. Then by the definition of polymer/Hilbert basis, P' is a splittable polymer which implies the existence of a saturated configuration α' of the reduced network such that $S(\alpha)+1 = S(f_{\text{conf}}(\alpha))+1 \leq S(\alpha')$ as α' can be obtained from $f_{\text{conf}}(\alpha)$ with at least one split operation. As shown in the proof of Lemma 5, the entropy of an arbitrary saturated configuration of the reduced network satisfies $S(\alpha') \leq |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$.

By using these two facts, we obtain $S(\alpha) < S(\alpha') \leq |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$ and by Theorem 6, α cannot be a stable configuration. Contradiction. ◀

As Lemma 7 restricts the images of polymers of a stable configuration to the unsplittable polymers of the reduced network, we can further eliminate more images:

▶ **Lemma 8.** *If α is a stable configuration of a given $\mathbb{SPN}$, for each polymer $P \in \alpha$, $f_{poly}(P) \in \{P_1, P_3, P_5, P_8, P_9\}$.*

Proof. If a monomer is mapped under f to an S1 monomer, it must contain exactly 2 starred sites; if it is mapped to an S2 monomer, it must contain exactly 4 starred sites; and if it is mapped to a U3 monomer, it must contain exactly 5 unstarred sites. Then, if $f_{\text{poly}}(P) = P_7$, P has at least one exposed starred site, which implies that P is not self-saturated. Similarly, if $f_{\text{poly}}(P) = P_{11}$, P has at least two exposed starred sites, which implies that P is not self-saturated. Hence, if there exists $P \in \alpha$ such that $f_{\text{poly}}(P) \in \{P_7, P_{11}\}$, α includes a polymer that is not self-saturated; hence α cannot be a saturated configuration, which implies that α is not a stable configuration.

Secondly, let the count of P_i in $f_{\text{conf}}(\alpha)$, where P_i is one of the polymers in Figure 3, be l_i. Suppose there exists $P \in \alpha$ such that $f_{\text{poly}}(P) \in \{P_2, P_4, P_6, P_{10}\}$. Then, at least for one $i \in \{2, 4, 6, 10\}$, $l_i > 0$. This implies, along with Equation (1) in Lemma 5, $l_8 + l_9 + l_{11} > \sum_{i=1}^{n} m_i$. Hence, $S(\alpha) = S(f_{conf}(\alpha)) < |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$ which, by Theorem 6, contradicts α being a stable configuration. ◀

Note that Theorem 6 assesses the entropy of a stable configuration and Lemma 8 shows the possible images of polymers in it for an $\mathbb{SPN}$ with an arbitrary set of reactions. From now on, we impose the following restrictions on the set of reactions of an $\mathbb{SPN}$:

- Two distinct reactions cannot share more than one species. For example, the pairs of reactions $A + B \longrightarrow C$ and $C + A \longrightarrow B$, as well as $A + B \longrightarrow C$ and $B \longrightarrow D + A$, are both forbidden. The only exception is when the reactions are reverses of each other; for instance, $A + B \longrightarrow C$ and $C \longrightarrow A + B$ are allowed.
- A reaction cannot include the same species more than once, e.g. $A+A \longrightarrow C$, $C+A \longrightarrow A$, $A \longrightarrow A + C$ or $A + A \longrightarrow A$ are all forbidden.

The next lemma shows that, with these restrictions, the set of possible image polymers can be further restricted:

▶ **Lemma 9.** *Let α be a stable configuration of a given $\mathbb{SPN}$. If $\mathcal{R}$ satisfies the restrictions above, $f_{poly}(P) \in \{P_1, P_3, P_5, P_8\}$ for each polymer $P \in \alpha$.*

Next, we will characterize the actual form of the polymers in a stable configuration of an $\mathbb{SPN}$ using the image polymers and the restrictions on the reactions. As a consequence of Lemma 9, only the polymers with image P_8 can include more than two monomers.

▶ **Lemma 10.** *In a stable configuration of the $\mathbb{SPN}$ including the reaction $A + B \longrightarrow C$ or $C \longrightarrow A + B$, if there is a polymer that includes the monomer $\{a_2^*, b_1^*, b_2^*, c_1^*\}$, it is either $\{\{a_2^*, b_1^*, b_2^*, c_1^*\}, \{a_2, b_1\}, \{b_2, c_1, c_2\}\}$ or $\{\{a_2^*, b_1^*, b_2^*, c_1^*\}, \{a_1, a_2, b_1, b_2, c_1\}\}$.*

▶ **Lemma 11.** *In a stable configuration of the $\mathbb{SPN}$ including the reaction $A + B \longrightarrow C$ or $C \longrightarrow A + B$, if there is a polymer including $\{a_1^*, a_2^*, b_1^*, b_2^*\}$, it is either $\{\{a_1^*, a_2^*, b_1^*, b_2^*\}$, $\{a_1, a_2, b_1, b_2, c_1\}\}$ or it includes three monomers: $\{a_1^*, a_2^*, b_1^*, b_2^*\}$ and two unstar monomers, one of which includes binding sites a_1 and a_2, and the other includes b_1 and b_2.*

Lemma 9, Lemma 10 and Lemma 11 also imply the form of polymers including the monomer $\{a_1, a_2, b_1, b_2, c_1\}$ and the ones including $\{a_2, b_1\}$.

▶ **Corollary 12.** *In a stable configuration of an $\mathbb{SPN}$ that includes the reaction $A + B \longrightarrow C$ or $C \longrightarrow A + B$, the module monomer $\{a_1, a_2, b_1, b_2, c_1\}$ is in a polymer either with the monomer $\{a_1^*, a_2^*, b_1^*, b_2^*\}$ or $\{a_2^*, b_1^*, b_2^*, c_1^*\}$.*

Proof. Lemma 9 shows that in a stable configuration, the only polymer that includes $\{a_1, a_2, b_1, b_2, c_1\}$ has image P_5, i.e. the monomer cannot be free. Lemma 10 and Lemma 11 imply that those polymers can only be of the form $\{\{a_1, a_2, b_1, b_2, c_1\}, \{a_1^*, a_2^*, b_1^*, b_2^*\}\}$ or $\{\{a_1, a_2, b_1, b_2, c_1\}, \{a_2^*, b_1^*, b_2^*, c_1^*\}\}$. ◀

▶ **Corollary 13.** *In a stable configuration of an $\mathbb{SPN}$ that includes a reaction $A + B \longrightarrow C$ or $C \longrightarrow A + B$, the module monomer $\{a_2, b_1\}$ is either in a polymer with $\{a_2^*, b_1^*, b_2^*, c_1^*\}$ and $\{b_2, c_1, c_2\}$ or free.*

Proof. Lemma 9 implies that $\{a_2, b_1\}$ is either free or in a polymer with image P_8. By Lemma 10 and Lemma 11, the only polymer with image P_8 containing $\{a_2, b_1\}$ is $\{\{a_2^*, b_1^*, b_2^*, c_1^*\}$, $\{a_2, b_1\}, \{b_2, c_1, c_2\}\}$. ◀

In light of Lemma 10, Lemma 11, Corollary 12 and Corollary 13, we extend the notion of an untriggered modular configuration. For the reaction $A + B \to C$, we define a modular configuration to be a *generalized untriggered modular configuration* if the monomer $\{c_1^*, c_2^*\}$ in Figure 1 (a) forms a polymer, which has image P_3, with an unstar monomer that includes the binding sites c_1 and c_2, i.e., a monomer of the form $\{c_1, c_2\}$ or $\{x_2, c_1, c_2\}$. On the other hand, for the reaction $C \to A + B$, we define a modular configuration to be a generalized untriggered modular configuration if the monomer $\{a_1^*, a_2^*, b_1^*, b_2^*\}$ in Figure 2 (a) forms a polymer, which has image P_8, with two unstar monomers, one including binding sites a_1 and a_2 and the other including b_1 and b_2, and the monomer $\{c_1^*, c_2^*\}$ forms a polymer, which has image P_3, with an unstar monomer including the binding sites c_1 and c_2. Observe that the untriggered modular configuration of a reaction is also a generalized untriggered modular configuration for that reaction. We denote a generalized untriggered modular configuration of a reaction R by C_R^{gen}. Next, we will characterize the structure of every stable configuration of an $\mathbb{SPN}$. First, we show the exact number of generalized untriggered modular configurations in a stable configuration as subconfigurations.

▶ **Lemma 14.** *Let $\mathbb{SPN}(\mathcal{R}, (m_i, M_i)_{i\in[n]}, k_I)$ be given. For each pair of reactions R_i and $R_{i'}$ that are reverses of each other, e.g. $A + B \longrightarrow C$ and $C \longrightarrow A + B$, the total number of $C_{R_i}^{gen}$ and $C_{R_{i'}}^{gen}$ as subconfigurations is exactly $m_i + m_{i'}$ in every stable configuration of the $\mathbb{SPN}$.*

Proof. W.l.o.g $R_i : A + B \longrightarrow C$ and $R_{i'} : C \longrightarrow A + B$ with m_i and $m_{i'}$ copies of module monomers for R_i and $R_{i'}$, possibly $m_i = 0$ or $m_{i'} = 0$. In total, $\mathbb{SPN}$ includes $m_i + m_{i'}$-many copies of each monomer $\{a_1, a_2, b_1, b_2, c_1\}$, $\{a_1^*, a_2^*, b_1^*, b_2^*\}$ and $\{a_2^*, b_1^*, b_2^*, c_1^*\}$.

By Lemma 11, the monomer $\{a_1^*, a_2^*, b_1^*, b_2^*\}$ is either in a polymer with two unstar monomers for which one includes binding sites a_1 and a_2 the other includes b_1 and b_2 (with image P_8) or in a polymer with $\{a_1, a_2, b_1, b_2, c_1\}$ (with image P_5). Let the number of the former in a stable configuration α be $t_{i'}$ and that of the latter be t_i, such that $t_i + t_{i'} = m_i + m_{i'}$. As there are $t_{i'}$-many $\{a_1, a_2, b_1, b_2, c_1\}$ left to consider, Corollary 12 implies that there are $t_{i'}$-many polymers of the form $\{\{a_1, a_2, b_1, b_2, c_1\}, \{a_2^*, b_1^*, b_2^*, c_1^*\}\}$. This, with Lemma 10, implies that there are t_i-many polymers of the form $\{\{a_2^*, b_1^*, b_2^*, c_1^*\}, \{a_2, b_1\}, \{b_2, c_1, c_2\}\}$. Furthermore, by Corollary 13, there are exactly $t_{i'}$-many free monomers of the form $\{a_2, b_1\}$.

Lastly, we consider the polymers including the monomer $\{c_1^*, c_2^*\}$. By Lemma 9, the only image polymer that can include $\{c_1^*, c_2^*\}$ is P_3. By restrictions, there are only two possible polymers $\{\{c_1, c_2\}, \{c_1^*, c_2^*\}\}$ and $\{\{x_2, c_1, c_2\}, \{c_1^*, c_2^*\}\}$, where x_2 could be of any species other than C, with image P_3. The total number $m_i + m_{i'}$ of $\{c_1^*, c_2^*\}$ in the modules for R_i and $R_{i'}$ implies that one can always consider t_i-many and $t_{i'}$-many of these polymers separately. By combining these results, α has exactly t_i-many $C_{R_i}^{gen}$ and $t_{i'}$-many $C_{R_{i'}}^{gen}$ as subconfigurations. As α is arbitrary, the lemma follows. ◀

Note that, in the proof of Lemma 14, for an arbitrary stable configuration, we consider the forms of every polymer that include a star monomer. As starred sites are limiting in the network, any polymer in a stable configuration that does not include a star monomer is necessarily a free unstar monomer. Each free unstar monomer of the form $\{a_2, b_1\}$ is already considered in a generalized untriggered modular configuration. This characterizes the structure of any stable configuration of an $\mathbb{SPN}$:

▶ **Corollary 15.** *Every stable configuration of an $\mathbb{SPN}$, whose underlying set of reactions satisfies the above restrictions, consists precisely of subconfigurations of the form C_R^{gen} for each reaction R, appearing with multiplicities as specified in Lemma 14, together with free unstarred monomers of the form $\{a_1, a_2\}$ or $\{x_2, a_1, a_2\}$.*

5 Simulating Chemical Reaction Networks with Signal Propagation Networks

▶ **Definition 16** (Discrete Chemical Reaction Network). *A* discrete chemical reaction network *(CRN) is a pair $\mathcal{C} = (\Lambda, \mathcal{R})$, where*

- *$\Lambda = \{X_1, \ldots, X_d\}$ is a finite set of* species*, and*
- *$\mathcal{R} \subset \mathbb{Z}^d$ is a finite set of* reactions*.*

Each reaction $r \in \mathcal{R}$ is described as $r = \beta_r - \alpha_r$ where $\alpha_r, \beta_r \in \mathbb{N}^d$ are vectors called the reactant *and* product *stoichiometries, respectively.*

A configuration *of the system is a vector $x \in \mathbb{N}^d$, where x_i represents the number of molecules of species X_i. A reaction r is* enabled *at configuration x if $x \geq \alpha_r$ (componentwise). If reaction r occurs, the system moves from configuration x to $x' = x - \alpha_r + \beta_r$. This is written as $x \xrightarrow{r} x'$.*

A reaction $r = \beta_r - \alpha_r$ of a CRN can also be represented in stoichiometric form R, e.g., CRN $\mathcal{C} = (\{A, B, C, ...\}, \{(-1, -1, 1, 0, 0, ..., 0), ...\})$ includes a reaction R of the form $A + B \longrightarrow C$. As the reaction modules are reversible in the sense described in Section 3,

we consider reversible discrete CRNs where for each reaction, the reverse reaction is also in the set, i.e., for all $r \in \mathcal{R}$, $-r \in \mathcal{R}$. From now on, we use CRN to refer to discrete reversible CRNs.

There is a natural correspondence between $\mathbb{SPN}$s and CRNs. Every $\mathbb{SPN}(\mathcal{R}, (m_i, M_i)_{i\in[n]}, k_I)$ for which $\mathcal{R}$ includes both the forward and backward directions for each reaction induces a CRN $\mathcal{C} = (\Lambda, \mathcal{R})$, where $\Lambda \leftrightarrow \bigcup_i I_{R_i} = I$, i.e., the input species of the CRN correspond exactly to the input monomer types of the $\mathbb{SPN}$. As $\mathcal{R}$ is reversible, I includes every input and output monomer in the $\mathbb{SPN}$. We refer to the CRN that is constructed from an $\mathbb{SPN}$ in this way as the *induced CRN*. Similarly, each CRN corresponds to a family of $\mathbb{SPN}$s that share the same set of reactions – including the associated module and input monomers for each reaction – and differ only in the quantities assigned to those monomers.

Furthermore, a natural mapping takes an $\mathbb{SPN}$ configuration and outputs the number of free input monomers – i.e., those that exist as single-monomer polymers, for each input monomer type. We will use a slightly more generalized mapping, consistent with the generalization of untriggered modular configuration, which outputs not only the number of free monomers of the form $\{a_1, a_2\}$ for species A but also those of the form $\{x_2, a_1, a_2\}$. Note that the monomer $\{x_2, a_1, a_2\}$ is included as a module monomer of a reaction of the form $_ + X \longrightarrow A$ or $A \longrightarrow _ + X$. For a given $\mathbb{SPN}$, we denote this mapping as $\phi_{\mathbb{SPN}}$. Next, we show that, under mapping $\phi_{\mathbb{SPN}}$, the behavior of a CRN can be explained by the thermodynamically favorable configurations of a suitable $\mathbb{SPN}$.

The theorem below gives one direction of a simulation desideratum in the sense that any path of a CRN can be mimicked by the corresponding $\mathbb{SPN}$.

▶ **Theorem 17.** *For every reversible CRN $\mathcal{C} = (\Lambda, \mathcal{R})$, where $\mathcal{R}$ satisfies the restrictions defined in Section 4, every initial configuration $x_{\text{init}} \in \mathbb{N}^{|\Lambda|}$ of the CRN, and every $t \in \mathbb{N}$, there exists an $\mathbb{SPN}\big(\mathcal{R}, \big((m_i, M_i)_{i\in[n]}, (k_I)\big)\big)$ and a stable configuration α of the $\mathbb{SPN}$ such that $\phi_{\mathbb{SPN}}(\alpha) = x_{\text{init}}$, with the following property: for every configuration c of the CRN that is reachable from x_{init} by a path of length at most t, there exists a stable configuration β of the $\mathbb{SPN}$ such that $\phi_{\mathbb{SPN}}(\beta) = c$, and β is reachable from α by a height-2 path.*

Proof. For an arbitrary initial configuration x_{init} of CRN $\mathcal{C} = (\Lambda, \mathcal{R})$ and a given $t \in \mathbb{N}$, construct $\mathbb{SPN}\big(\mathcal{R}, \big(t, M_i\big)_{i\in[n]}, x_{init}\big)$. We consider a configuration α of the $\mathbb{SPN}$ where all the input monomers with counts in x_{init} are free polymers and the rest of the configuration consists of t-many untriggered modular configurations for each reaction $R \in \mathcal{R}$. First, note that $\phi_{\mathbb{SPN}}(\alpha) = x_{init}$. Secondly, as each untriggered modular configuration performs one merge of unstar monomers, which causes a drop of 1 from the maximum possible entropy, $|U_{\mathbb{SPN}}|$, $S(\alpha) = |U_{\mathbb{SPN}}| - tn$. By Theorem 6, α is a stable configuration.

Let c be a CRN configuration s.t. there exists a path $x_{init} = x_0 \xrightarrow{R_{i_0}} x_1 \xrightarrow{R_{i_1}} x_2 \xrightarrow{R_{i_2}} \ldots \xrightarrow{R_{i_{t'-1}}} c = x_{t'-1}$ where $t' \leq t$ and each step corresponds to a transition implied by a reaction $R_{i_k} \in \mathcal{R}$. Consider the transitions in the $\mathbb{SPN}$ such that at each step k, a previously unfired reaction module for R_{i_k} is fired. Let the corresponding sequence of $\mathbb{SPN}$ configurations be $\alpha \to \gamma_1 \to \gamma_2 \to \ldots \to \gamma_{t'-2} \to \beta$. As α includes t-many reaction modules for each $R \in \mathcal{R}$ and $t' \leq t$, it is always possible to find an unfired module along the sequence. Hence, $\phi_{\mathbb{SPN}}(\gamma_j) = x_j$ for $j \in \{1, 2, \ldots, t'-1\}$ and $\phi_{\mathbb{SPN}}(\beta) = c$. Note that firing a reaction module does not change the total number of unstar merges in the system, as each reaction module includes exactly one polymer merging two unstar monomers (Figure 1 (b) to (d) and Figure 2 (b) to (d)). Recall that the entropies of two configurations are equal if and only if the total number of unstar merges is the same in the configurations. Hence, $S(\alpha) = S(\gamma_1) = S(\gamma_2) = \ldots = S(\gamma_{t'-1}) = S(\beta)$. As firing a reaction results in another

saturated configuration ((Figure 1 (d) and Figure 2 (d)) and α is a stable configuration, γ_1, γ_2,...,$\gamma_{t'-1}$ and β are all stable configurations. Finally, as firing each module necessitates a separate height-2 path, the overall path that takes α to β is also a height-2 path. ◀

Next, we consider the reverse direction and show that the corresponding images, under $\phi_{\mathbb{SPN}}$, of each pair of stable configurations of an $\mathbb{SPN}$ are reachable, with an appropriate definition of reachability, from one another in the induced CRN.

▶ **Definition 18** (Weak reachability in CRNs). *Let a CRN $\mathcal{C} = (\Lambda, \mathcal{R})$ where $\mathcal{R} = (r_1, r_2, ..., r_k)$ be given. We consider two configurations c_1 and c_2 of $\mathcal{C}$ and say c_2 is weakly reachable from c_1 if there exists $\gamma_1, \gamma_2, ..., \gamma_k \in \mathbb{N}$ s.t. $c_2 = c_1 + \sum_{i=1}^{k} \gamma_i r_i$.*

Our definition of weak reachability corresponds to being in the same stoichiometric compatibility class as defined in the Chemical Reaction Network Theory literature [1].

When the CRN is reversible, c_2 being weakly reachable from c_1 also implies c_1 being weakly reachable from c_2. Next, we show that the stable configurations of an arbitrary $\mathbb{SPN}$ correspond to configurations of the induced CRN that are weakly reachable.

▶ **Theorem 19.** *For an arbitrary $\mathbb{SPN}(\mathcal{R}, (m_i, M_i)_{i\in[n]}, k_I)$, where $\mathcal{R}$ satisfies the restrictions in Section 4, let $\mathcal{C} = (\Lambda, \mathcal{R})$ be the induced CRN. For every pair α and β of stable configurations of the $\mathbb{SPN}$, $\phi_{\mathbb{SPN}}(\alpha)$ and $\phi_{\mathbb{SPN}}(\beta)$ are weakly reachable in the CRN.*

Proof. For $0 < i \leq \frac{n}{2}$, let $R_i : X_{i_1} + X_{i_2} \longrightarrow X_{i_3}$ and $R_{i'} : X_{i_3} \longrightarrow X_{i_1} + X_{i_2}$, where $X_{i_1}, X_{i_2}, X_{i_3}$ denote the input species with indices $i_1, i_2, i_3 \in [|\Lambda|]$. Let g_i and $g_{i'}$ be the amount of subconfigurations of $C^{gen}_{R_i}$ and $C^{gen}_{R_{i'}}$ in α and let h_i and $h_{i'}$ be those in β. By Lemma 14, $g_i + g_{i'} = h_i + h_{i'} = m_i + m_{i'}$, which implies $g_i - h_i = h_{i'} - g_{i'}$. Note that Lemma 14 also holds when $m_i = 0$ and/or $m_{i'} = 0$ and hence, if $\mathcal{R}$ is not reversible, it can be expanded to include reverse reactions without any copies of module monomers for those reactions.

We will argue based on the counts of the monomers, with image under $\phi_{\mathbb{SPN}}$ when free, in α and β. Note that $\phi_{\mathbb{SPN}}(\alpha)$ and $\phi_{\mathbb{SPN}}(\beta)$ already capture the counts of the free monomers of the form $\{a_1, a_2\}$ and $\{x_2, a_1, a_2\}$ in these configurations. In each $C^{gen}_{R_i}$, there are exactly two monomers which are mapped to X_{i_3} and no monomers mapped to X_{i_1} or X_{i_2} under $\phi_{\mathbb{SPN}}$, when free. In each $C^{gen}_{R_{i'}}$, there are exactly three monomers mapped to X_{i_1}, X_{i_2} and X_{i_3}, respectively, under $\phi_{\mathbb{SPN}}$, when free. Hence, by the conservation of the number of monomers in α and β:

$$\phi_{\mathbb{SPN}}(\alpha)+\sum_{i=1}^{n/2}(2g_ie_{i_3}+g_{i'}e_{i_1}+g_{i'}e_{i_2}+g_{i'}e_{i_3}) = \phi_{\mathbb{SPN}}(\beta)+\sum_{i=1}^{n/2}(2h_ie_{i_3}+h_{i'}e_{i_1}+h_{i'}e_{i_2}+h_{i'}e_{i_3}) \quad (2)$$

where $e_{i_1}, e_{i_2}, e_{i_3}$ are the i_1st, i_2nd and i_3rd unit vectors of dimension $|\Lambda|$. Using $g_i - h_i = h_{i'} - g_{i'}$ and rearranging the terms in Equation (2) yields

$$\phi_{\mathbb{SPN}}(\alpha) + \sum_{i=1}^{n/2}((g_i - h_i)e_{i_3} + (g_{i'} - h_{i'})e_{i_1} + (g_{i'} - h_{i'})e_{i_2}) = \phi_{\mathbb{SPN}}(\beta)$$

If $g_i - h_i = -(g_{i'} - h_{i'}) \geq 0$, $((g_i - h_i)e_{i_3} + (g_{i'} - h_{i'})e_{i_1} + (g_{i'} - h_{i'})e_{i_2}) = (g_i - h_i)r_i$ and if $g_i - h_i = -(g_{i'} - h_{i'}) \leq 0$, $((g_i - h_i)e_{i_3} + (g_{i'} - h_{i'})e_{i_1} + (g_{i'} - h_{i'})e_{i_2}) = (g_{i'} - h_{i'})r_{i'}$ where r_i and $r_{i'}$ are the CRN transitions corresponding to the reactions R_i and $R_{i'}$, respectively. The theorem follows. ◀

6 Two-reactant, two-product reversible signal propagation modules

The reversible signal propagation module studied above has been previously described and experimentally implemented [3, 17]. However, it has the limitation of having only one product (or reactant for the backward module). In Appendix A, we introduce a new reversible signal propagation module with two reactants and two products, potentially enabling a more compact realization of CRNs. Importantly, testifying to the generality of our proof technique, we show that an appropriately chosen reduction mapping can be used to determine the exact entropy of a stable configuration of an $\mathbb{SPN}$ with an arbitrary number of reaction modules for this system as well.

The reduction mapping we employ accounts for the total number of 1-unstarred (starred) and 2-unstarred (starred) sites in each monomer. In contrast, if one considers an alternative reduction mapping that tracks only the number of 1-unstarred (starred) sites, which gives the same reduced network if 2-unstarred (starred) sites or 3-unstarred (starred) sites are considered as the construction is symmetric, the resulting lowest upper bound on the entropy of a saturated configuration becomes trivial and thus uninformative. This demonstrates that the choice of reduction mapping for analyzing the entropy of stable configurations in an $\mathbb{SPN}$ is non-trivial. For space considerations, full details are deferred to Appendix A and Appendix B.

7 Discussion

We present a general method for analyzing thermodynamically favorable configurations of modular signal propagation networks, with the goal of showing that modules verified in isolation remain correct under composition. Our approach relies on constructing a reduction mapping that transforms the original system into a coarser network, enabling tight upper bounds on the entropy of saturated configurations. We show that this reduction can map networks composed of arbitrarily many reaction modules to a single reduced module.

The reduced system is small enough to make the computation of its Hilbert basis and the characterization of its stable configurations tractable, thereby avoiding the computational hardness of the original problem. Using this framework, we lift entropy bounds and structural properties from the reduced system back to the original network, establishing that modular $\mathbb{SPN}$s can be safely composed. We subsequently show that under certain restrictions on the reaction set, these networks can simulate a class of reversible chemical reaction networks: every bounded-length CRN execution can be realized by a height-2 path between stable TBN configurations, and any two stable TBN configurations map to CRN states in the same stoichiometric compatibility class.

Besides applying our method to the previously described reversible signal propagation module, an additional contribution of this paper is to introduce a novel reversible signal propagation module with two inputs and two outputs. In general, we caution the reader that the choice of reduction mapping is critical. While an appropriate reduction preserves enough structure to yield tight entropy bounds, coarser reductions fail to capture essential combinatorial information and lead only to trivial bounds (as shown in Appendix B).

Several directions remain open. First, it is natural to ask whether the constructions in this paper can be extended to simulate arbitrary reversible CRNs directly without structural restriction, and how the proof technique can be adapted accordingly. Second, adding redundancy in binding-site types and monomer structure, may allow control over the distance to stability of leaked configurations [7]; we conjecture that this can be captured via extensions of our reduction-based analysis. Finally, it would be interesting to characterize all height-2 paths in an $\mathbb{SPN}$ and relate them to trajectories in the corresponding CRN.

References

1 David F Anderson, Gheorghe Craciun, and Thomas G Kurtz. Product-form stationary distributions for deficiency zero chemical reaction networks. *Bulletin of mathematical biology*, 72(8):1947–1970, 2010.

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

3 Keenan Breik, Chris Thachuk, Marijn Heule, and David Soloveichik. Computing properties of stable configurations of thermodynamic binding networks. *Theoretical Computer Science*, 785:17–29, 2019. doi:10.1016/J.TCS.2018.10.027.

4 Robert Brijder. Computing with chemical reaction networks: a tutorial. *Natural Computing*, 18(1):119–137, 2019. doi:10.1007/S11047-018-9723-9.

5 Cameron Chalk, Jacob Hendricks, Matthew J Patitz, and Michael Sharp. Thermodynamically favorable computation via tile self-assembly. In *International Conference on Unconventional Computation and Natural Computation*, pages 16–31. Springer, 2018. doi:10.1007/978-3-319-92435-9_2.

6 David Doty. Theory of algorithmic self-assembly. *Communications of the ACM*, 55(12):78–88, 2012. doi:10.1145/2380656.2380675.

7 David Doty, Trent A Rogers, David Soloveichik, Chris Thachuk, and Damien Woods. Thermodynamic binding networks. In *International Conference on DNA-Based Computers*, pages 249–266. Springer, 2017. doi:10.1007/978-3-319-66799-7_16.

8 David Haley. StableTBN. https://github.com/drhaley/stable_tbn, 2021. Accessed: 2025-09-03.

9 David Haley and David Doty. Computing properties of thermodynamic binding networks: An integer programming approach. In Matthew R. Lakin and Petr Šulc, editors, *27th International Conference on DNA Computing and Molecular Programming (DNA 27)*, volume 205 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 2:1–2:16, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.DNA.27.2.

10 Austin Luchsinger. Brief announcement: Barrier-1 reachability for thermodynamic binding networks is PSPACE-complete. In *1st Symposium on Algorithmic Foundations of Dynamic Networks (SAND 2022)*, pages 24–1. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPIcs.SAND.2022.24.

11 Joshua Petrack, David Soloveichik, and David Doty. Thermodynamically driven signal amplification. In Ho-Lin Chen and Constantine G. Evans, editors, *29th International Conference on DNA Computing and Molecular Programming (DNA 29)*, volume 276 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 8:1–8:22, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.DNA.29.8.

12 Andrew Phillips and Luca Cardelli. A programming language for composable DNA circuits. *Journal of the Royal Society Interface*, 6(suppl_4):S419–S436, 2009.

13 David Soloveichik. TBNExplorer2: A Python library and command-line tools for analyzing thermodynamic binding networks, 2025. URL: https://github.com/david-soloveichik/TBNExplorer2.

14 David Soloveichik, Matthew Cook, Erik Winfree, and Jehoshua Bruck. Computation with finite stochastic chemical reaction networks. *Natural Computing*, 7(4):615–633, 2008. doi:10.1007/S11047-008-9067-Y.

15 David Soloveichik, Georg Seelig, and Erik Winfree. DNA as a universal substrate for chemical kinetics. *Proceedings of the National Academy of Sciences*, 107(12):5393–5398, 2010.

16 David Soloveichik and Capstone Team. stablegen.net: Free energies and concentrations of domain-level DNA designs. https://stablegen.net, 2024. Accessed: 2024-08-30.

17 Boya Wang, Cameron Chalk, David Doty, and David Soloveichik. Molecular computation at equilibrium via programmable entropy. *Science Advances*, 12(2):eadx3969, 2026.

18 And Kaan Ata Yilmaz, David Soloveichik, and Austin Luchsinger. Thermodynamic binding networks fully capture equilibrium monomer-polymer thermodynamics. In *32nd International Conference on DNA Computing and Molecular Programming (DNA32)*, 2026. doi:10.4230/LIPIcs.DNA.32.9.

A Two-reactant, two-product reversible signal propagation modules

In this section, we introduce a novel reversible signal propagation module. Unlike the system of Section 3, the new construction directly implements a reversible reaction with two reactants and two products such as $A + B \longrightarrow C + D$ and thus could realize CRNs more compactly, without requiring a decomposition of such a reaction into $A + B \longrightarrow X$ and $X \longrightarrow C + D$. Additionally, we are motivated to demonstrate that our proof technique for showing the correctness of such modules is general enough to cover multiple distinct constructions.

The new system consists of reaction modules of the form $R : A + B \longrightarrow C + D$. The *untriggered modular configuration* for this reaction module is shown in Figure 4 (a) and the two stable configurations in the presence of input monomers for the module are shown in Figure 4 (b), (c). The correspondence between species and monomers is the following: $\{x_1, x_2, x_3\} \leftrightarrow X$. As in Section 3, the fact that (a)-(c) describe the unique stable configurations of an isolated module can be easily checked via tools such as `StableGen` or `stable_tbn` [3, 9, 16, 8].

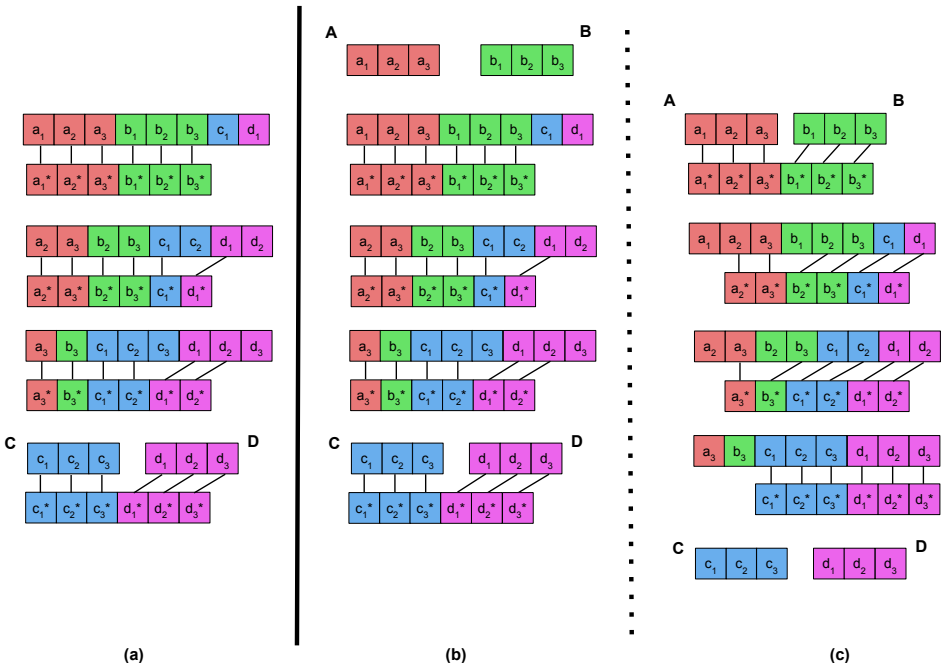

Figure 4 (a) The only stable configuration of the module for the reaction $A + B \longrightarrow C + D$; (b), (c) two stable configurations when both A and B are present.

Module monomers, input monomers, output monomers and $\mathbb{SPN}\big(\mathcal{R}, ((m_i, M_i)_{i\in[n]}), k_I\big)$, are defined analogously to the first system. The reduction map f that we use for this system outputs a monomer of length equal to the total number of 1-unstarred (starred) and 2-unstarred (starred) sites for unstar (star) monomers. Explicitly,

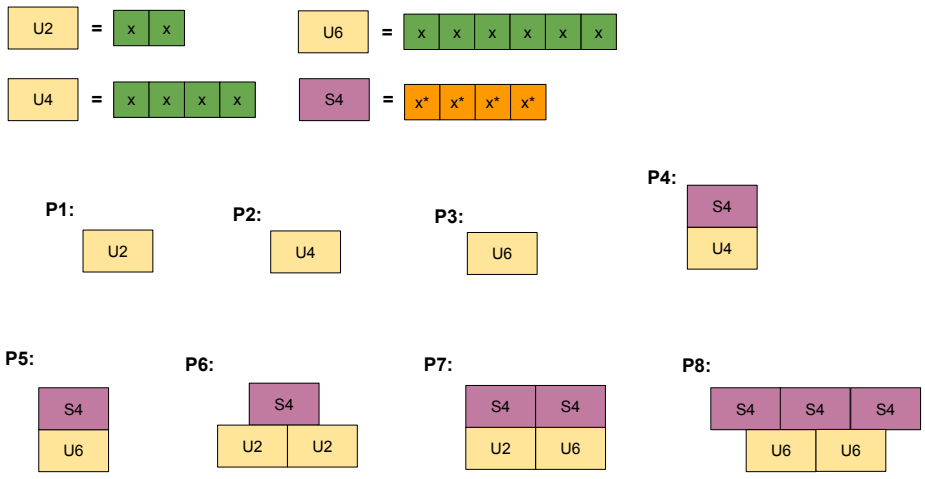

Figure 5 Unsplittable polymers (Hilbert basis) of the reduced network for the new system.

- $f(\{a_1, a_2, a_3, b_1, b_2, b_3, c_1, d_1\}) = f(\{a_2, a_3, b_2, b_3, c_1, c_2, d_1, d_2\}) = \{x, x, x, x, x, x\}$
- $f(\{a_1^*, a_2^*, a_3^*, b_1^*, b_2^*, b_3^*\}) = f(\{a_2^*, a_3^*, b_2^*, b_3^*, c_1^*, d_1^*\}) = f(\{a_3^*, b_3^*, c_1^*, c_2^*, d_1^*, d_2^*\}) = f(\{c_1^*, c_2^*, c_3^*, d_1^*, d_2^*, d_3^*\}) = \{x^*, x^*, x^*, x^*\}$
- $f(\{a_3, b_3, c_1, c_2, c_3, d_1, d_2, d_3\}) = \{x, x, x, x\}$
- $f(\{a_1, a_2, a_3\}) = f(\{b_1, b_2, b_3\}) = f(\{c_1, c_2, c_3\}) = f(\{d_1, d_2, d_3\}) = \{x, x\}$

f_{poly} and f_{conf} are defined accordingly. As each monomer type has an image under f and f preserves the total number of 1-unstarred (starred) and 2-unstarred (starred) sites of unstar (star) monomers, if a polymer P is self-saturated, $f_{\text{poly}}(P)$ is also self-saturated, which further implies that if a configuration α of the $\mathbb{SPN}$ is saturated, then $f_{\text{conf}}(\alpha)$ is also saturated. Moreover, $S(\alpha) = S(f_{\text{conf}}(\alpha))$. Then, we can invoke Lemma 4 and find an upper bound on the entropy of the reduced network by analyzing its unsplittable polymers. The unsplittable polymers (Hilbert basis) of the reduced network are shown in Figure 5.

Next, we will assess the entropy of a stable configuration, with respect to the parameters of the $\mathbb{SPN}$ in a way that is the direct analogue of Lemma 3 and Lemma 5.

▶ **Lemma 20.** *If α is a stable configuration of $\mathbb{SPN}(\mathcal{R}, (m_i, M_i)_{i\in[n]}, k_I)$, then $S(\alpha) \geq |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$.*

Proof. The proof follows the same lines as the proof of Lemma 3. Consider configuration α' that consists of the untriggered modular configuration of each reaction $R \in \mathcal{R}$ (Figure 4 (a)), with respective multiplicities, as subconfigurations and free input monomers. As each polymer is self-saturated, α' is saturated. Furthermore, as each untriggered modular configuration includes exactly one polymer that merges two unstar monomers, $S(\alpha') = |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i \leq S(\alpha)$. ◀

▶ **Lemma 21.** *If α is a saturated configuration of an $\mathbb{SPN}(\mathcal{R}, (m_i, M_i)_{i\in[n]}, k_I)$, $S(\alpha) \leq |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$.*

Proof. As in Lemma 5, we will use the unsplittable polymers of the reduced network to show an upper bound on the entropy of a stable configuration of the reduced network. Take an arbitrary unsplittable configuration α' of the reduced network by assigning counts $l_1, l_2, ..., l_8$ to the 8 unsplittable polymers, $P_1, P_2, ..., P_8$, respectively, shown in Figure 5. Each module includes two U6, four S4 and one U4 monomer.

The number of U6 monomers is:

$$l_3 + l_5 + l_7 + 2l_8 = 2\sum_{i=1}^{n} m_i \tag{3}$$

The number of S4 monomers is:

$$l_4 + l_5 + l_6 + 2l_7 + 3l_8 = 4\sum_{i=1}^{n} m_i \tag{4}$$

The number of U4 monomers is:

$$l_2 + l_4 = \sum_{i=1}^{n} m_i \tag{5}$$

Note that only P_6, P_7 and P_8 merge two unstar monomers, each of which reduces the entropy by 1 from the maximum possible value, $|U_{\mathbb{SPN}}|$, upon polymer formation. Hence, $S(\alpha') = |U_{\mathbb{SPN}}| - (l_6 + l_7 + l_8)$.

If we subtract Equation (3) from Equation (4), we get $l_6 + l_7 + l_8 + l_4 - l_3 = 2\sum_{i=1}^{n} m_i$, which implies $l_6 + l_7 + l_8 + l_4 \geq 2\sum_{i=1}^{n} m_i$ as $l_3 \geq 0$. By Equation (5), $l_4 \leq \sum_{i=1}^{n} m_i$, which implies $l_6 + l_7 + l_8 \geq \sum_{i=1}^{n} m_i$. Hence, $S(\alpha') \leq |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$.

Therefore, for any saturated configuration α of $\mathbb{SPN}$, $S(\alpha) = S(f_{\text{conf}}(\alpha)) \leq |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$. ◀

▶ **Theorem 22.** *Let α be a stable configuration of $\mathbb{SPN}(\mathcal{R}, (m_i, M_i)_{i\in[n]}, k_I)$, then $S(\alpha) = |U_{\mathbb{SPN}}| - \sum_{i=1}^{n} m_i$.*

Proof. Direct consequence of Lemma 20 and Lemma 21. ◀

B The Choice of Reduction Mapping is Non-Trivial

We will show that applying the same method with different reduction mappings might not always be able to assess the entropy of a stable configuration of the $\mathbb{SPN}$. To show this, for the $\mathbb{SPN}$ consisting of modules in Figure 4, instead of the reduction mapping that outputs a monomer with length equal to the total number of 1-unstarred (starred) and 2-unstarred (starred) sites for an unstar (star) monomer, we consider f' which outputs a monomer with length equal to the number of 1-unstarred (starred) sites. Explicitly,

- $f'(\{a_1, a_2, a_3, b_1, b_2, b_3, c_1, d_1\}) = \{x, x, x, x\}$
- $f'(\{a_1^*, a_2^*, a_3^*, b_1^*, b_2^*, b_3^*\}) = f(\{a_2^*, a_3^*, b_2^*, b_3^*, c_1^*, d_1^*\}) = f'(\{a_3^*, b_3^*, c_1^*, c_2^*, d_1^*, d_2^*\}) = f'(\{c_1^*, c_2^*, c_3^*, d_1^*, d_2^*, d_3^*\}) = \{x^*, x^*\}$
- $f'(\{a_3, b_3, c_1, c_2, c_3, d_1, d_2, d_3\}) = f'(\{a_2, a_3, b_2, b_3, c_1, c_2, d_1, d_2\}) = \{x, x\}$
- $f'(\{a_1, a_2, a_3\}) = f'(\{b_1, b_2, b_3\}) = f'(\{c_1, c_2, c_3\}) = f'(\{d_1, d_2, d_3\}) = \{x\}$

The unsplittable polymers (Hilbert basis) of the system are shown in Figure 6. We will show that, by using the reduced network induced by f', the smallest upper bound on the entropy of a saturated configuration α is $|U_{\mathbb{SPN}}|$:

▶ **Lemma 23.** *Let α' be a saturated configuration of the reduced network under f'. $S(\alpha') \leq |U_{\mathbb{SPN}}|$ and the bound is tight, i.e., there exists a saturated α' such that $S(\alpha') = |U_{\mathbb{SPN}}|$.*

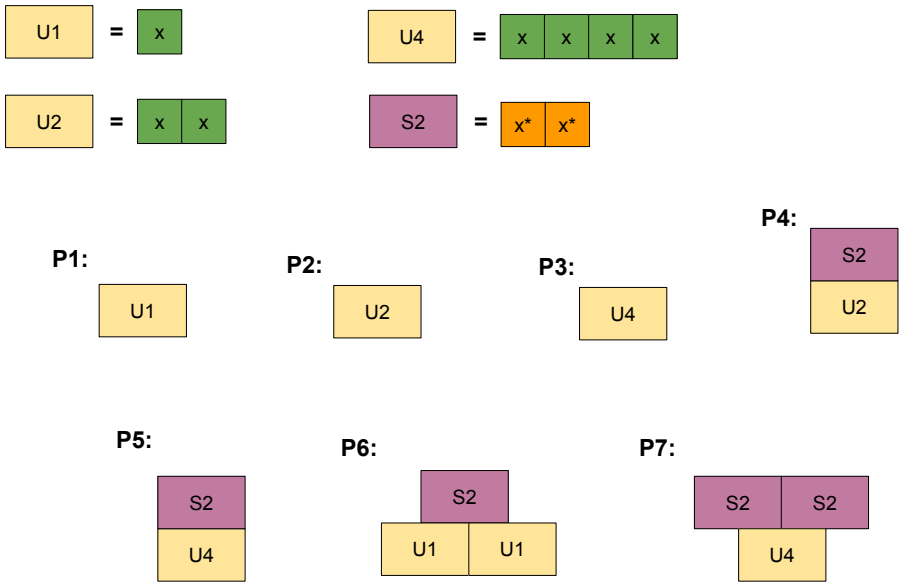

Figure 6 Unsplittable polymers (Hilbert basis) of the reduced network, under f', for the new system.

Proof. Take an unsplittable configuration α' of the reduced network by assigning counts $l_1, l_2, ..., l_7$ to the 7 unsplittable polymers, $P_1, P_2, ..., P_7$, shown in Figure 6. Each module includes one U4, four S2 and two U2 monomers.

The number of U4 monomers is:

$$l_3 + l_5 + l_7 = \sum_{i=1}^{n} m_i$$

The number of S2 monomers is:

$$l_4 + l_5 + l_6 + 2l_7 = 4\sum_{i=1}^{n} m_i$$

The number of U2 monomers is:

$$l_2 + l_4 = 2\sum_{i=1}^{n} m_i$$

Note that the assignment $l_2 = l_3 = l_5 = l_6 = 0$, $l_7 = \sum_{i=1}^{n} m_i$ and $l_4 = 2\sum_{i=1}^{n} m_i$ satisfies all constraints. Consider a saturated configuration α' which consists of $(2\sum_{i=1}^{n} m_i)$-many P_4 and $(\sum_{i=1}^{n} m_i)$-many P_7, with all U1 monomers free, i.e. in polymer P_1. As α' does not include P_6, $S(\alpha') = |U_{\mathbb{SPN}}| - l_6 = |U_{\mathbb{SPN}}|$. ◀

C Omitted Proofs

Proof of Lemma 9. We will do a case-by-case analysis on the two possible forms of S2 monomers to show that P_9 is not an allowed image under the restrictions. Note that, as two S2 monomers include a total of eight starred sites, the same argument based on self-saturation presented in Lemma 8 can be invoked to show that U1 and U3 monomers that saturate two S2 monomers include three and five unstarred sites, respectively, i.e. an unstar monomer of the form $\{x_1, x_2\}$ cannot be included in a polymer with image P_9 and hence, the polymer has no exposed sites.

For each case, we will argue on the basis of the U1 monomer with three unstarred sites. Hence, all monomers that we consider below are included in the modules of a reaction and of its reverse. As the restrictions, when applicable, are not direction-dependent, without loss of generality, we will show contradictions for the reactions in the forward direction.

Case 1: S2 $= \{x_1^*, x_2^*, y_1^*, y_2^*\}$ and S2 $= \{t_1^*, t_2^*, z_1^*, z_2^*\}$.

The reactions that can include these monomers as their module monomers have the form: $X + Y \longrightarrow _$ and $Z + T \longrightarrow _$. The character '_' is a placeholder for an arbitrary species (and for an arbitrary unstarred site in the following):

Case 1.1: If U1$= \{x_2, _, y_2\}$, there exists a reaction $_ + Y \longrightarrow X$ or $_ + X \longrightarrow Y$ which are not allowed by the restrictions. A symmetric argument can be made for a monomer of the form $\{t_2, _, z_2\}$.

Case 1.2: If U1$= \{x_2, t_1, t_2\}$, U3$= \{y_1, y_2, z_1, z_2, x_1\}$. U3 monomer implies the existence of a forbidden reaction $Y + Z \longrightarrow X$ or $Z + Y \longrightarrow X$. The same argument can be made for a monomer of the form $\{x_2, z_1, z_2\}$, $\{y_2, z_1, z_2\}$, $\{y_2, t_1, t_2\}$,$\{z_2, x_1, x_2\}$, $\{z_2, y_1, y_2\}$, $\{t_2, x_1, x_2\}$, $\{t_2, y_1, y_2\}$.

Case 1.3: If U1$= \{x_2, y_1, t_2\}$, U3$= \{t_1, y_2, z_1, z_2, x_1\}$. U1 implies either $X = Y$, which is forbidden, or $Y = T$, which would imply U3 monomer being in the form $\{y_1, y_2, z_1, z_2, x_1\}$ and hence, the existence of a reaction of the form $Y + Z \longrightarrow X$ (or $Z + Y \longrightarrow X$). The same argument can be made for $\{x_2, z_1, t_2\}, \{x_2, t_1, z_2\}, \{x_2, y_1, z_2\}, \{y_2, z_1, t_2\}, \{y_2, x_1, t_2\}$, $\{y_2, x_1, z_2\}, \{y_2, t_1, z_2\}$.

Case 2: S2$= \{x_1^*, x_2^*, y_1^*, y_2^*\}$ and S2 $= \{t_2^*, z_1^*, z_2^*, u_1^*\}$

The reactions that can include these monomers as their module monomers have the form $X + Y \longrightarrow _$ and $T + Z \longrightarrow U$.

Case 2.1: If U1 monomer is $\{x_2, _, y_2\}$, there exists a reaction $_ + X \longrightarrow Y$ or $_ + Y \longrightarrow X$. The same argument can be made for a monomer of the form $\{t_2, _, z_2\}$.

Case 2.2: If U1$= \{t_2, x_1, x_2\}$, U3 $= \{y_1, y_2, z_1, z_2, u_1\}$. These forms of U1 and U3 imply two reactions $_ + T \longrightarrow X$ and $Y + Z \longrightarrow U$ (or $Z + Y \longrightarrow U$, which is already forbidden). If $T \neq Y$, the second reaction is forbidden. Else, the first reaction is forbidden. The same argument can be made for $\{t_2, y_1, y_2\}$.

Case 2.3: If U1$= \{x_2, y_1, t_2\}$. U1 monomer implies either $X = Y$, which is forbidden or $T = Y$, which would imply the existence of a reaction $_ + X \longrightarrow Y$ that is forbidden. The same argument can be made for $\{y_2, x_1, t_2\}, \{x_2, z_1, t_2\}, \{y_2, z_1, t_2\}, \{y_2, u_1, t_2\}, \{x_2, u_1, t_2\}$.

Case 2.4 : If U1 $= \{x_2, x_1, z_2\}$, U3 $= \{y_1, y_2, z_1, t_2, u_1\}$. U3 monomer implies that either $Z = T$ or $Z = U$. The same argument can be made for $\{y_2, y_1, z_2\}$.

Case 2.5: If U1 $= \{x_2, y_1, z_2\}$, U3 $= \{x_1, y_2, z_1, t_2, u_1\}$. U1 monomer implies either $X = Y$, which is forbidden, or $Y = Z$. If $Y = Z$, U3 monomer implies either $T = U$, which is forbidden, or $T = X$. $Y = Z$ and $T = X$ implies an existence of a reaction $_ + T \longrightarrow Z$, through the U1 monomer, which is forbidden. The same argument can be made for U1$= \{y_2, x_1, z_2\}$.

Case 2.6 : If U1 $= \{x_2, z_1, z_2\}$, U3 $= \{x_1, y_1, y_2, t_2, u_1\}$. U3 monomer implies either $T = U$, which is forbidden, or $T = X$. If $T = X$, U1 monomer implies the existence of a reaction either $_ + T \longrightarrow Z$ or $_ + Z \longrightarrow T$, which are both forbidden. The same argument can be made for $\{y_2, z_1, z_2\}, \{x_2, u_1, z_2\}, \{y_2, u_1, z_2\}$.

Case 3: S2 $= \{x_2^*, y_1^*, y_2^*, z_1^*\}$ and S2 $= \{t_2^*, u_1^*, u_2^*, v_1^*\}$ The reactions that can include these monomers as their module monomers have the form $X + Y \longrightarrow Z$ and $T + U \longrightarrow V$

Case 3.1: If U1 $= \{x_2, _, y_2\}$, there exists a reaction of the form $_ + X \longrightarrow Y$ or $_ + Y \longrightarrow X$, which are both forbidden. The same argument is valid for U1 $= \{t_2, _, u_2\}$.

Case 3.2: If U1 $= \{x_2, y_1, t_2\}$, U3 $= \{y_2, z_1, u_1, u_2, v_1\}$. U1 monomer implies either $X = Y$, which is forbidden, or $Y = T$. If $Y = T$, U3 monomer implies either $Z = Y$, which is forbidden or $T = Y = V$, which is also forbidden. The same argument is valid for $\{x_2, u_1, t_2\}$.

Case 3.3: If U1 $= \{x_2, z_1, t_2\}$, U3 $= \{y_1, y_2, u_1, u_2, v_1\}$. U1 monomer implies either $X = Z$, which is forbidden or $Z = T$. If $Z = T$, the reaction $T + U \longrightarrow V$ can be written as $Z + U \longrightarrow V$, which is forbidden with the reaction implied by the U3 monomer, $Y + U \longrightarrow V$ or $U + Y \longrightarrow V$, as $Z \neq Y$. The same argument is valid for $\{x_2, v_1, t_2\}$.

Case 3.4: If U1 $= \{y_2, y_1, u_2\}$, U3 $= \{x_2, z_1, t_2, u_1, v_1\}$. U3 monomer implies that $Z = T$ as $T = U$ and $T = V$ are forbidden. If $X = V$, U3 monomer implies the existence of a reaction $T + V \longrightarrow U$ or $V + T \longrightarrow U$ which are both forbidden. If $X = U$, U1 monomer implies the existence of a reaction, $_ + X \longrightarrow Y$, which is forbidden. The same argument can be made for $\{y_2, u_1, u_2\}$.

Case 3.5: If U1 $= \{y_2, z_1, u_2\}$, U3 $= \{x_2, y_1, t_2, u_1, v_1\}$. U1 monomer implies that either $Y = Z$, which is forbidden, or $Z = U$. U3 monomer implies that $Y = T$ as $T = U$ and $T = V$ are forbidden. $Z = U$ and $Y = T$ imply the existence of a reaction $X + T \longrightarrow U$, which is forbidden. The same argument can be made for $\{y_2, v_1, u_2\}, \{x_2, y_1, u_2\}, \{t_2, u_1, y_2\}$.

Case 3.6: If U1 $= \{x_2, z_1, u_2\}$, U3 $= \{y_1, y_2, t_2, u_1, v_1\}$. U3 monomer implies either $T = U$ or $T = V$, which are both forbidden. The same argument can be made for $\{t_2, v_1, y_2\}$.

Case 3.7: If U1 $= \{x_2, v_1, u_2\}$, U3 $= \{y_1, y_2, z_1, u_1, t_2\}$. U3 monomer implies that either $U = T$, which is forbidden, or $Z = T$. If $Z = T$, U3 monomer implies the existence of a reaction $T + Y \longrightarrow U$ or $Y + T \longrightarrow U$, which are both forbidden. The same argument can be made for $\{t_2, z_1, y_2\}, \{x_2, u_1, u_2\}, \{t_2, y_1, y_2\}$. ◀

Proof of Lemma 10. The monomer $\{a_2^*, b_1^*, b_2^*, c_1^*\}$ is an S2 monomer. By Lemma 9, the monomer is in a polymer either with a U3 monomer (with image P_5) or two U1 monomers (with image P_8). For the former, the self-saturation immediately implies that U3 monomer includes binding sites a_2, b_1, b_2, and c_1. Then U3 monomer is necessarily $\{a_1, a_2, b_1, b_2, c_1\}$.

For the latter, we will argue by the partition, of cardinality 2, of starred sites of the monomer $\{a_2^*, b_1^*, b_2^*, c_1^*\}$ that two unstar monomers can use to form bonds. First, if the partition includes a single 2-starred site, i.e. $\{a_2^*\}$ or $\{b_2^*\}$, there does not exist an unstar monomer, with image U1, that can form bonds with the rest of the starred sites. If the partition includes $\{b_1^*\}$, the polymer necessarily includes the unstar monomer $\{a_2, b_2, c_1\}$ which implies $A = C$ or $B = C$, both of which are forbidden. Similarly, if the partition

includes $\{c_1^*\}$, the polymer includes the unstar monomer $\{a_2, b_1, b_2\}$, which would imply the existence of a reaction of the form $_ + A \longrightarrow B$ which is forbidden. Hence, the partition cannot be made up of subsets of one-element and three-elements of the starred binding sites of $\{a_2^*, b_1^*, b_2^*, c_1^*\}$.

As the largest unstar monomer that is mapped to U1 under f has length 3, the only other case we need to consider is when the partition is composed of two two-element subsets of the starred binding sites. There are 3 such partitions of a set with 4 elements. The partition $\{\{a_2^*, b_2^*\}, \{c_1^*, b_1^*\}\}$ is eliminated as there is no unstar monomer that is mapped to U1 and includes two 1-starred sites. The partition $\{\{a_2^*, c_1^*\}, \{b_1^*, b_2^*\}\}$ is also eliminated as $\{a_2^*, c_1^*\}$ implies the existence of a reaction $A + C \longrightarrow _$, if the unstar monomer is $\{a_2, c_1\}$ or, $_ + A \longrightarrow C$, if the unstar monomer is $\{a_2, c_1, c_2\}$, which are both forbidden. The only other case is $\{\{a_2^*, b_1^*\}, \{b_2^*, c_1^*\}\}$. If one of the unstar monomers is of the form $\{a_2, b_1, b_2\}$, there exists a reaction of the form $_ + A \longrightarrow B$ which is forbidden. Hence, one of the unstar monomers that forms bonds with $\{a_2^*, b_1^*\}$ is necessarily $\{a_2, b_1\}$. For the other unstar monomer, which forms bonds with $\{b_2^*, c_1^*\}$, if it is $\{b_2, c_1\}$, there exists a reaction $B + C \longrightarrow _$, which is forbidden. Hence, the second unstar monomer is necessarily $\{b_2, c_1, c_2\}$. ◀

Proof of Lemma 11. We will proceed similarly to Lemma 10. The self-saturation implies that in the polymer with image P_5, U3 monomer necessarily includes binding sites a_1, a_2, b_1 and b_2. If the fifth unstarred site is x_1, which is not c_1, there is a reaction of the form $A + B \longrightarrow X$, which is forbidden.

For the polymer with image P_8, the partition of the sites of $\{a_1^*, a_2^*, b_1^*, b_2^*\}$ cannot include a single 2-starred site, i.e. $\{a_2^*\}$ or $\{b_2^*\}$ as there does not exist an unstar monomer, with image U1, that can form bonds with the rest of the starred sites. If the partition includes a single 1-starred site $\{a_1^*\}$ (or $\{b_1^*\}$), the other element of the partition implies that the polymer includes the unstar monomer $\{a_2, b_1, b_2\}$ (or $\{b_2, a_1, a_2\}$), which further implies the existence of a reaction $_ + A \longrightarrow B$ (or $_ + B \longrightarrow A$) which is forbidden. Hence, the partition cannot be composed of one-element and three-element subsets of the starred sites and as in Lemma 10, it is enough to consider partitions into two two-element subsets of the starred sites. The partition $\{\{a_1^*, b_1^*\}, \{a_2^*, b_2^*\}\}$ is eliminated as there is no unstar monomer that is mapped to U1 and includes two 1-starred sites. The partition $\{\{a_1^*, b_2^*\}, \{a_2^*, b_1^*\}\}$ is also eliminated as an unstar monomer including unstarred sites a_1 and b_2 implies the existence of a reaction $B + A \longrightarrow _$, if the monomer is $\{b_2, a_1\}$, or $_ + B \longrightarrow A$, if the monomer is $\{b_2, a_1, a_2\}$, which are forbidden. Hence, only the partition $\{\{a_1^*, a_2^*\}, \{b_1^*, b_2^*\}\}$ is left and the result follows. ◀