Molecular Machines from Topological Linkages

Keenan Breik
The University of Texas at Austin, TX, USA

Austin Luchsinger
The University of Texas at Austin, TX, USA

David Soloveichik
The University of Texas at Austin, TX, USA

Abstract

Life is built upon amazingly sophisticated molecular machines whose behavior combines mechanical and chemical action. Engineering of similarly complex nanoscale devices from first principles remains an as yet unrealized goal of bioengineering. In this paper we formalize a simple model of mechanical motion (mechanical linkages) combined with chemical bonding. The model has a natural implementation using DNA with double-stranded rigid links, and single-stranded flexible joints and binding sites. Surprisingly, we show that much of the complex behavior is preserved in an idealized topological model which considers solely the graph connectivity of the linkages. We show a number of artifacts including Boolean logic, catalysts, a fueled motor, and chemo-mechanical coupling, all of which can be understood and reasoned about in the topological model. The variety of achieved behaviors supports the use of topological chemical linkages in understanding and engineering complex molecular behaviors.

2012 ACM Subject Classification Theory of computation → Models of computation; Theory of computation → Computational geometry

Keywords and phrases chemical computation, mechanical computation, bioengineering, models of biochemistry, molecular machines, mechanical linkages, generic rigidity

Digital Object Identifier 10.4230/LIPIcs.DNA.27.7

Funding This work was supported by NSF grant CCF-1901025.

Acknowledgements We thank Tosan Omabegho for introducing us to chemical linkages and for extensive discussions.

1 Introduction

Living cells, by far, show the most complex chemical behavior known. Their primary functional parts are molecular machines that derive their behavior from internal mechanical motion [2]. There are large gaps in our understanding of how function originates from mechanical reconfiguration, and engineering such machines from scratch stands as the grand challenge of bioengineering.

Even simple chemistry involves intricate physics, which makes it a challenge to model molecular machines. But we may not need that intricacy to capture their rich behavior. It would be interesting if the essence of their behavior could be reproduced by a simple mechanical model. But how simple can the model be, what features should it include, and how should its parts be assembled?

We look to linkages, a simple tool from mechanical engineering. A linkage is a set of rigid rods (called links) connected at rotary joints. This well-studied tool has been shown to be capable of very complex motion, such as tracing any arbitrary curve [9, 8]. Reference [4] provides an excellent overview of the capabilities of linkages. Linkages have also found use in studying biological mechanisms. In reference [14], linkages are used to model the mechanical behavior of proteins. Most closely related to this paper, however, is Omabegho's work

introducing chemical linkages where joints can form chemical bonds [12]. There, chemical linkages are used to model the role of allostery in enzyme behavior. Motivated by the above work, we reimagine chemical linkages as a basic model for rich chemical behavior, and explore the variety of behaviors that can be captured by the model.

Linkages seem like a minimal mechanical model of biochemistry, but we show how to simplify them further and still produce complex behavior. The link lengths in a linkage determine how the linkage can move. This in turn controls which joints can chemically bind. Surprisingly, in many cases which joints can bind could be fully determined by just the topology of the underlying graph. Focusing on simple graphs makes systems easier to design and analyze.

On top of having interesting behavior theoretically, chemical linkages could also lead to real molecular machines, like artificial enzymes. Their rich behavior comes from simple parts: links, joints, binding sites. These may be possible to build directly. Double-stranded DNA may be rigid enough to act as a link. Single-stranded DNA might implement a joint as a so-called compliant mechanism. Orthogonal DNA sequences could act as binding sites. However, this work does not further explore practical implementation.

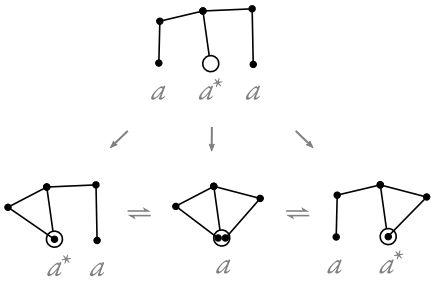
The contributions of this work are as follows. We formalize chemical linkages, previously described only informally in [12]. We also introduce a new topological model, which bases bond formation on simple graph topology. To articulate our new model, we expand on traditional characterizations of graph rigidity given by Laman [10] and Henneberg [5]. We design artifacts (which work in both models) showing that surprisingly complex behavior can be developed from first principles. The constructions include Boolean logic and signal propagation, catalytic splitting reminiscent of ATP hydrolysis, fueled directed cycles, and chemo-mechanical coupling. The latter constructions are motivated by the coupling of fuel consumption with other processes prevalent in biological molecular machines such as kinesin, myosin, and dynein [1, 3].

2 Examples

Mechanical linkages are well studied and common in mechanical engineering. Even one of the most basic linkages, the lever, is found as a component in countless machines and tools. Linkages have also been shown to be capable of very complex motion [11]. Adding binding sites makes linkages interesting as a model of chemical machines. This section shows how chemical linkages work by example, while Section 3 defines them formally.

In this work we consider a single-copy regime where, unless otherwise stated, a single copy of each linkage is present in a given system (see Conclusion for additional discussion).

▶ **Example 1** (Binding sites)**.** As the following example shows, joints can have binding sites. The star $*$ means a (solid dot) and a^* (hollow circle) are complements and so can bind. For their sites to bind, joints have to overlap. When an a and a^* overlap, we may label them together with just a. Unless noted otherwise, we consider strong bonds which may not break. (Appendix A discusses the approximation of strong bonds using bonds that may break.) Although bonds are strong, joints with the same binding sites may displace one another. So to get from the left bound state to the right bound state, the right a^* joint must colocalize and displace the left a^* joint. The four possible states of the system are the following.



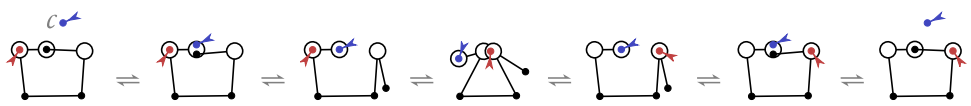
We consider a state to change only when the set of bonds changes. Physically, the shape of a linkage can move among an infinite continuum of conformations. But it would move rapidly and randomly among the conformations allowed by its bonds as Brownian motion and low Reynolds number dominate molecular dynamics. This is why we say this example has four discrete states and not an infinite continuum of states.

▶ **Example 2** (Allostery)**.** The following example shows that geometry can prevent complements from binding. There are two bonds that can form, the a bond and the b bond. But after one forms, the other can no longer reach to overlap.

Note that links are allowed to cross over each other. Allowing link crossing simplifies the mathematical model and is standard in the analysis of mechanical linkages. In a physical realization, the links might be offset at different heights to allow such crossing.

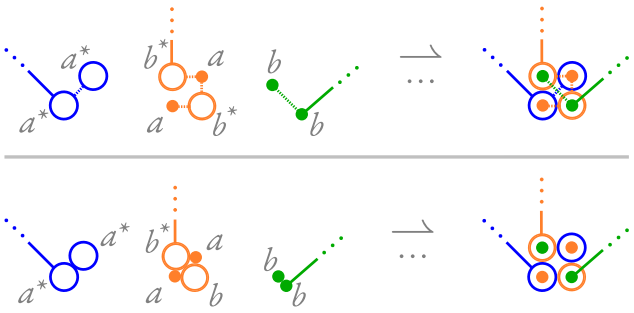
▶ **Example 3** (Simple catalysis)**.** The following example is a system where there are two states, left and right, that cannot reach each other. For each state, two of its (infinitely many) conformations are shown with $\approx$ between them. The link lengths keep the small red linkage from getting close enough to the opposite joint to displace onto it. (A small flag represents irrelevant omitted parts, so the red linkage shown could be a portion of a larger linkage that includes the joint with the flag.) So the red linkage stays bound to whichever joint it starts bound to.

But we can add a linkage that allows the states to reach each other. In the following example, consider adding the blue linkage at the top to mediate the state change. The blue linkage with binding site c^* can enter and displace the center bond. This allows the red linkage to reach the opposite joint. Afterward, the blue linkage is displaced by the center bond and leaves. Thus the blue linkage is unchanged, but the black linkage has changed state. The blue linkage acts as a catalyst.



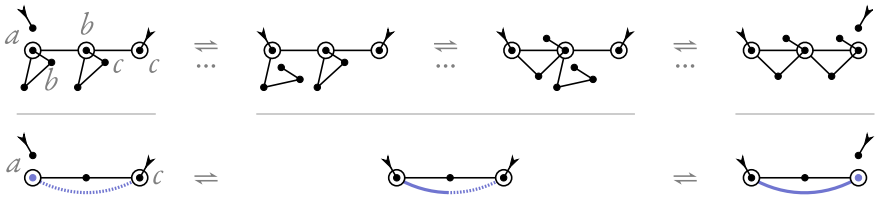
▶ **Example 4** (Compound catalysis). The following is a catalytic system that involves three catalysts, or one compound catalyst, depending on the reader's perspective. The linkage can again be in one of two states unable to reach each other. But when all three of the catalytic linkages are present, the state can change. The intermediate displacement states are left out.

▶ **Example 5** (Many binding sites). While each joint may only have one binding site, the effect of multiple binding sites per joint may be achieved via zero-length edges. Below, zero-length edges between joints are indicated by dashed lines. We omit these edges for visual clarity but instead rely on color to disambiguate which binding sites are connected together with zero-length edges. (Although some information is lost going from the top to bottom figure, the bottom notation will be sufficient for our purposes.)

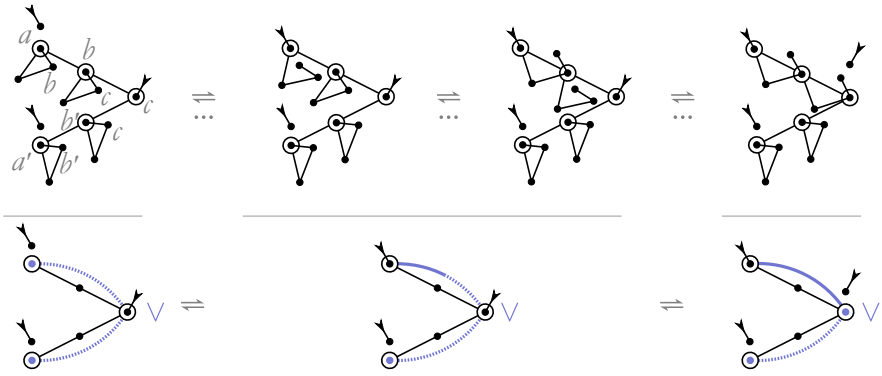


This effectively allows a joint to form multiple bonds at the same time.

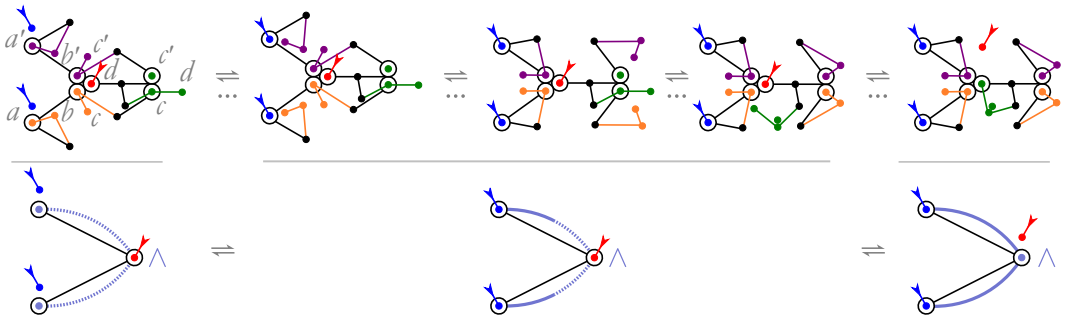
▶ **Example 6** (Signal cascade). The following is a system that transmits the signal of whether a bond is formed. The signal travels along a sequence of links. The effect is that the flag with binding site a and the flag with binding site c can never be free at the same time. By repeating this pattern, we can get this effect across any number of links. Since signal cascading is useful as a modular gadget, we abstract it with visual notation using blue arcs as shown in the bottom half of the figure.



▶ **Example 7** (Advanced cascades). The following systems show how cascades can be combined for various effects. By combining two signal cascades with a common joint, signal cascades may behave like a logical OR. The rightmost flag may be freed after either of the leftmost flags have bonded.



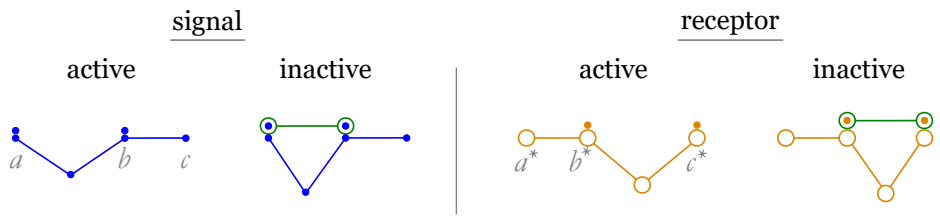
A logical AND may be achieved with signal cascades as well. Both of the leftmost flags (a and a') must bond to the linkage before the d flag can be freed. Using the notation introduced in Example 5, directly adjacent binding sites indicate a single joint with multiple binding sites.



By reversing the AND mechanism shown above, we can effectively implement a fanout.

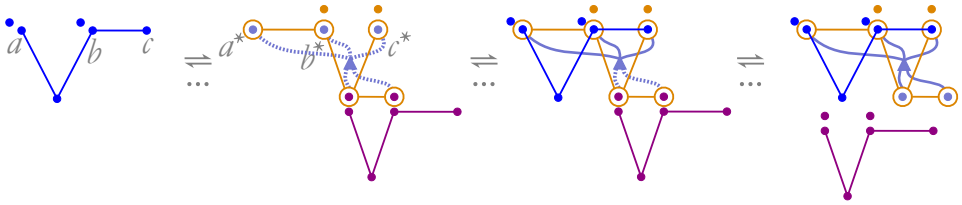
▶ **Example 8** (Active/inactive signal receptors and sequential AND gate)**.** We now show a construction for intermolecular signals and their corresponding receptors, with both capable of activation and inactivation by other signals. In particular, we show an example which mimics the sequential AND gates of reference [15] operating via DNA strand displacement.

In our construction, whether signals and their receptors are active depends on whether their three binding sites can be simultaneously bound or not (shown below). We later refer to such complete binding as *docking*. Docking is prevented by geometrical constraints if either the signal (blue) or receptor (orange) is inactive. Intuitively, the distance between the joints within signals and receptors needs to match exactly for them to dock. But this cannot happen when either the signal or receptor has a joint distance that is fixed by another linkage. In the figure, the green linkage fixes the distance between the a and b joints (or b^* and c^* for the receptor) to a length that is too short. So only active signals and receptors may dock.

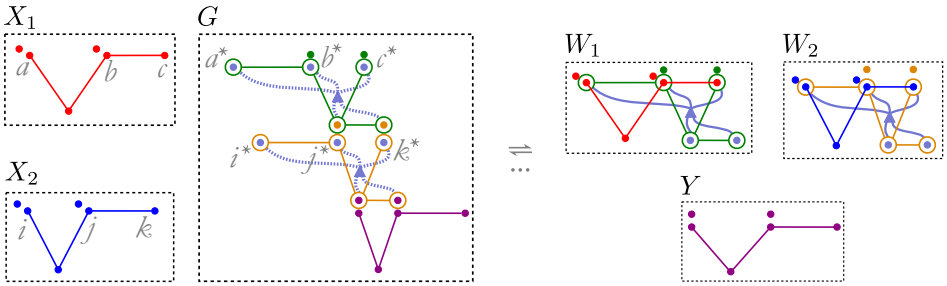


The following figure shows a modification of the orange receptor which holds the output (purple) signal linkage inactive until the input (blue) docks with it. Here, the blue triangular symbol represents a cascade combination of an AND and a fanout. All three signal cascades

entering the top of the triangle must fire before the two signals on bottom can fan out. This gadget ensures that all three bonds a, b, and c must form between the signal and receptor before the purple linkage is released. Notice that the output signal linkage can be an input to another receptor downstream, and so such systems are composable.



Modifying this scheme further, the following example shows a composable sequential AND gate.



Written in terms of abstract chemical reactions, the above system implements the following behavior:

$$X_1 + G \rightleftharpoons W_1 + H$$
$$X_2 + H \rightleftharpoons W_2 + Y$$

where X_1 and X_2 are the input signal linkages and Y is the output signal linkage. Observe that the purple output signal linkage and the orange receptor are initially both in their inactive states. The red input linkage must dock with the green receptor first in order to activate the orange receptor (splitting complex G in the process). Then the blue input signal linkage docks with the now active orange receptor, displacing and activating the purple output signal linkage. Appendix B shows the state-change details for this construction.

Note that DNA strand displacement cascades are based upon the toehold sequestering mechanism which allows activating and inactivating both signals and their displacement targets. Targets (receptors) are activated by opening their toehold domains, while signals are activated by opening their toehold binding domains. The above construction shows one way in which signal and receptor activation can be recapitulated by chemical linkages.[1]

[1] In the toehold sequestering mechanism, activation is a kinetic effect: toehold binding increases the effective local concentration of the signal strand near the target, which promotes displacement. Displacement can still occur without preceding toehold binding, but is much slower. Our chemical linkages implementation does not attempt to capture such physical details of toehold sequestering but rather the higher-level activation/inactivation behavior. The physical correlate for activation in our model is geometric compatibility rather than an increase in local concentration. (Although beyond the scope of this paper, a kinetic model of chemical linkages, for example operating via Gillespie kinetics, with weak bonds representing toeholds (see Appendix A) is needed to capture the kinetic mechanism of toehold sequestering.)

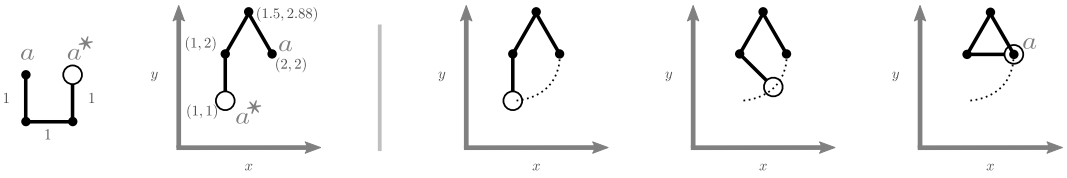


Figure 1 (Left) A chemical linkage and a conformation p of that linkage. (Right) A motion q of conformation p. Left to right: conformation $q(0)$ of q, conformation $q(t)$ where $0 < t < 1$, and conformation $q(1)$. Notice q is a binding motion that forms a new bond.

3 Formal model

The previous section relied on intuitive explanations of chemical linkages, as does prior work [12]. It would be useful to have a precise, general definition. This would support engineering, guide simulations, and enable proofs. This section formally defines chemical linkage systems and their state space.

3.1 Chemical linkages

A *mechanical linkage* is a pair (G, ℓ). G is a connected graph with vertices V and edges E. We also call vertices *joints* and edges *links*. $\ell : E \to \mathbb{R}_{\geq 0}$ is a map that gives each link a length. Link lengths alone are not enough to define how the graph sits in space. So to uniquely determine the shape of a linkage, we use conformations.

A *conformation* of a linkage (G, ℓ) is a map $p : V \to \mathbb{R}^2$ where $|p_u - p_v| = \ell(u, v)$ for each pair u and v of linked joints.[2] Intuitively, a conformation is a drawing of a linkage in the plane with the right link lengths. In some drawings, joints may overlap. The *overlap* of a conformation is the partition of its joints where each element is a set of joints that overlap. For example, the partition $\{\{u, v\}, \{w\}, \{x\}\}$ of $\{u, v, w, x\}$ would mean that joints u and v overlap while no joint overlaps w and no joint overlaps x.

A *chemical linkage* is a mechanical linkage with a function $d : V \to \Sigma$ that puts a binding domain on each joint. Σ is an alphabet of starred and unstarred symbols called *domains*. Domains x and x^* are said to be complementary and thus capable of binding. Intuitively, this evokes a chemistry where opposites bind like complementary DNA domains.[3] Note that we use the phrase "multiple binding sites" to refer to zero-length edges effectively allowing a joint to have more than one binding domain (see Example 5). In this paper we use the word linkage to refer to mechanical or chemical linkage if clear from context.

A *matching* of a conformation is a set of unordered pairs of its joints such that (1) each pair consists of overlapping joints which have complementary domains and (2) no two pairs share a common joint. A matching of a conformation is a *binding* if it is not a subset of any other matching of that conformation (i.e. it is a maximal matching). Intuitively, elements of the binding represent a bond between two joints. Since a binding is a *maximal* matching, we consider bonds to form as soon as joints become overlapping. In the case where three or more joints overlap, a conformation could have multiple bindings (for example consider the lower middle state in Figure 3).

[2] For simplicity, we focus on two dimensions in this work. It is an open question how some of this work generalizes to three dimensions (see also Conclusion).

[3] Other choices of domain chemistry are of course possible, where binding might be like-like. Such binding rules are not explored in this work.

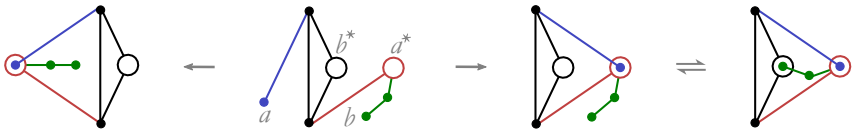


Figure 2 This figure illustrates why bindings alone do not sufficiently capture the behavior of chemical linkages. Middle left: a state where bonds a and b can form. If a bonds first, it can lead to two different states. In one state (left), the binding of a prevents the binding of b due to link lengths. In the other state (middle right), the green link can still reach a state where b is bonded (right). Notice the left and middle right conformations have the same binding, but the conformation geometry dictates which states can be reached.

The natural notion of motion captures how transformations may be applied to conformations. Let $[0, 1]$ be the interval of real numbers from 0 to 1. A *reconfiguration* of a conformation p is a map $q : V \to ([0, 1] \to \mathbb{R}^2)$ where each $q(t)$ is a conformation and $q(0) = p$. Note that for convenience, $q(t) = u \mapsto q_u(t)$ is the conformation at time t, and $q_u = t \mapsto q_u(t)$ is the trajectory of joint u. A *motion* of a conformation p is a reconfiguration q where each q_u is continuous. Intuitively, this means a motion preserves link lengths and never flips parts of a linkage: for example in Figure 2, transitioning from the left state to the middle right state, without breaking bonds, is not a motion.

A motion q is a *step motion* if there exists a same binding of conformation $q(t)$ for all $t \in [0, 1)$. A step motion is a *binding motion* if conformation $q(0)$ has some binding that is a subset of a binding of conformation $q(1)$. Intuitively, a step motion maintains the overlap of bound joints and a binding motion only ever (potentially) creates more bonds. For conformations x and y, we write $x \to y$ if there exists a binding motion from x to y. We write $\to^*$ to mean the reflexive, transitive closure of $\to$. We write $x \leftrightarrow y$ if $x \to y$ and $y \to x$. Similarly, we write $x \leftrightarrow^* y$ if $x \to^* y$ and $y \to^* x$.

Although this work only uses binding motions, we can also define motions which are allowed to reduce the number of bonds – for completeness. A step motion is a *breaking motion* if conformation $q(0)$ has a binding that is a superset of a binding of conformation $q(1)$.

3.2 Chemical linkage states and systems

Binding motions fully capture the behavior of chemical linkages. But most binding motions do not lead to interesting changes in a given conformation. Such changes arise when a conformation's binding is altered by the binding motion (e.g., see Figure 2). In a sense, some conformations are equivalent, while others are not. We use the notion of states to capture conformation equivalence and identify the significant binding motions.

A state is an equivalence class of conformations defined as follows. Two conformations are in the same state if they can both reach all of the same conformations through a binding motion. Formally, two conformations x and y are in the same state if for every other conformation z, $x \to z$ if and only if $y \to z$. State b is *directly reachable* from a, written $a \rightharpoonup b$, if there exist conformations x in state a and y in state b such that $x \to y$. If $a \rightharpoonup b$ and $b \rightharpoonup a$, we write $a \rightleftharpoons b$. We say state b is *reachable* from a, written $a \underset{\cdots}{\rightharpoonup} b$, if $x \to^* y$. If $a \underset{\cdots}{\rightharpoonup} b$ and $b \underset{\cdots}{\rightharpoonup} a$, we write $a \underset{\cdots}{\rightleftharpoons} b$. Figure 3 (bottom) shows an example of state reachability.

We can treat a set of linkages exactly like a single linkage by forming the disjoint union of its linkages. Intuitively, we pretend all its linkages are one big linkage. This way we can use all the vocabulary of linkages for a set of linkages. We refer to sets of chemical linkages as a system. Formally, a *chemical linkage system* is a pair (C, s) where C is a set of

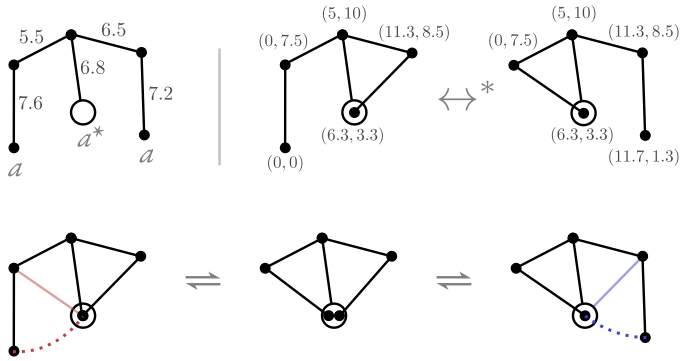


Figure 3 (Top) A chemical linkage c and two conformations of c which may be transformed into one another via a sequence of binding motions. (Bottom) Three distinct states of c. In the middle state, both of the two a binding sites and the a^* binding site overlap. The left and middle states are directly reachable from one another via the binding motion shown in red. Likewise, the middle and right states are directly reachable from one another via the blue binding motion. The left and right states are reachable from one another, but not directly reachable (since the binding changes in the middle state).

chemical linkages and s is an initial state. As Figure 2 shows, the initial state is important for determining state reachability for a given chemical linkage system. In a conformation of a linkage system, we often call a set of linkages a *complex* if there exists a binding which makes that set of linkages connected.

3.3 Complexity of simulation

With the formal model established, we now have a notion of how to describe the behavior of chemical linkages. To understand and predict this behavior, we need to know a given linkage's state space. Computing the entire state space for a system is certainly a hard problem. But being able to easily check if one state is directly (one step) reachable from another would make it easier to design and analyze chemical linkage systems.

Unfortunately, the problem of deciding direct reachability between two states is PSPACE-hard because the subproblem of finding a motion between conformations is PSPACE-hard [6][4]. This does not bode well for the future development of simulation tools for this model. For such simulation tools, we would want to have a fast algorithm to check direct reachability. This is part of our motivation for introducing the topological linkages model described in the next section.

4 Topological linkages

Linkages provide a model of physical constraints that seems minimal. They involve only two simple physical parts, links and joints, neither of which can be removed. Despite that intuition, this section describes a surprising new simpler model that still has interesting complex behavior. In fact, our simpler model captures the behavior of all of the examples in Section 2.

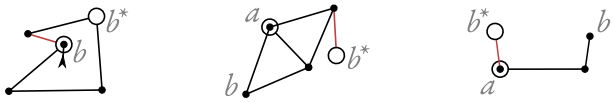
[4] While the general mover's problem (reachability) was shown to be PSPACE-hard in [13], we can more easily adopt the formulation of [6] by fixing some joints' relative positions to each other via a rigid "frame." A study of motion planning can be found in [11].

To distinguish the two models, we call the original metric and the simpler topological.

4.1 Topological motivation

In this section we build the case for a version of the chemical linkages model which ignores link lengths. The utility of such a model may seem surprising given that the constructions developed in previous sections relied on some joint being able or not being able to "reach" to another joint in metric space. Rather than fixing link lengths beforehand and asking whether a joint can reach another, we instead ask how *constrained* must the link lengths be for such reach to be possible.

Recall Examples 2 and 3 from Section 2 and how certain states were not reachable due to the choices of link lengths. We present these two examples again below, along with a third example.



In each example, we consider the potential bonding of the b and b^* joints. Without link lengths being previously fixed, observe the link length constraints implied for each edge. Once the lengths of the black links are fixed, we ask how constrained the red link lengths are.

In the right example, a range of lengths for the red link will allow bond b to form (as long as the link has sufficient length). On the other hand, in the two examples on the left, bond b may be formed only if the red links happen to be the *exact* right length. Assuming link lengths are somehow "generically" chosen, bond b will not form in the left two examples because the red link lengths will not reach. Note that from a practical perspective, avoiding such exact-length coincidences is easy. Contrarily, setting lengths exactly in a molecular implementation may be onerous.

Our key observation is that this exact-length constraint arises in these examples because the forming of bond b would cause the red link to connect two joints which are already rigidly connected (we refer to this later as overbracing). Note that we say two joints are rigidly connected if their distance is fixed for every motion. In the left example it is clear that the top-most link already fixes the distance of its two joints. In the middle example the rigidity of the linked rhombus fixes the distance of the bottom-left and top-right joints.

Each of the constructions presented in Section 2 exhibit the behavior of these left two examples. We will define the topological linkage model with regard to overbracedness, rather than metric lengths. To formally define overbracing edges we must first discuss the notion of rigidity.

4.2 Rigidity

Intuitively, a graph is rigid if all of its joints have to move together. A rigid graph is minimally rigid if removing any edge results in a non-rigid graph. These properties naturally generalize to subgraphs as well.

We provide a definition of graph rigidity based on the characterization captured by Henneberg operations [5, 4]. Henneberg characterized minimally rigid graphs as the graphs which can be constructed, starting from a single edge, by executing some sequence of two types of operations. A V operation adds a new node u and two new edges that connect u to two existing nodes in the graph. A T operation adds a new node u that splits an existing edge and adds a new edge from u to an existing node. Figure 4 shows an example.

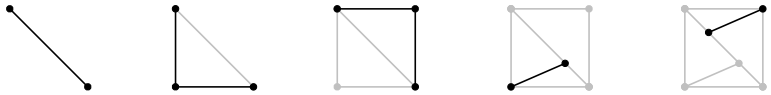


Figure 4 A sequence of Henneberg operations. An edge to start, then two V operations, then two T operations. The graph produced at each step is minimally rigid.

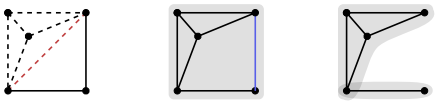


Figure 5 An overbraced graph (left) and its overbracing edges (dashed). If we remove an overbracing edge (red), this graph happens to become minimally rigid (center). We can verify this with Henneberg operations. If we remove a non-overbracing edge (blue), the graph is no longer minimally rigid (right). We can verify this with Laman's theorem.

So we say a graph G is *minimally rigid* if some sequence of Henneberg operations turns a single edge into G. We build upon this definition to formalize overbracing discussed previously. A graph G is *overbraced* if it has a subgraph H that is minimally rigid and G contains an additional edge between two nodes of H. This edge is called an *overbracing edge*. Figure 5 shows an example.

Naturally, these Henneberg operations are useful in verifying if a given graph is minimally rigid. What they do not do, however, is provide an easy method for showing a graph is not minimally rigid.

Luckily, when a graph is not rigid, Laman's theorem [10] guarantees there is a simple proof. To state this powerful theorem, say the *excess* of a graph with v vertices and e edges is $e - (2v - 3)$. Recall that for a graph G, an induced subgraph I is a graph formed from a subset of vertices from G taking all edges connecting pairs of vertices in that subset.

▶ **Theorem 9** (Laman's theorem). *A graph G is minimally rigid iff (1) the excess of G is 0, and (2) the excess of any induced subgraph of G is at most 0.*

Laman's theorem lends itself to proving that a graph is not minimally rigid. We just show that the graph has non-zero excess, or we give an induced subgraph with positive excess.

We now want to transition from checking minimal rigidity to checking overbracedness, which is the real object of our attention since we use overbracedness as a substitute for "being unable to reach". First, the following corollary formally confirms our intuition that a minimally rigid graph cannot be overbraced. Via this corollary, we can use Henneberg operations or Laman's theorem to show that a graph is not overbraced.

▶ **Corollary 10.** *If a graph G is overbraced, then G is not minimally rigid.*

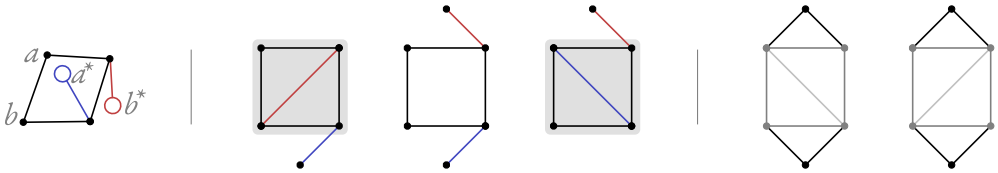
Proof. Let G be an overbraced graph. By definition of overbraced, G has some minimally rigid subgraph $H = (V, E)$ and some additional edge e between two nodes of H. By Theorem 9, the excess of H is 0. Let $I = (V, E')$ be the induced subgraph of G on nodes V. Since $e \in E'$ and $e \notin E$, the excess of I must be greater than 0. So, G is not minimally rigid. ◀

Our constructions may not be minimally rigid as a whole (may have non-rigid parts) as in Figure 5 (right). In this case, to argue that a graph is not overbraced we rely on the following Lemma.

▶ **Lemma 11.** *If a graph G is a subgraph of some minimally rigid graph M, then G is not overbraced.*

Proof. We prove this by contrapositive. Let G be an overbraced graph. By definition of overbraced, if we add edges and/or vertices to an overbraced graph it will still be overbraced. This means that any graph M which has G as a subgraph must also be overbraced. By Corollary 10, M is not minimally rigid. ◀

For the examples in Section 4.1, we relied on mechanical intuition to see that the endpoints of the red edge are rigidly connected. We are now ready to show how formal rigidity arguments can be used to show this.



The figure above shows the linkage from Example 2 (left) and its three states (middle). We can prove that none of the three states are overbraced using Henneberg operations and Lemma 11. Two minimally rigid graphs are shown (right), constructed via Henneberg operations. The edge shading shows the order of the operations. This shows that each of the three states is a subgraph of a minimally rigid graph. Thus, by Lemma 11, they are not overbraced. However the following would-be state, not shown in the example, is prevented not just by reach, but by mere rigidity.

This figure shows that the state (left) has a minimally rigid subgraph H and an edge between two vertices in H (right). By definition, this state is overbraced.

The figure below shows similar analysis of the prevented state from Example 3.

Example 3 (left) and its prevented state (right). The state on the right is shown to be overbraced, as it has a minimally rigid subgraph (solid black edge) and an edge connecting two vertices from that subgraph (dashed edge). In fact, any graph with multiple edges between vertices is overbraced by our definition. However, in the presence of the catalyst no state in the transition sequence is overbraced. The reader can verify this again using Henneberg operations. In fact, this kind of analysis may be applied to *all* the examples in Section 2, verifying that the topological behavior of each follows the metric behavior.

Finally, we touch upon the computational complexity of checking overbracedness. Recall that one of the motivations for the topological linkages model is that checking direct (one-step) reachability in our original metric model was PSPACE-hard. Can we improve upon this with the simplification of the topological model? Laman's Theorem does not suggest a fast algorithm since there might be exponentially many induced subgraphs. However, a well-known algorithm called the pebble game does just that [7]. The algorithm runs in quadratic time and reports whether the given graph is rigid, what its rigid components are if it is not, and where it is overbraced.

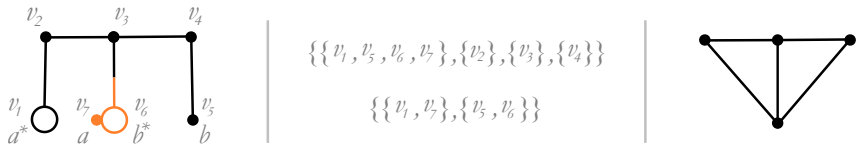


Figure 6 A topological linkage (left). Note that zero-length edges are denoted by adjacent same-colored joints. An overlap and a binding for the given linkage (middle). The collapse of the given overlap (right). Notice that the collapse is not overbraced, so it is a state.

4.3 Formal model

We define a kind of chemical linkage that completely ignores lengths. Figure 6 illustrates these definitions. A *topological linkage* is a triple (G, d, ℓ). G is a connected graph with vertices V and edges E. $d : V \to \Sigma$ is a map that puts a binding domain on each joint and $\ell : E \to \{0, +\}$ is a map that labels each edge as a zero-length edge or positive-length edge. Recall that the use of zero-length edges in the metric model effectively describes multiple binding sites on one joint. We distinguish between zero-length and positive-length edges for the same effect here. A topological linkage has no other specified lengths or geometry. Its only structure comes from the topology of its graph.

We define the state space of a topological linkage with no appeal to motion or conformations. Instead, we use a partition of the topological linkage's joints that represents which joints are meant to overlap. An *overlap* of a topological linkage is a partition of its joints such that (1) no two joints are in the same partition part if they are connected by a positive-length edge and (2) any two joints connected by a zero-length edge are in the same partition part. Intuitively, joints connected by a zero-length edge already overlap, while joints connected by a positive-length edge cannot overlap. A *matching* of an overlap is a set of unordered pairs of its joints such that (1) each pair consists of joints from the same part of the overlap and which have complementary domains and (2) no two pairs share a common joint. A matching of an overlap is a *binding* if it is not a subset of any other matching of that overlap. Note that an overlap may have multiple bindings. The *collapse* of a topological linkage relative to an overlap is the graph that results from the following operations: (1) remove all zero-length edges, and (2) perform vertex contraction on the vertices in each part of the overlap.[5] A *state* is an overlap whose collapse is not overbraced.[6]

Similar to metric linkages, we define a notion of reachability for topological linkage states. State b is *directly reachable* from a, written $a \rightharpoonup b$, if a has a binding that is a subset of a binding of b. If $a \rightharpoonup b$ and $b \rightharpoonup a$, we write $a \rightleftharpoons b$. We define $\underset{\cdots}{\rightharpoonup}$ to be the reflexive, transitive closure of $\rightharpoonup$ and say state b is *reachable* from a if $a \underset{\cdots}{\rightharpoonup} b$. If $a \underset{\cdots}{\rightharpoonup} b$ and $b \underset{\cdots}{\rightharpoonup} a$, we write $a \underset{\cdots}{\rightleftharpoons} b$. Also similar to metric linkages, we define a *topological linkage system* as a pair (T, s), where T is a set of topological linkages, and s is an initial state for that set of linkages.

[5] We consider vertex contraction that may lead to a multigraph in cases where two vertices to be contracted, v_i and v_j, are both adjacent to some other vertex w.

[6] Another reasonable approach may be to allow a system's initial state to be overbraced but to disallow forming bonds that lead to additional overbracing. However, without loss of generality overbracing edges could be removed from the initial state without affecting the behavior.

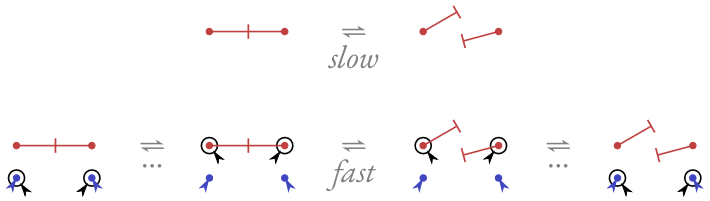
5 Fueled machines

One of the goals of this paper is to recreate some of the rich behavior of molecular machines. Thus far, we have presented constructions in the metric model (Section 2) and have shown that these constructions also work in the topological model (Section 4.2). This section develops additional complex behavior abstracting the ability of biological machines to consume fuel and couple this consumption with driving other processes. These constructions, although explained in the topological model, can also be understood in the metric model.

5.1 Hydrolysis

The molecular machines in living cells are fueled largely by ATP hydrolysis. We can imagine the molecule ATP as composed of two parts, ADP and P_i. For our purposes, we write this as ATP $\rightleftharpoons$ ADP + P_i. The forward reaction is hydrolysis, which splits ATP. Normally, hydrolysis and its reverse are slow, which makes ATP stable in isolation. But if ATP docks with certain catalysts, both directions become fast. To make sure that hydrolysis happens more than its reverse, cells keep the concentration of the wastes ADP and P_i low.

It is not clear how to engineer systems to turn hydrolysis into work. But we can start by figuring out how to do so with linkages. The following linkage system abstracts a hydrolysis-like splitting event. The top red bar plays the role ATP. The two small linkages below it represent the catalyst that docks with it. Once the catalyst docks, the red bar can split.

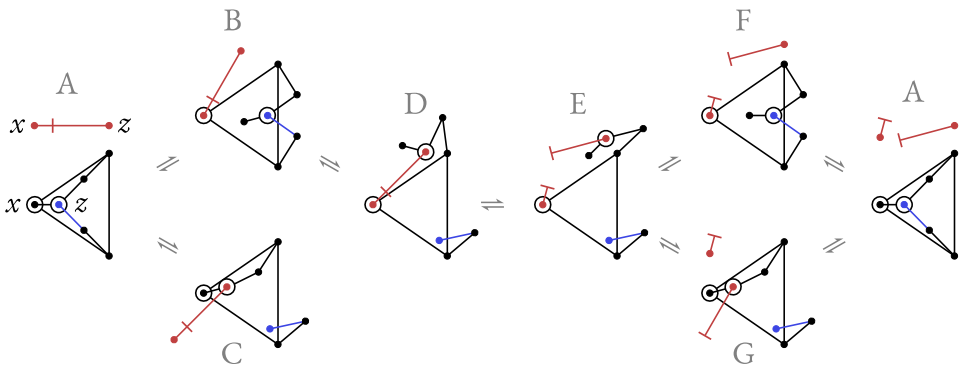


For now, we will take these splitting pseudo-linkages as a primitive, as does prior work [12], and we will focus on a construction that uses it.

5.2 Motor

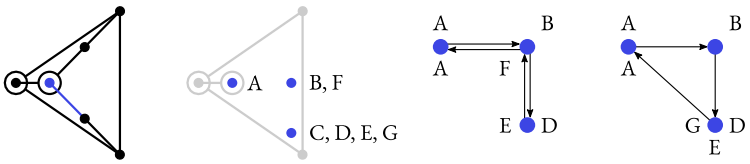
Mechanical work can be coupled to the motion of the catalyst if the catalyst undergoes an overall cyclic motion. A catalyst for binding and splitting ATP is shown below.

While we discuss the example in terms of topological states, we continue to use a visual notation which contains implicit link lengths. In this way our visual representation shows a particular metric implementation which remains compatible with the metric model.



The catalyst is asymmetric in a way that yields the following behavior. If the catalyst first binds ATP on the left (state B above), then it can subsequently bind on the right (state D). However, if the catalyst first binds the ATP on the right (state C), it is prevented from subsequently binding it on the left because that displacement passes through an overbraced state (not shown). After ATP splits into ADP and P_i (state E), the catalyst can unbind in any order (since the two binding sites are now split, no overbracing occurs).

Observe that the catalyst itself is always in one of three distinct states. The motion of the catalyst is determined by the order of detachment for ADP and P_i. Shown below is a depiction of how mechanical work is coupled to the motion of the catalyst from state to state.



If the catalyst unbinds on the right and then on the left, which is in the opposite order of binding, then it undoes any mechanical work done in the process of binding. However, half the time, the catalyst unbinds on the left and then on the right. This results in an overall biased work cycle, capable of driving mechanical work.

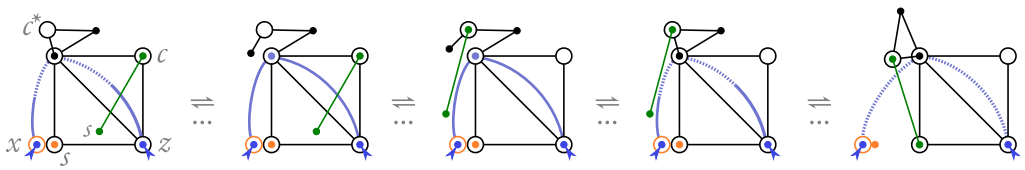
5.3 ATP from linkages

The following construction shows that we do not need to assume ATP hydrolysis as a primitive. Instead, here we present a pure linkage system that behaves from the outside just like the primitive. So we can actually treat the primitive not as an assumption, but as an abstraction.

The two states, whole (left) and split (right), cannot reach each other. Recall that a dashed arc represents the gadget from Example 6. At least one of the binding sites at its endpoints must be bound. The gadget does not physically attach the halves, so the split state is two separate parts, despite a dashed arc appearing to connect them.

To go from whole to split, the long green link bound at the joint marked c would have to relocate to the unbound joint marked c^*. That would have to break a bond, which is not possible. But by adding the following catalyst, the long link can relocate.

When bound to the whole, the catalyst enables the following path between whole and split.

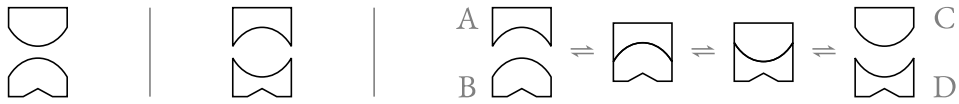


It might seem that the above ATP construction and the motor construction of Section 5.2 require different parities for the ATP-catalyst interaction. In the motor, the ATP displaces joints in the catalyst. In the ATP in this section, the catalyst displaces joints in the ATP. Luckily, the two displacements can be combined as shown in Appendix C.

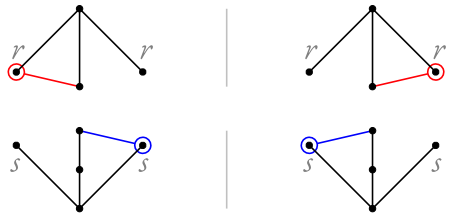
5.4 Chemo-mechanical coupling

Chemical coupling is a powerful tool. By chemical coupling we mean a reaction like $A + B \rightleftharpoons C + D$ with no side reactions like $A \rightleftharpoons C$ or $B \rightleftharpoons D$. Such a coupling allows a high concentration of A to behave as fuel to drive B into D, even when B turning into D is thermodynamically unfavorable.

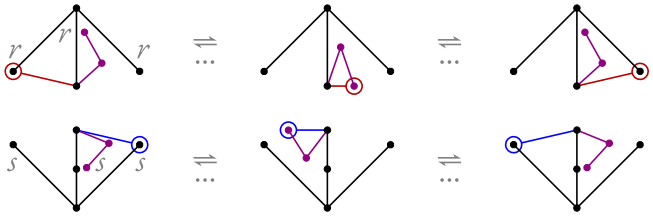
Here we will develop a construction that achieves chemical coupling. The following abstract diagram illustrates the target behavior. When the two linkages meet, they can only dock if their states are complementary. While docked they can switch states as long as they stay complementary. Otherwise, there is no docking and no state change.



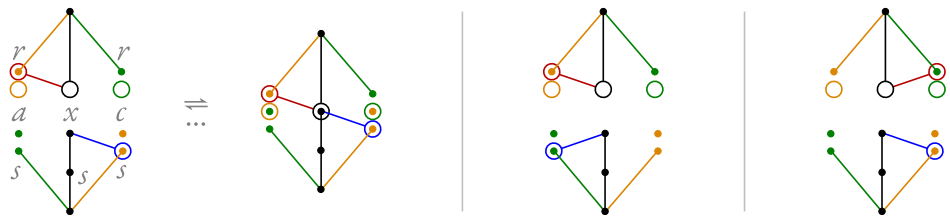
The following are two simple linkages each with two states, left and right. For each linkage, there is a barrier between its two states. The left state cannot reach the right state.



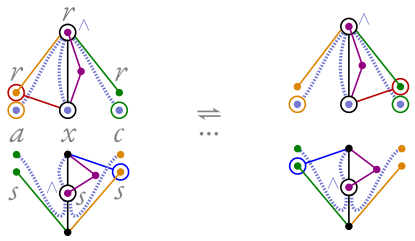
We can remove the barrier if we add an internal catalytic part. The following construction shows an example of this. It adds an arm with two links. The new arm can displace the matching domain of the original arm and carry it to the opposite side.



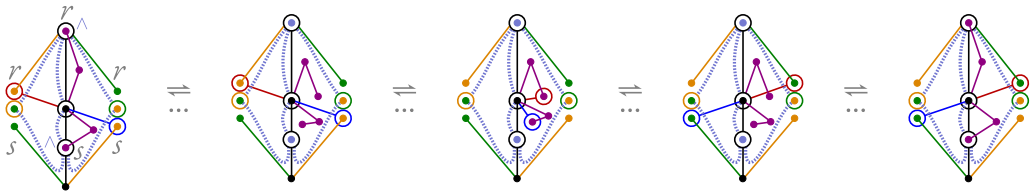
We can also allow the two linkages to interact by docking. The following construction shows an example of this. It adds matching domains, a, x, c, to three joints on each linkage. The dashed circle indicates a joint with two domains. The two linkages dock when the three joints all bind their partner. Notice that the two linkages can dock only when their states are complements since the result would be overbraced otherwise.



The following construction shows how we can prevent catalysis unless docked. Recall that the dashed lines represent the AND gadget from Example 7. So the internal catalytic arm can come free only when all of the docking sites are bound. This way the following two system states can reach each other, but only because their linkage states are complements.



The following figure shows the sequence of system states that flips the linkage states atomically, as a unit. This implements the chemo-mechanical coupling $A + B \rightleftharpoons C + D$ that we had as our goal.



6 Conclusion

Along with defining the metric and topological chemical linkage models, we have provided several examples of the complex behavior captured by them.

Throughout this work we have assumed a single-molecule regime where exactly one copy of the linkages shown is present in the system. Indeed, having multiple copies introduces potential problems. For example, in Example 3, two copies of the system can catalyze each other's state change even in the absence of the blue catalyst. Nonetheless, we imagine that an implementation of a chemical linkage would utilize other kinds of geometry to prevent such issues (e.g., through volume exclusion not captured by our linkage model). Indeed, linkages have a history of being used as a small part of a whole system (e.g., the steam engine is not entirely a linkage system, but the linkage model provided valuable insight into its function).

Some important theoretical and practical questions remain. One of the most immediate questions is whether or not the topological model captures the full power of the metric model. Are some behaviors easier to achieve when using explicit edge lengths? Also, this work considered two-dimensional linkages. Can this be generalized to three dimensions? Note that minimal rigidity can be generalized to 3D via Henneberg-like operations [16].

While the topological model simplifies the design and analysis of chemical linkages, what about their actual construction? The lengths that were removed for topological analysis will have to be added back in the real world. For this, we give the following conjecture:

▶ **Conjecture 12.** *Given any topological chemical linkage system, there exists a metric chemical linkage system which has the same reachable state space.*

Ultimately, we believe that chemical linkages and other simple chemo-mechanical models hold promise. Maybe they can help us understand the behaviors we see in living cells. And maybe they can help us mimic them.

References

1 R. Dean Astumian, Shayantani Mukherjee, and Arieh Warshel. The physics and physical chemistry of molecular machines. *ChemPhysChem*, 17(12):1719–1741, 2016.

2 E. Branscomb, T. Biancalani, N. Goldenfeld, and M. Russell. Escapement mechanisms and the conversion of disequilibria; the engines of creation. *Physics Reports*, 677:1–60, 2017.

3 Aidan I. Brown and David A. Sivak. Theory of nonequilibrium free energy transduction by molecular machines. *Chemical Reviews*, 120(1):434–459, January 2020.

4 Erik D. Demaine and Joseph O'Rourke. *Geometric folding algorithms: linkages, origami, polyhedra*. Cambridge university press, 2007.

5 Lebrecht Henneberg. *Die graphische Statik der starren Systeme*, volume 31. BG Teubner, 1911.

6 John Hopcroft, Deborah Joseph, and Sue Whitesides. Movement problems for 2-dimensional linkages. *SIAM journal on computing*, 13(3):610–629, 1984.

7 Donald J. Jacobs and Michael F. Thorpe. Generic rigidity percolation: the pebble game. *Physical review letters*, 75(22):4051, 1995.

8 Michael Kapovich and John J. Millson. Universality theorems for configuration spaces of planar linkages. *Topology*, 41(6):1051–1107, 2002.

9 Alfred B. Kempe. On a general method of describing plane curves of the nth degree by linkwork. *Proceedings of the London Mathematical Society*, pages 213–216, 1875.

10 Gerard Laman. On graphs and rigidity of plane skeletal structures. *Journal of Engineering mathematics*, 4(4):331–340, 1970.

11 Steven M. LaValle. *Planning algorithms*. Cambridge university press, 2006.

12 Tosan Omabegho. Allosteric linkages that emulate a molecular motor enzyme. *bioRxiv*, 2021. URL: `https://www.biorxiv.org/content/10.1101/2021.04.20.440673v1`.

13 John H. Reif. Complexity of the mover's problem and generalizations. In *20th Annual Symposium on Foundations of Computer Science (sfcs 1979)*, pages 421–427, 1979.

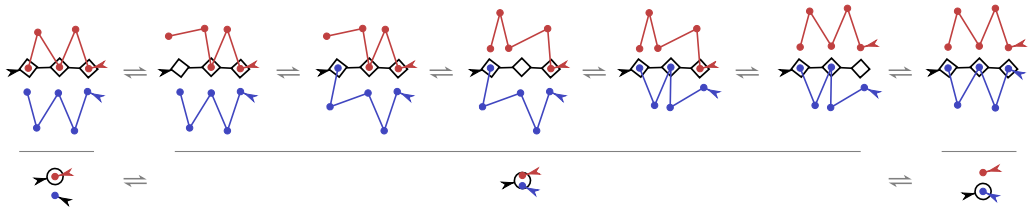
14 Jason W. Rocks, Nidhi Pashine, Irmgard Bischofberger, Carl P. Goodrich, Andrea J. Liu, and Sidney R. Nagel. Designing allostery-inspired response in mechanical networks. *Proceedings of the National Academy of Sciences*, 114(10):2520–2525, 2017.

15 Georg Seelig, David Soloveichik, David Yu Zhang, and Erik Winfree. Enzyme-free nucleic acid logic circuits. *Science*, 314(5805):1585–1588, 2006.

16 Tiong-Seng Tay and Walter Whiteley. Generating isostatic frameworks. *Structural Topology 1985 Núm 11*, 1985.

A Weak bonds

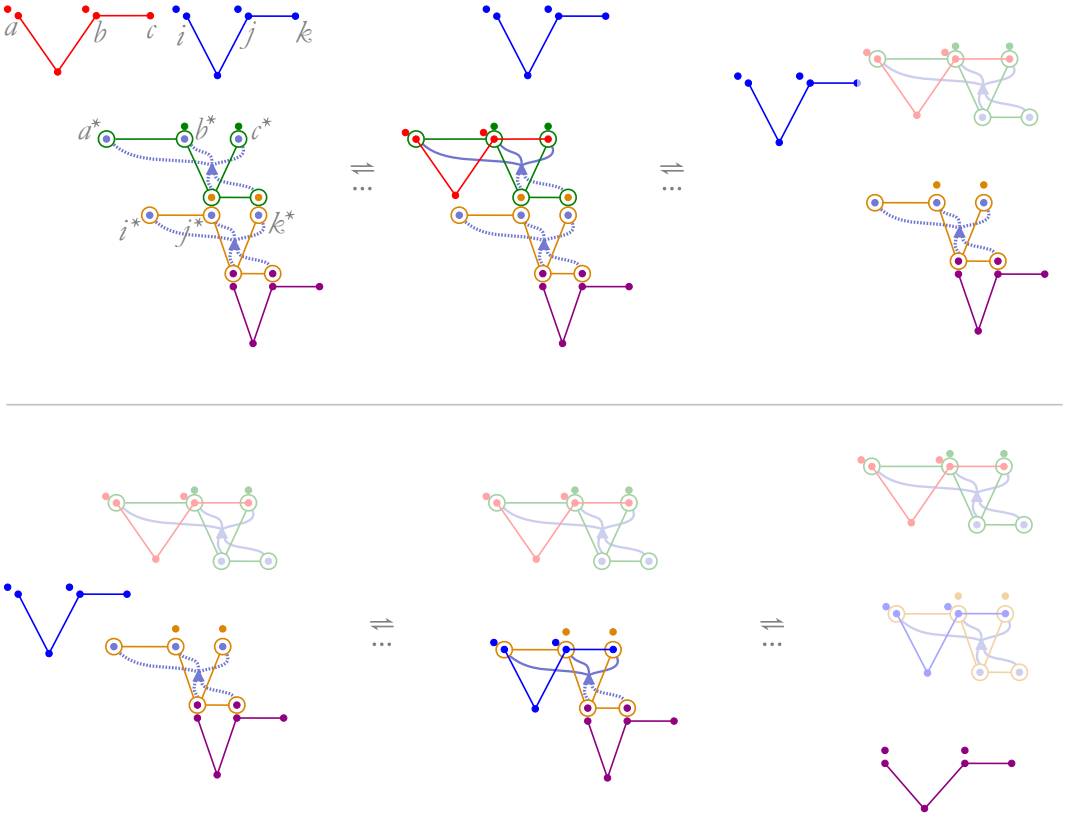
In the work above, we rely on strong unbreakable bonds. We can extend consideration to weak bonds which may break and reform. Indeed, we could consider weak bonds as a starting point of our model and construct strong bonds from weak bonds. The following example shows how a group of weak bonds (indicated as a diamond rather than a circle) can mimic strong bond displacement.



This example combines three weak bonds. Each can break individually, but they are unlikely to all break at the same time. So the only way the bottom blue linkage is likely to replace the top red linkage is by gradually displacing it. Note that this is very similar to DNA strand displacement. The bottom of the figure shows our standard representation of strong bond displacement.

B Sequential AND details

Here we show the state change sequence for the sequential AND construction.



In the initial system state (top left), the red linkage and blue linkage are separate from the green, orange, and purple linkages (which are all bonded to one another). Since the red signal and the green receptor are both active, they may dock with one another (top center). This bonding triggers the AND/FANOUT gadget, displacing the linkage pair and activating the orange receptor (top right). With the blue signal and the orange receptor both active (bottom left), they may dock to trigger the displacement mechanism (bottom center). This results in the purple signal linkage becoming active (bottom right) only when the red and blue signal links have bonded with their corresponding receptors (in the correct order).

C Modified ATP and catalyst

In order to apply the ATP from Section 5.3 to the motor from Section 5.2, we need to ensure that binding of the ATP and the catalyst results in displacement of joints both in the ATP and the catalyst. We can achieve this by splitting the original binding sites x and z into two different binding sites each of opposite complementarity allowing for a separation of

responsibility. As shown below with respect to the x binding site, one site (x_1) would be responsible for triggering the signal cascades within the ATP linkage. The other (x_2) would similarly displace the black joint in the motor.