



---

## The Building Game: From Enumerative Combinatorics to Conformational Diffusion

Daniel Johnson-Chyzykhov<sup>1,2</sup> · Govind Menon<sup>1</sup>

Received: 16 June 2015 / Accepted: 3 March 2016  
© Springer Science+Business Media New York 2016

**Abstract** We study a discrete attachment model for the self-assembly of polyhedra called the building game. We investigate two distinct aspects of the model: (i) enumerative combinatorics of the intermediate states and (ii) a notion of Brownian motion for the polyhedral linkage defined by each intermediate that we term conformational diffusion. The combinatorial configuration space of the model is computed for the Platonic, Archimedean, and Catalan solids of up to 30 faces, and several novel enumerative results are generated. These represent the most exhaustive computations of this nature to date. We further extend the building game to include geometric information. The combinatorial structure of each intermediate yields a systems of constraints specifying a polyhedral linkage and its moduli space. We use a random walk to simulate a reflected Brownian motion in each moduli space. Empirical statistics of the random walk may be used to define the rates of transition for a Markov process modeling the process of self-assembly.

**Keywords** Self-assembly · Polyhedra · Molecular cages · Polyhedral linkages · Brownian motion on manifolds

---

Communicated by Robert V. Kohn.

---

Partially supported by NSF DMS Grants 1148284 and 1411278.

---

✉ Govind Menon  
govind\_menon@brown.edu  
Daniel Johnson-Chyzykhov  
daniel\_johnson@alumni.brown.edu

<sup>1</sup> Division of Applied Mathematics, Brown University, 182 George St., Providence, RI 02912, USA

<sup>2</sup> Present Address: Hudson River Trading, 32 Old Slip, 30th Floor, New York, NY 10005, USA

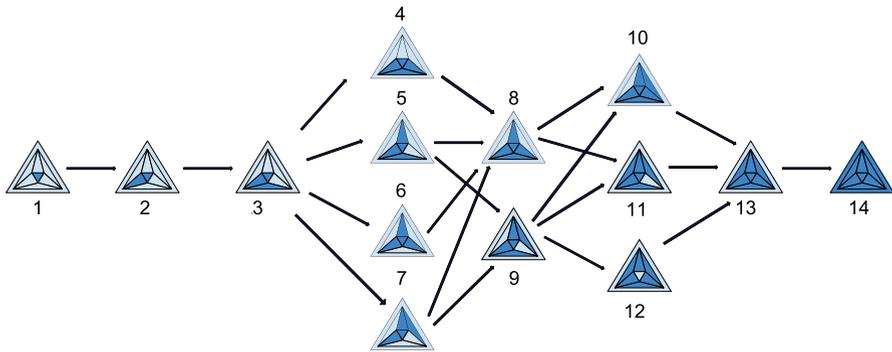
## 1 Introduction

The formation of closed-shell structures provides striking examples of self-assembly in nature. For instance, the discovery of  $C_{60}$ , buckminsterfullerene, revolutionized the understanding of fundamental forms of carbon (Kroto et al. 1985). In a similar vein, the study of viral capsids with icosahedral symmetry has played an important role in molecular biology (Caspar and Klug 1962; Crick and Watson 1956). These natural examples have inspired the development of several synthetic examples on larger scales, such as supramolecular polyhedral clusters (Liu et al. 2011; Sun et al. 2010; Li et al. 2011), synthetic DNA containers (Bhatia et al. 2009; Douglas et al. 2009), and self-folding polyhedra (Pandey et al. 2011).

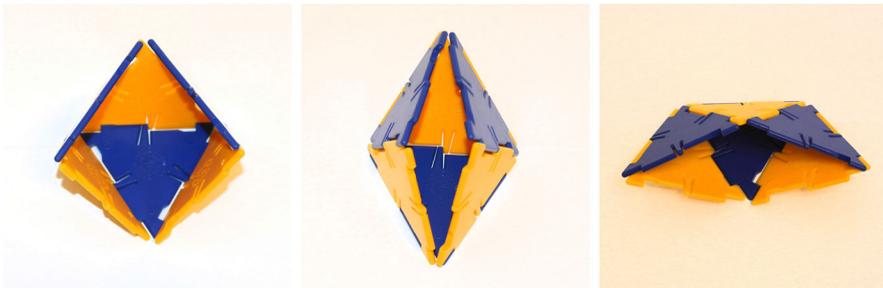
Our purpose in this paper, and a companion paper (Russell and Menon 2016), is to present a set of mathematical results inspired by the self-assembly of closed-shell structures. In this paper, we study mathematical properties of the *building game* (Wales 1987; Zlotnick 1994). In Russell and Menon (2016), we present a mathematical model for recent experiments on synthetic supramolecular cages. Both papers share a common perspective rooted in discrete geometry and statistical physics. However, the papers address distinct challenges and may be read independently.

The building game is a discrete growth-by-attachment model for a fixed polyhedron that may be described informally as follows. The process begins with a cluster consisting of a single face. At each step, a face of the polyhedron that is not in the cluster but shares an edge with it is chosen and attached to the cluster. The process terminates when all faces of the polyhedron are part of the cluster. The *configurations* or *intermediates* of the system consist of all possible clusters formed by this process modulo the symmetry group of the polyhedron. The *combinatorial configuration space*,  $\mathcal{C}$ , is the set of all configurations, equipped with the structure of a graph: Two configurations that differ by a single face are neighbors. Pathways of assembly are paths in the combinatorial configuration space that begin with a single face and terminate in the complete polyhedron. These notions are illustrated in Fig. 1.

In a recent article, one of the authors (GM), along with other co-authors, illustrated the utility of such discrete geometric models of self-assembly for two problems: the self-assembly of the bacteriophage MS2 and surface tension-driven self-folding (Kaplan et al. 2014). While the work presented in the current paper is inspired by similar scientific questions, it is focused primarily on mathematical properties of the building game. The novel results in this work relate to two distinct issues: (i) enumerative problems and (ii) random motion of polyhedral linkages. In the first two sections of this paper, we formalize and present a set of new enumerative results on the building game. Despite the fact that the size of the configuration space explodes combinatorially, it is necessary to perform exhaustive computations in all tractable cases, in order to develop heuristics and sampling techniques for situations where the configuration space is computationally intractable. The results presented here are the most extensive such computations to date. We expect that these computational results will be useful independent of any particular experimental system, much in the same



**Fig. 1** Combinatorial configuration space for the octahedron. Assembly proceeds by the attachment of one face at a time. Each configuration consists of a contiguous cluster of triangular faces attached at edges. It is convenient to graphically represent each three-dimensional cluster by projecting it into two-dimensions as a Schlegel diagram. In the figure above, each cluster is represented by the *dark blue triangles*. The combinatorial configuration space is the directed graph of all configurations. The directed edges are denoted by *arrows above*. An assembly pathway is a pathway from intermediate 1 (consisting of one face) to the intermediate 14 (the completely formed octahedron) (Color figure online)



**Fig. 2** Geometry of embedded configurations. Each intermediate in the combinatorial configuration space determines a set of polynomial constraint equations as explained in Sect. 4. The solution set (over the reals) of these constraint equations is a real algebraic variety. Points on this variety that also satisfy the constraint of nonself-intersection correspond to geometric embeddings of a polyhedral linkage. In this figure, we illustrate three embeddings of intermediate 10 in Fig. 1. It is simplest to first embed this configuration by removing two contiguous faces from the standard embedding of the complete octahedron (*left figure*) and to then squash it into the two embeddings on the *right*. This linkage has one internal degree of freedom. In order to visualize ‘conformational diffusion’ the reader should imagine a continuous random jiggling that explores embeddings like those above

way that detailed computations of idealized models (e.g., configurations of minima in the Lennard–Jones system with  $n$  particles) are useful in a variety of problems.<sup>1</sup>

The second, and main, theme in this work is to relate the geometry of intermediates to rates of transition between them. The adjective *combinatorial* in the definition of the configuration space reflects the fact that each configuration is described completely by a 2-coloring of the faces of the polyhedron. It is included to contrast purely combinatorial notions with the geometric properties of intermediates, related to the manner in which they are embedded in  $\mathbb{R}^3$ . As seen in Fig. 2, each configuration determines a

<sup>1</sup> These data are publicly available at the Brown Digital Repository (Johnson and Menon 2016).

polyhedral linkage. *Our main thesis is that the kinematics of linkages should play an important role in self-assembly.* This insight is motivated by experiments on synthetic self-assembly. We noted the importance of rigidity of intermediates in experiments with self-folding polyhedra (Kaplan et al. 2014; Pandey et al. 2011). In a similar vein, the importance of kinematics of collections of rigid spherical clusters has been noted in Holmes-Cerfon et al. (2013).

While the building game intermediates are not the same as those seen in self-folding, the size of the combinatorial configuration space is much smaller, making it a more attractive model for a mathematical study. Further, in instances where the building game is used as a model for the formation of closed-shell structures, it is natural to use the kinematics of a linkage to develop a notion of random internal motion of each intermediate. Indeed, perhaps the simplest physical caricature of the formation of a closed-shell structure is the following. We assume that the fundamental unit in the closed-shell structure ('monomers') is elastic, but stiff, triangles with sticky edges in a thermal bath. The monomers diffuse at random in the bath, sticking with a certain probability when they collide. When two or more triangles meet, they form intermediate states. These intermediate states are subject to two forms of random motion. On the one hand, their center of mass diffuses in space. On the other hand, they also undergo internal vibrations, adopting different conformations. As a first approximation of such conformational diffusion, it is natural to consider the random motion of a *rigid* triangular polyhedral linkage as considered in this paper. Further, since the monomers are assumed to be stiff, it is natural to expect that the probability of attachment or detachment of monomers depends on the internal conformation, with high probability of attachment when the intermediate adopts a conformation that is close to a partially formed shell.

We formalize these ideas by introducing a stochastic process that corresponds to the exploration of internal degrees of freedom by intermediates composed of rigid faces. Each configuration defines a polyhedral linkage, or equivalently, an algebraic variety. We call the collection of algebraic varieties fibered over the combinatorial configuration space, the *geometric configuration space*,  $\mathcal{G}$ . We study reflected Brownian motion on each algebraic variety within the geometric configuration space.<sup>2</sup>

The link between the enumerative problems and conformational diffusion is this: The kinetics of self-assembly are naturally modeled by a Markov process that hops between states in the combinatorial configuration space,  $\mathcal{C}$ . However, the combinatorial explosion in the size of the configuration space makes it impossible to determine the transition rates of this Markov process from experimental data (since there is one rate constant for each edge in the combinatorial configuration space). It is necessary to determine rates by computation, and reflected Brownian motion allows us to make precise the notion of 'typical shape' for a configuration and to compute rates of transition between states based on Brownian motion on each algebraic variety within  $\mathcal{G}$ .

The remainder of this article is organized as follows. A mathematical formalism for the building game is introduced in the next section. This is followed by a set of enumerative results in Sect. 3. We introduce the moduli space of a configuration and

<sup>2</sup> More precisely, we study a discretized reflected random walk on each algebraic variety. It has *not* been proven in general that these random walks converge to a reflected Brownian motion on each variety.

the geometric configuration space,  $\mathcal{G}$ , in Sect. 4. The notion of degrees of freedom is explored in Sect. 5. This is followed by a study of random walks on moduli space and the extraction of transition rates in Sect. 6.

As noted above, the paper naturally breaks into two distinct parts: the first on enumerative results and the second on the kinematics and random motion of polyhedral linkages. These two parts are largely independent of one another. The primary purpose of Sect. 2 is to provide precise definitions of configurations obtained by the building game, in a way that accounts for symmetry, since these definitions underlie Sect. 3. The reader interested primarily in the random motion of polyhedral linkages is advised to simply skim over the next two sections, noting the intuitive meaning of the building game in Fig. 1 and the combinatorial explosion implicit in the enumerative results tabulated in Sect. 3.

## 2 The Building Game: Mathematical Formalism

### 2.1 Introduction

The building game was introduced by Wales (1987) as a model for the formation of fullerenes. Zlotnick subsequently used it to model the assembly of polyhedral viral capsids (Zlotnick 1994). The model consists of a graph of intermediate configurations and a Markov process defined on this graph. In what follows, we focus on the description of the graph of intermediate configurations.

In a nutshell, the configurations may be naturally described by a coloring process. Given a polyhedron with each face painted red, choose a face and paint it blue. At each subsequent step, choose a red face that is adjacent to a blue face and paint it blue; stop when all faces are blue. Each step of this process yields a connected cluster of blue faces that represents a partially formed state in the assembly process. We further identify all states that are equivalent under the action of the rotation group of the polyhedron, and call an equivalence class of states a configuration. The abstract notions in this subsection are simply included in order that these equivalence classes are defined precisely.

The principal virtue of the building games is that it provides a simple formulation for essential features of self-assembly such as a description of partially formed states, pathways, and kinetics of assembly. However, it does not account for more realistic features such as the formation of malformed capsids. The reader is referred to Kaplan et al. (2014) for a more detailed description of the advantages and drawbacks of this model in the context of capsid formation. In this paper, we focus on its mathematical properties.

### 2.2 Definitions

Assume that given a polyhedron,  $\mathcal{P}$ , whose labeled faces, edges, and vertices are the sets  $\mathcal{F}$ ,  $\mathcal{E}$ , and  $\mathcal{V}$ , respectively. Also assume that given a symmetry group,  $\mathbf{G}$ , of the polyhedron. In this paper,  $\mathcal{P}$  is a Platonic, Archimedean, or Catalan solid and  $\mathbf{G}$  is its group of rotational symmetries. Let  $2^{\mathcal{F}}$  denote the set of all subsets of  $\mathcal{F}$ . We denote

a subset  $x \in 2^{\mathcal{F}}$  by  $x = \{f_1, \dots, f_{|x|}\}$ ,  $f_j \in \mathcal{F}$ ,  $1 \leq j \leq |x|$ . We say that a subset  $x \in 2^{\mathcal{F}}$  is *connected* if each face  $f_i \in x$  shares an edge with another face  $f_j \in x$ . Let  $X \subset 2^{\mathcal{F}}$  denote the set of connected subsets. A *state* for the building game is an element  $x \in X$ . The states  $x \in X$  may also be described by 2-colorings. A 2-coloring of the faces of the polyhedron is a function  $c : \mathcal{F} \rightarrow \{0, 1\}$ . We adopt the convention that 0 is ‘red’ and 1 is ‘blue.’ Then each state  $x$  of the building game is a subset  $x = c^{-1}\{1\} \subset \mathcal{F}$  such that  $x$  is connected.

We have indicated the connection with colorings, since the enumeration of colorings is a classic problem in discrete group theory (Rotman 1995) and it allows us to make contact with (Russell and Menon 2016). The group  $\mathbf{G}$  acts naturally on the set  $X$  as follows: given  $g \in G$  and  $x = \{f_1, \dots, f_{|x|}\} \in X$ , the set  $gx$  consists of the faces  $\{gf_1, \dots, gf_{|x|}\}$ . We say that two states  $x$  and  $y$  are equivalent, written  $x \sim y$ , if there exists  $g \in G$  such that  $gx = y$ . Finally, we define a *configuration* to be the set of equivalence classes  $[x]$  of states  $x \in X$ , under the equivalence relation  $\sim$ .

Recall that the *orbit* of an element  $x \in X$  is the subset  $G \cdot x \doteq \{g \cdot x : g \in \mathbf{G}\}$ . Thus, each configuration  $[x]$  in the building game may be considered as the group orbit  $\mathbf{G} \cdot x$ . Group orbits are classically enumerated via Burnside’s lemma: The number of configurations (orbits),  $|X/\mathbf{G}|$ , is given by

$$|X/G| = \frac{1}{|\mathbf{G}|} \sum_{g \in \mathbf{G}} |X^g| \tag{1}$$

where  $X^g = \{x \in X : g \cdot x = x\}$  is the set of states fixed by  $g$ . The size of any equivalence class (orbit) is enumerated in a similar manner. Recall that the *stabilizer subgroup* of an element  $x \in X$  is the subgroup  $\mathbf{G}_x \doteq \{g \in \mathbf{G} : gx = x\}$  of  $\mathbf{G}$  that fixes  $x$  (Rotman 1995). Then by the orbit stabilizer theorem and Lagrange’s theorem, the size of any group orbit is

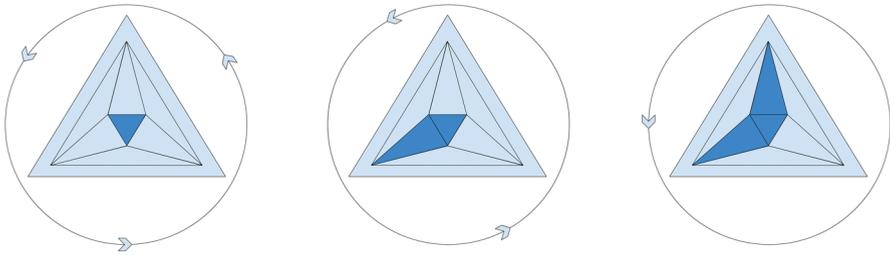
$$|\mathbf{G} \cdot x| = \frac{|\mathbf{G}|}{|\mathbf{G}_x|}. \tag{2}$$

We say that the *symmetry number*  $r_x$  of a state  $x$  is the order of its stabilizer subgroup  $|\mathbf{G}_x|$ . The symmetry number  $r_x$  is the same for all states in the equivalence class  $[x]$ . Indeed, if  $x$  and  $y$  are states in  $[x]$ , we have

$$r_x = |\mathbf{G}_x| = \frac{|\mathbf{G}|}{|\mathbf{G} \cdot x|} = \frac{|\mathbf{G}|}{|\mathbf{G} \cdot y|} = |\mathbf{G}_y| = r_y. \tag{3}$$

In Fig. 3, we see three states of the octahedron and their orbit stabilizer subgroups. The first state, with only a single face, has a symmetry number of 3 since any rotation by a multiple of  $\frac{2\pi}{3}$  fixes the face. The second has a symmetry number of 2 since only the identity and a rotation by  $\pi$  will fix the faces. The final state, with three faces, cannot be fixed by any rotation other than the identity and thus has symmetry number 1.

There is a natural notion of neighbors on the set  $X$ . We say that two states  $x$  and  $y$  are neighbors if they differ by a single face. There are two possibilities. Either  $x \subset y$ ,



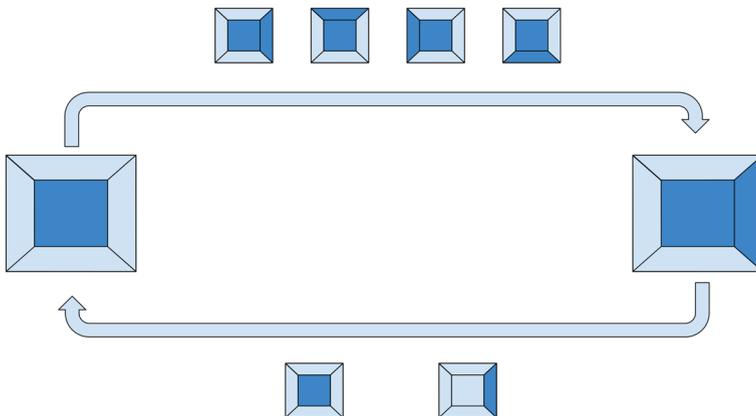
**Fig. 3** Stabilizer subgroups for three octahedron states. States are represented by Schlegel diagrams as in Fig. 1

and  $|y| = |x| + 1$  and we say that  $y$  is obtained from  $x$  by a forward step; or  $y \subset x$ ,  $|x| = |y| + 1$  and we say that  $y$  is obtained from  $x$  through a backward step. If  $x$  and  $y$  are neighbors, so are  $gx$  and  $gy$  for all  $g \in G$ . We say that the orbits  $[x]$  and  $[y]$  are neighbors if there exist states  $x \in [x]$  and  $y \in [y]$  such that  $x$  and  $y$  are neighboring states. Two neighboring configurations may be linked by the attachment or detachment of several possible faces. If  $x$  and  $y$  are neighboring states, the state  $x$  may be connected to other states  $z \in [y]$ . The set of different faces

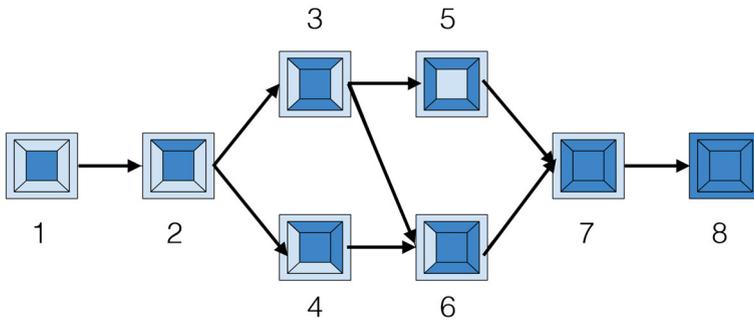
$$F_{xy} := \{f \notin x : x \cup \{f\} \in [y]\} \cup \{f \in x : x \setminus \{f\} \in [y]\}$$

that can be added or removed from  $x$  to get an element of  $[y]$  is called the *degeneracy set*, and the number of such faces  $S_{xy} := |F_{xy}|$  is called the *degeneracy number*.

It is important to note that  $F_{xy}$  is in general not the same as  $F_{yx}$  (see Fig. 4). Further, while it is clear from the definition that  $F_{xy}$  depends only on  $[y]$ , it may also be checked that  $F_{xy}$  depends only on  $[x]$  and that the degeneracy and symmetry number are related by Johnson (2015, Theorem 4)



**Fig. 4** Degeneracy is not symmetric. A face may be attached to one of the four edges to obtain the intermediate on the right from the one on the left. However, the removal of a face can take place in only two ways



**Fig. 5** Combinatorial configuration space for the cube. The pathways 1, 2, 4, 6, 7, 8 and 1, 2, 3, 6, 7, 8 provide shellings of the cube. However, intermediate 5 is not shellable and 1, 2, 3, 5, 7, 8 is not a shelling pathway

$$r_x S_{xy} = r_y S_{yx} \tag{4}$$

**Definition 1** The *combinatorial configuration space*, denoted  $\mathcal{C}$ , for  $\mathcal{P}$  is the set of configurations  $X/G$  equipped with edges linking neighboring configurations.

In order to avoid any confusion with the edges,  $\mathcal{E}$ , of the polyhedron, we will refer to these edges as *connections*. An *assembly pathway*, or *pathway* for short, is a sequence of intermediates  $[x^0], [x^1], \dots, [x^{|\mathcal{F}|}]$  such that  $[x^j]$  is connected to  $[x^{j+1}]$ ,  $|x^k| = k$ ,  $[x^0] = \emptyset$ , and  $x^{|\mathcal{F}|} = \mathcal{F}$ . Figures 1 and 5 show the combinatorial configuration spaces for the octahedron and cube, respectively.

### 2.3 Shellable Configurations

This section may be omitted on a first reading. It is included to connect the building game with the classical notion of shelling of polytopes.

The building game is a physically motivated schema for the assembly of a polyhedron from its faces. More broadly, the idea that complex polytopes can be constructed by gluing simpler polytopes, such as simplices, is fundamental to the theory of polytopes and combinatorial topology. The simplest examples of such results are provided by triangulations of two-dimensional polygons. A fundamental fact about triangulations, which underlies the usual proof of Euler’s formula  $V - E + F = 2$  for connected planar graphs, is that the removal of a triangle from a triangulation does not change the Euler characteristic of the triangulation. The extension of this idea to higher dimensions is subtle for the following reason: There exist simplicial decompositions of the tetrahedron for which the Euler characteristic changes when one simplex is removed (Rudin 1958).

Loosely speaking, a shelling of a polytope is a construction of the polytope by the attachment of basic units (e.g., vertices, faces, simplices) such that all the intermediates along the pathway are homeomorphic. All the polyhedra considered in this work can be viewed as spherical polyhedra (i.e., tilings of the sphere, with edges given by great arcs). For such polyhedra, a shelling is defined as follows (Whiteley 1994, p. 152).

We view each face as a closed subset of the sphere, and note that the boundary of each face determines a spherical polygon.

**Definition 2** A shelling of  $\mathcal{P}$  is an ordering of the faces (as polygons),  $f_1, f_2, \dots, f_k, f_{k+1}, \dots, f_{\mathcal{F}}$  such that the intersection of each face  $f_{k+1}$  with the union of the previous faces  $\cup_{j=1}^k f_j$  is a single, connected polygonal path with at least one edge.

Although the notion of shelling is usually applied to a full polyhedron, we may adapt it to partially formed polyhedra in a natural manner. It is only necessary to quotient out the group action of  $\mathbf{G}$  to obtain the appropriate notion of shellability for configurations.

**Definition 3** A configuration  $[x]$  is **shellable** if for every state  $x \in [x]$  there is a linear ordering  $f_1, f_2, \dots, f_{|x|}$  on the faces of  $x$  such that this ordering is the leading sequence of a shelling  $f_1, f_2, \dots, f_{|x|}, \dots, f_{|F|}$  of the polyhedron  $\mathcal{P}$ .

We also say that a connection is **ashellable connection** if the two configurations it connects are both shellable. A **shellable pathway** is a pathway composed entirely of shellable configurations  $[x^0], \dots, [x^{|\mathcal{F}|}]$ .

We illustrate these ideas for the cube and octahedron. The combinatorial configuration space for the cube is shown in Fig. 5. Both 1, 2, 4, 6, 7, 8 and 1, 2, 3, 6, 7, 8 are shelling pathways. However, not each pathway provides a shelling. When a face is attached to state 3 to obtain state 5, the intersection of its boundary with the union of previous faces yields two disconnected lines. Thus, intermediate 5 is not shellable, and 1, 2, 3, 5, 7, 8 is not a shelling pathway. Similarly, the reader may check that the octahedral configuration 9, 11, and 12 in Fig. 1 are not shellable (the closed subsets of the sphere defined by 7 and 9 are not homeomorphic, because 9 includes a hole). Only the pathways that contain the sequence 8, 10 are shellable—this gives 4 shellable pathways for the octahedron (cf. Table 2).

As a demonstration of the utility of building game configuration space, we present the following enumerative result. For any polyhedron  $\mathcal{P}$ , the number of shellings is obtained by recursively summing over all shellable pathways, weighting each path by the degeneracy of the states along the path and we find (Johnson 2015, Theorem 7)

$$\#(\text{shellings}) = \sum_{\substack{\text{shellable} \\ \text{pathways}}} |[x_1]| \prod_{j=1}^{|\mathcal{F}|-1} S_{[x_j][x_{j+1}]} \tag{5}$$

Thus,  $\mathcal{C}$  always admits a subgraph consisting of shellable intermediates and connections. Since shellability is somewhat tangential to the discussion in this paper, the interested reader is referred to Johnson (2015) for a proof of this result.

### 3 Enumerative Results

While the problem of computing the combinatorial configuration space,  $\mathcal{C}$ , is clearly related to classical enumerative problems in graph theory, such as the enumeration

**Table 1** Enumerative results for the Platonic, Archimedean, and Catalan solids

Polyhedron	$ F $	Intermediates	Connections	Pathways
Tetrahedron	4	4	3	1
Cube	6	8	9	3
Octahedron	8	14	21	14
Dodecahedron	12	73	263	17,696
Icosahedron	20	2,649	17,241	57,396,146,640
Truncated tetrahedron	8	28	63	402
Cuboctahedron	14	340	1,634	10,170,968
Truncated cube	14	499	2,729	101,443,338
Truncated octahedron	14	555	3,069	68,106,377
Rhombicuboctahedron	26	638,850	6,459,801	164,068,345,221,515,292,308
Truncated cuboctahedron	26	1,525,658	17,672,374	13,837,219,462,483,379,105,902
Triakis tetrahedron	12	98	318	38,938
Rhombic dodecahedron	12	127	493	76,936
Triakis octahedron	24	12,748	81,296	169,402,670,046,670
Tetrakis hexahedron	24	50,767	394,377	4,253,948,297,210,346
Deltoidal icositetrahedron	24	209,675	1,989,548	418,663,242,727,526,726
Pentagonal icositetrahedron	24	345,938	3,544,987	2,828,128,000,716,774,492
Rhombic triacontahedron	30	2,423,212	26,823,095	161,598,744,916,797,017,978,128

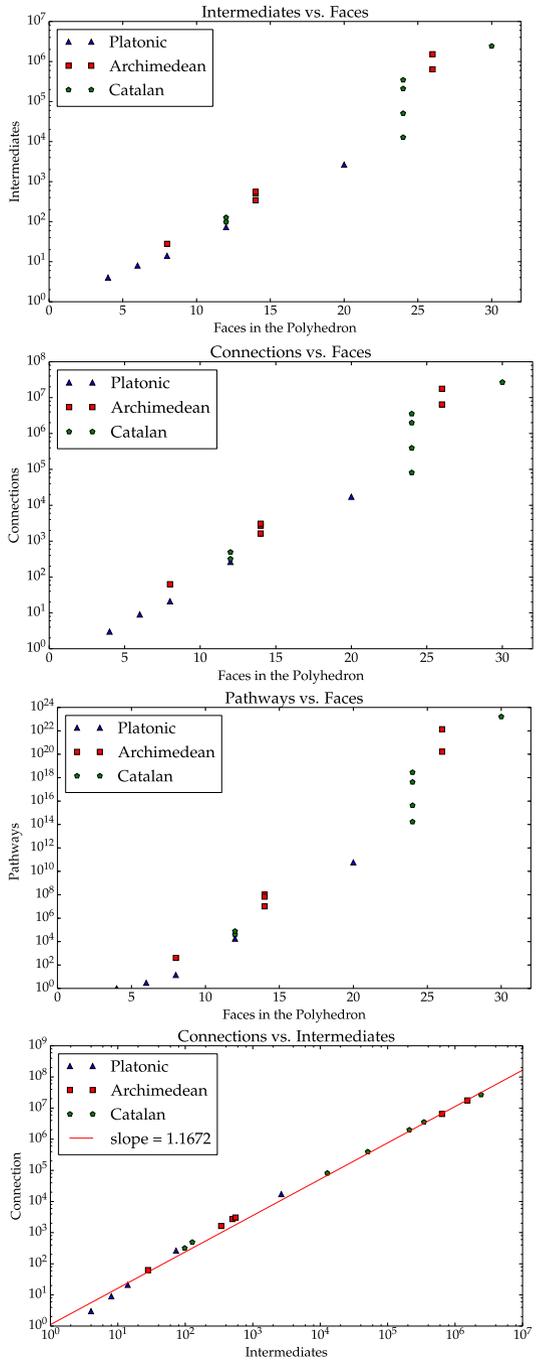
of monochromatic subsets of colored graphs, we are unaware of any theoretical enumerative results on the building game.<sup>3</sup> The first computational result is due to David Wilson, who enumerated the number of configurations for the icosahedron (Sloane 2003, sequence A030138) (this computation was recreational, motivated by polyominoes). In the context of viral self-assembly, Zlotnick and his co-workers enumerated the configurations for the Platonic solids (Endres et al. 2005).

In Table 1 below, we present comprehensive enumerative results on the building game for all Platonic, Archimedean, and Catalan solids. Our results also include an enumeration of connections and pathways. The most obvious feature of the problem is the combinatorial explosion in the number of intermediates for polyhedra of increasing complexity. This is illustrated with several graphs in Fig. 6. The number of intermediates is roughly of the same order of magnitude as the number of connections between intermediates, but the number of pathways is far larger.

We have also compared building game intermediates with shellable intermediates. A dynamic programming procedure was implemented to enumerate the number of shellings via formula (5). Tables 2 and 3 detail the shellability statistics for the Platonic, Archimedean, and Catalan solids. Because of the added restriction of shellability, the

<sup>3</sup> By ‘theoretical’ we mean results that are obtained without explicit enumeration on a computer. Closest in spirit to our work are enumerative results on polyominoes, configurations of  $n$  attached squares on a planar lattice.

**Fig. 6** Relation between number of faces and intermediates, connections and pathways in the combinatorial configuration space,  $C$



**Table 2** Shellable intermediates for the Platonic, Archimedean, and Catalan solids

Polyhedron	$ F $	Shellable intermediates	Shellable connections	Shellable pathways
Tetrahedron	4	4	5	1
Cube	6	7	7	2
Octahedron	8	11	13	4
Dodecahedron	12	52	155	2,166
Icosahedron	20	469	1,985	105,999,738
Truncated tetrahedron	8	21	40	174
Cuboctahedron	14	136	468	477,776
Truncated cube	14	247	1,000	5,232,294
Truncated octahedron	14	342	1,464	5,704,138
Rhombicuboctahedron	26	70,887	462,721	64,308,526,503,247,584
Truncated cuboctahedron	26	515,335	4,070,813	13,890,723,216,176,694,816
Triakis Tetrahedron	12	48	115	5,012
Rhombic dodecahedron	12	67	195	6,258
Triakis octahedron	24	1,021	4,237	210,459,770,300
Tetrakis hexahedron	24	4,224	21,125	5,894,431,702,846
Deltoidal icositetrahedron	24	33,046	208,317	703,619,122,996,096
Pentagonal icositetrahedron	24	95,326	657,013	7,572,459,719,248,765
Rhombic triacontahedron	30	97,741	702,219	7,057,239,571,753,327,764

**Table 3** Number of shellings for the Platonic, Archimedean, and Catalan solids with at most 30 faces

Polyhedron	$ F $	Shellings
Tetrahedron	4	24
Cube	6	480
Octahedron	8	4,224
Dodecahedron	12	19,041,600
Icosahedron	20	1,417,229,099,520
Truncated tetrahedron	8	9,216
Cuboctahedron	14	113,055,744
Truncated cube	14	654,801,408
Truncated octahedron	14	937,087,104
Rhombicuboctahedron	26	4,728,400,467,971,102,208
Truncated cuboctahedron	26	688,499,026,944,479,645,952
Triakis tetrahedron	12	587,040
Rhombic dodecahedron	12	5,836,800
Triakis octahedron	24	66,063,419,534,592
Tetrakis hexahedron	24	1,389,323,257,015,296
Deltoidal icositetrahedron	24	125,987,819,253,281,472
Pentagonal icositetrahedron	24	1,144,572,832,023,047,616
Rhombic triacontahedron	30	15,574,782,555,813,226,074,240

number of intermediates and connections is lower than in the general case, but the combinatorial growth as the number of faces increases is similar.

All the results included here are based on exhaustive ‘brute-force’ computation, though various techniques are necessary to streamline the computation and organize the data. The methods for similar computations have been discussed in print before (Pandey et al. 2011) and are not repeated here. The reader interested in the details of the computation is referred to Johnson (2015).

## 4 The Geometric Configuration Space

### 4.1 The Problem of Rates

The combinatorial explosion in the number of connections in the configuration space,  $\mathcal{C}$ , starkly highlights the challenge of using the building game as a model for self-assembly. In order to define a Markov process on  $\mathcal{C}$ , we must associate a rate constant to each connection between neighbors in  $\mathcal{C}$ . If the rates are to be determined from experimental data, we are faced with a parameter estimation problem that is impossible to resolve for even the simple polyhedra considered here.

Various strategies may be adopted to resolve this problem, while still retaining the essential simplicity of the building game. One possibility is to prune the graph,  $\mathcal{C}$ , retaining only a subgraph that we hope will capture the essential configurations in self-assembly. Zlotnick and co-workers adopt this strategy, pruning the set of assembly pathways for the Platonic solids to a single path linking neighboring configurations that have minimal energy (Endres et al. 2005). Numerical experiments reveal that such procedures, while admittedly *ad hoc*, may be quite robust—we observed in Pandey et al. (2011), Kaplan et al. (2014) that many natural choices of rate functions led to pathways that focus through the same configurations. A second strategy is to resolve rates by computation, rather than (physical) experiments. In this approach, it is necessary to augment the purely combinatorial description of configurations in  $\mathcal{C}$  with geometric, chemical, or physical properties that determine the rates of transition between states. This is the approach we adopt below.

Let us fix an embedding of  $\mathcal{P}$  in  $\mathbb{R}^3$  and view states  $x \in 2^{\mathcal{F}}$  as subsets of  $\mathbb{R}^3$ . Then, each configuration  $[x]$  in  $\mathcal{C}$  admits an embedding in  $\mathbb{R}^3$  given by a subset  $x \in \mathcal{P}$ , for any  $x \in [x]$ . We call this the *standard embedding*. It is unique modulo the action of  $\mathbf{G}$ . However, viewing  $x$  as a polyhedral linkage, we see that the standard embedding is simply one among infinitely many if the linkage has nontrivial internal degrees of freedom (see for example, Fig. 2). The set of embeddings of  $x$  is described by an algebraic variety. We define a stochastic process in this variety and use it to define rates of transition between configurations.

Our approach focuses on the geometric properties of configurations, ignoring (at present) physical modeling. There are several reasons for this choice: First, it is natural to expect that physical modeling of configurations will only augment our basic model—for example, we may introduce attachment and detachment probabilities based on whether the embedded configuration is close to a standard state consisting of a connected subset of faces of  $\mathcal{P}$ . Second, in recent work (Pandey et al. 2014)

we explored the role of rigidity and degrees of freedom in self-folding polyhedra. The current study allows us to explore mathematical ideas implicit in that paper in greater detail. Finally, the theory of linkages provides an elegant approach to the mathematical problem of constructing manifolds with prescribed geometry. Each manifold is identified with the moduli space of a set of constraint equations (Thurston and Weeks 1984). Thus, it is of intrinsic mathematical interest to explore the geometric properties of the manifolds defined by each configuration  $[x] \in \mathcal{C}$ .

### 4.2 The Constraint Equations

In this section, we assume that a state  $x \subset \mathcal{P}$  is fixed. We always assume that the edges and faces of the polyhedron are rigid. No conditions are imposed at points where faces meet, except that of connectivity between the faces (i.e., vertices on the common edge are identified). Thus, an embedding of the state  $x$  into  $\mathbb{R}^3$  is completely prescribed by a list of coordinates for each vertex in  $x$  that is in accordance with the constraints of rigidity of each edge and face. Let  $s_f$  denote the number of vertices in a face  $f \subset x$ , and let  $N_x = \sum_{f \in x} s_f$ . We denote the  $k$ th vertex of the  $j$ th face of  $x$  by  $v^{jk} = (v_x^{jk}, v_y^{jk}, v_z^{jk})$ . It is convenient to list all coordinates in a single vector,  $z$ , of length  $n = 3N_x$ , defined by

$$z = \begin{bmatrix} v^{1,1} \\ \vdots \\ v^{1,s_{f_1}} \\ \vdots \\ v^{|x|,1} \\ \vdots \\ v^{|x|,s_{f_{|x|}}} \end{bmatrix} \in \mathbb{R}^n. \tag{6}$$

Each embedding of  $x$  is a point  $z \in \mathbb{R}^n$  that solves the constraint equation  $c(z) = 0$  where the constraints  $c^{j,k}(z)$  are listed in Table 4. As explained below, these equations are not independent in general (but this does not affect the definitions or our computation). Finally, when treating linkages it is important to prevent self-intersection of faces. That is, while each embedding solves the equations in Table 4, the only solutions that correspond to a physical linkage are those that are also nonself-intersecting.

We now explain the quadratic constraints in Table 4 in greater detail. The first constraint fixes the length of edges. If the  $k$ th edge is defined to be that between the  $(k - 1)$ st and  $k$ th vertices, we use the following function to constrain its lengths to a known value  $\ell^{jk}$ :

$$c_{edge}^{j,k}(z) = |v^{j,k} - v^{j,k-1}|^2 - (\ell_{j,k})^2. \tag{7}$$

**Table 4** Constraints defining the moduli space  $\mathcal{M}_{[x]}$

Constraint type	Polynomial
Edge length	$c_{edge}^{j,k}(z) =  v^{j,k} - v^{j,k-1} ^2 - (\ell_{j,k})^2$
Angle	$c_{ang}^{j,k}(z) = (v^{j,k-1} - v^{j,k}) \cdot (v^{j,k+1} - v^{j,k}) - \ell^{j,k} \ell^{j,k+1} \cos(\theta^{j,k})$
2D face	$c_{2D}^{j,k}(z) = v^{1,k} + \ell^{j,k,1} R(v^{j,0} - v^{j,1}) - v^{j,k}$
Vertex identification	$c_{ident}^{j_1,k_1,j_2,k_2,d}(z) = v_d^{j_1,k_1} - v_d^{j_2,k_2}$

The constraint that the angle at each vertex of each face is fixed may be written as a polynomial constraint using the dot product

$$c_{ang}^{j,k}(z) = (v^{j,k-1} - v^{j,k}) \cdot (v^{j,k+1} - v^{j,k}) - \ell^{j,k} \ell^{j,k+1} \cos(\theta^{j,k}). \tag{8}$$

Here,  $\theta^{j,k}$  is the angle between the  $k$ th and  $(k + 1)$ st edges at the vertex  $v^{j,k}$ .

In practice, we need only explicitly enforce the lengths of the first two edges ( $k = 1, 2$ ) and the first angle on each face. Indeed, this fixes a rigid triangle on the face consisting of the first three vertices. Thus, there are a total of  $2|x|$  independent edge length constraints and  $|x|$  independent angle constraints. Once the positions of three distinct vertices on each face has been determined, it is somewhat simpler to use a rotation matrix and a template for each polygonal face to determine the remaining vertices (rather than solve the nonlinear systems (7) and (8)). The template is a set of coordinates for the vertices in a fixed reference configuration. We denote it by the list  $\hat{v}^{j,0}, \hat{v}^{j,1}, \hat{v}^{j,2}, \dots, \hat{v}^{j,k}, \dots$ . Once the locations for  $v^{j,0}, v^{j,1}$ , and  $v^{j,2}$  are determined, we can identify the location of  $v^{j,k}$  for  $k > 2$  using the following length and angle constants

$$\ell^{j,k_1,k_2} \doteq |\hat{v}^{j,k_1} - \hat{v}^{j,k_2}|, \tag{9}$$

$$\phi^{j,k_1,k_2,k_3} \doteq \cos^{-1} \left( \frac{(\hat{v}^{j,k_1} - \hat{v}^{j,k_2}) \cdot (\hat{v}^{j,k_3} - \hat{v}^{j,k_2})}{(\ell^{j,k_1,k_2})(\ell^{j,k_3,k_2})} \right). \tag{10}$$

We first place a point  $\bar{v}^{j,k}$  in the span of  $v^{j,0} - v^{j,1}$  at a distance of  $|\bar{v}^{j,k} - v^{j,1}| = \ell^{j,k,1}$ . The choice  $\bar{v}^{j,k} = v^{j,1} + \frac{\ell^{j,k,1}}{\ell^{j,0,1}}(v^{j,0} - v^{j,1})$  will work, since

$$|\bar{v}^{j,k} - v^{j,1}| = \left| \frac{\ell^{j,k,1}}{\ell^{j,0,1}}(v^{j,0} - v^{j,1}) \right| = \frac{\ell^{j,k,1}}{\ell^{j,0,1}} |v^{j,0} - v^{j,1}| = \ell^{j,k,1}. \tag{11}$$

Next, a rotation matrix is used to rotate  $\bar{v}^{j,k}$  by the correct angle into its position  $v^{j,k}$ . The rotation matrix is centered at  $v^{j,1}$ , and its axis of rotation is defined by  $u = \frac{1}{\ell^{j,0,1}\ell^{j,2,1}}(v^{j,0} - v^{j,1}) \times (v^{j,2} - v^{j,1})$ . Similarly, the angle of rotation  $\phi^{j,0,1,k}$  is the angle created by the two line segments in the template  $(\hat{v}^{j,0}, \hat{v}^{j,1})$  and  $(\hat{v}^{j,2}, \hat{v}^{j,1})$ . Thus, using  $R = R(\phi^{j,0,1,k}, u)$  our equation for  $v^{j,k}$  is

$$v^{j,k} = v^{1,k} + R(\bar{v}^{j,k} - v^{j,1}) = v^{1,k} + \ell^{j,k,1} R(v^{j,0} - v^{j,1}). \tag{12}$$

Since  $R$  is a polynomial in  $v^{j,0}$ ,  $v^{j,1}$ , and  $v^{j,2}$ , we get the following polynomial 2D face constraint for each  $k > 2$ .

$$c_{2D}^{j,k}(z) = v^{1,k} + \ell^{j,k,1} R(v^{j,0} - v^{j,1}) - v^{j,k} \tag{13}$$

The final constraint type, vertex identification, is used to enforce that the edge between two faces is a hinge. To do this, we simply need to ensure that corresponding vertices on each edge share identical locations. This results in the constraints

$$c_{ident}^{j_1,k_1,j_2,k_2,d}(z) \doteq v_d^{j_1,k_1} - v_d^{j_2,k_2}, \tag{14}$$

where  $v^{j_1,k_1}$  and  $v^{j_2,k_2}$  are corresponding vertices from the faces  $j_1$  and  $j_2$  meeting at a hinged edge. If there are  $|E_x|$  hinges in a state  $x$ , there are  $6|E_x|$  corresponding vertex identification constraints.

We do not have to use the angle and 2D face constraints when all the faces of the polyhedron are triangles (e.g., for the tetrahedron, octahedron, and icosahedron). Further, rather than explicitly using vertex identification constraints, we can either treat them as length constraints with zero length between identified vertices or we can simply reindex the vertices so that identified vertices are actually treated as a single vertex. With either choice, in the triangular case, we may only deal with length constraints if we wish. This property simplifies computations with such polyhedra and is utilized in Sect. 6.

### 4.3 The Geometric Configuration Space

We have shown in the previous subsection that each state  $x$  determines a set of quadratic constraints  $c(z; x)$  (the dependence on the state  $x$  was suppressed in the notation for simplicity). The solution set of the constraint equations is invariant under rigid body motions. Precisely, suppose  $Q \in SO(3)$  is a rotation matrix and  $c \in \mathbb{R}^3$ , let  $Qz + c$  denote the vector in  $\mathbb{R}^{3N_x}$  obtained by the transformation  $v \mapsto Qv + c$  for each vertex contained in  $z$ . Then, the form of the constraint equations in Table 4 implies that  $c(Qz + c; x) = 0$ .

We define the *moduli space*,  $\mathcal{M}_{[x]}$ , to denote the solution set

$$\mathcal{M}_{[x]} = \{z \in \mathbb{R}^{3N_x} \mid c(z; x) = 0\}. \tag{15}$$

The moduli space is well defined since each  $y \in [x]$  is of the form  $y = gx$ , for some  $g \in SO(3)$ . Further, since  $\mathcal{M}_{[x]}$  is the solution set of a family of quadratic equations determined by  $[x]$ , it is an algebraic variety. Each point  $z$  in the moduli space that corresponds to a polyhedral linkage that does not self-intersect is called a *conformation*. Finally, we define the *geometric configuration space*,  $\mathcal{G}$ , to be the set of moduli spaces  $\mathcal{M}_{[x]}$  fibered over configurations  $[x]$  in the graph  $\mathcal{C}$ .

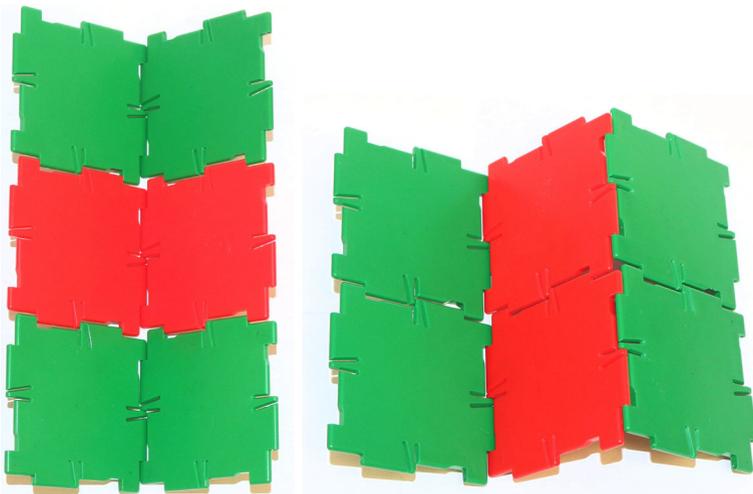
In what follows, we explore some properties of  $\mathcal{G}$  in order to convince the reader that it is the natural setting for the study of properties of configurations that rely on their embedding in space (e.g., energy functions that depend on the embedding of a state  $x$  into  $\mathbb{R}^3$ ).

## 5 Degrees of Freedom

### 5.1 Definitions

Roughly speaking, the degrees of freedom of a linkage are the number of parameters needed to specify its position completely. In more mathematical terms, this should correspond to the dimension of the moduli space. However, this is not entirely correct: The moduli space  $\mathcal{M}_{[x]}$  is an *affine variety*, not a manifold; it may possess singular points, and the degrees of freedom of a linkage may depend on the conformation. That this is not an esoteric mathematical objection is revealed in Fig. 7. Several intriguing linkages with this property have been studied in the mechanics literature (Galletti and Fanghella 2001; Wohlhart 1996).

There are in fact three distinct nonnegative integers that could be viewed as the ‘number of independent coordinates’: the dimension of the variety  $\mathcal{M}_{[x]}$ ; the dimension of the variety  $\mathcal{M}_{[x]}$  at a point  $z \in \mathbb{R}^n$ ; and finally the dimension of the tangent space  $T_z \mathcal{M}_{[x]}$ . In order to explain the subtleties involved, we review some basic algebraic geometry following (Cox et al. 1992). Our algorithm in Sect. 6 does not directly



**Fig. 7** How many degrees of freedom? A linkage of six squares admits two families of embeddings into  $\mathbb{R}^3$ . The embedding on the left has one degree of freedom; the embedding on the right has two degrees of freedom. By generalizing this construction, one sees that a corrugated linkage with  $2n$  squares may have 1 or  $n - 1$  degrees of freedom corresponding to the modes above. We are unaware of a natural notion of Brownian motion of this linkage that allows it to flip between these distinct conformations

use these ideas, but there are inconsistencies in our approach that should be rectified in future work.

We consider polynomials in  $n$  variables  $z_1, \dots, z_n$  with coefficients in a field  $\mathbb{F}$ . Given  $s$  polynomials  $f_1, \dots, f_s$ , the affine variety  $\mathbf{V} = \mathbf{V}(f_1, \dots, f_s)$  is the set of zeros  $\{(z_1, \dots, z_n) \in \mathbb{F}^n \mid f_j(z_1, \dots, z_n) = 0, 1 \leq j \leq s\}$ . The set of all polynomials that vanish on  $\mathbf{V}$  is an ideal, denoted  $I(\mathbf{V})$ . Let  $\langle f_1, \dots, f_s \rangle$  denote the ideal generated by the polynomials  $f_1, \dots, f_s$ . Clearly, all polynomials in  $\langle f_1, \dots, f_s \rangle$  vanish on  $\mathbf{V}$ , so that  $\langle f_1, \dots, f_s \rangle \subset I(\mathbf{V})$ . But, in general, these ideals are *not* the same, though they are related through Hilbert's Nullstellensatz if the field  $\mathbb{F}$  is algebraically closed (Cox et al. 1992, Ch. 4). The algorithmic approach to linkages is complicated by the fact that  $\mathbb{R}$  is not algebraically closed.

In this context, the dimensions of interest are defined as follows.

1. The dimension of a variety  $\mathbf{V}$ , denoted  $\dim \mathbf{V}$ , is the degree of the affine Hilbert polynomial of the corresponding ideal  $I(\mathbf{V})$  (Cox et al. 1992, Defn. 7, p. 430).
2. The dimension of  $\mathbf{V}$  at a point  $p$ , denoted  $\dim_p \mathbf{V}$ , is the maximum dimension of an irreducible component of  $\mathbf{V}$  containing  $p$  (Cox et al. 1992, Defn. 6, p. 460).
3. Finally, the tangent space  $T_p \mathbf{V}$  is itself an affine variety (Cox et al. 1992, Defn. 1, p. 455). It may be defined as a translate of a linear subspace of the space of polynomials in  $n$  variables  $z_1, \dots, z_n \in \mathbb{F}$ , without requiring that the field be  $\mathbb{R}$  or  $\mathbb{C}$ . Since it is an affine variety, its dimension is defined as in (1) and denoted  $\dim T_p \mathbf{V}$ .
4. A point  $p \in \mathbf{V}$  is *non-singular* when  $\dim_p \mathbf{V} = \dim T_p \mathbf{V}$ . If not, the point is *singular*. The set of singular points  $p \in \mathbf{V}$  is called the *singular locus*.

The above definitions rely only on the algebraic properties of the ideal  $I(\mathbf{V})$ . When the field  $\mathbb{F} = \mathbb{R}$ , or  $\mathbb{C}$ , it remains to determine whether the tangent space to the variety, as defined above, agrees with the geometric definition of the tangent space to a real, or complex, manifold. If a point  $p \in \mathbf{V}$  is non-singular, then these two notions agree. But when  $\mathbb{F} = \mathbb{R}$ , it may also happen that a point  $p \in \mathbf{V}$  is singular but that the tangent space (in the sense of manifolds) is well defined.

But how are we to determine if a point is singular or not? As the above discussion makes clear, the fundamental definitions of singular and non-singular points on a real affine variety  $\mathbf{V}(f_1, \dots, f_s)$  rely on purely algebraic notions (e.g., the computation of Hilbert polynomials), rather than the use of differential calculus (e.g., the computation of the rank of the  $s \times n$  Jacobian matrix  $Df$ ). When the field  $\mathbb{F} = \mathbb{C}$ , these computations yield equivalent results: A point  $p \in \mathbf{V}(f_1, \dots, f_s) \subset \mathbb{C}^n$  is non-singular if and only if  $\dim_p \mathbf{V} = n - \text{rank}(Df)$ . But these two computations are not equivalent when  $\mathbb{F} = \mathbb{R}$ . It is this subtlety that complicates the analysis of degrees of freedom in practice.

## 5.2 Computing Degrees of Freedom

We now return to the moduli space  $\mathcal{M}_{[x]}$  and the issue of defining degrees of freedom for a polyhedral linkage. Since the conformations of polyhedral linkages are points on an affine variety, it is natural to define the number of degrees of freedom in accordance with one of the definitions of dimension above. But which one? From a computational

standpoint, if one is using only numerical algorithms, it is easy to compute the rank of the Jacobian of the constraint matrix fast. For this reason, we adopt the following *working*

**Definition 4** The number of **degrees of freedom** of a configuration  $x$  at a non-singular point  $z \in \mathcal{M}_{[x]}$  is defined to be

$$\dim(T_z \mathcal{M}_{[x]}) = n - \text{rank}(C(z)), \tag{16}$$

where  $C(z)$  is the Jacobian matrix of the constraint equations  $c(z; x)$ . If  $z$  is a singular point of  $\mathcal{M}_{[x]}$ , then the number of degrees of freedom is undefined.

As the discussion in the previous subsection has made clear, this definition is incomplete. A complete analysis of the building game must rely on the computation of the ideals, the affine Hilbert polynomial, and singular locus defined by the constraint equations for each configuration  $[x]$ . While these computations may be implemented for a given linkage using Groebner bases, in order to obtain a tractable computation for the whole configuration space, it is also necessary to utilize the inclusion relations between the moduli spaces  $\mathcal{M}_{[x]}$  and  $\mathcal{M}_{[y]}$  for two neighboring configurations  $[x]$  and  $[y]$  in the combinatorial configuration space. We have not pursued this approach in this work. An algebro-geometric analysis of the building game is of fundamental interest and is certain to reveal interesting phenomena.

### 5.3 Results

Assume that a configuration  $[x] \in \mathcal{C}$  is fixed. For brevity, let  $n = 3N_x$  denote the number of coordinates of vertices, and let  $m$  denote the number of constraint equations listed in Table 4. The explicit form of the Jacobian is easily computed since  $C(z)$  is a set of quadratic equations. Formula 16 is then numerically implemented as follows. First, a solution  $z$  to the constraint equation is computed numerically. Next, we compute the SVD of the Jacobian matrix  $C(z)$ . The rank is the number of nonzero singular values.

We implicitly distinguish between the six parameters needed to prescribe the rigid body modes of a linkage and the (nontrivial) internal degrees of freedom. More formally, in line with the definition above, for each  $x \in \mathcal{C}$  it is necessary to consider the quotient space  $\mathcal{M}_{[x]}/E(3)$  where  $E(3)$  denotes the Euclidean group, and the dimension of the associated tangent space at a point  $[z] \in \mathcal{M}_{[x]}/E(3)$ . In our calculations, we have not implemented this idea formally (see [Holmes-Cerfon et al. 2013, SI](#) for a related calculation).

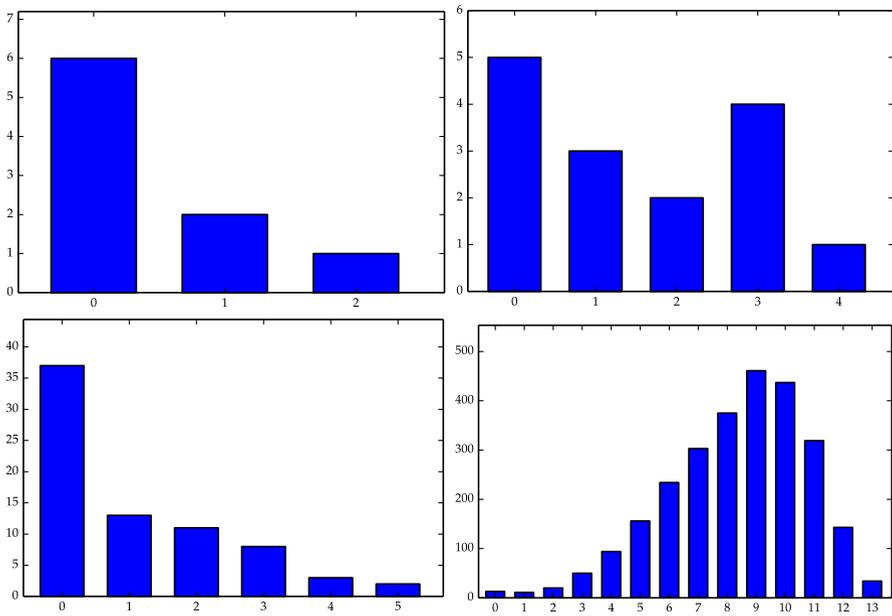
We have used this method to compute the number of degrees of freedom for all configurations of the Platonic solids at the standard embedding. In order to test that this embedding is non-singular, we check that the rank of the Jacobian does not change under small random perturbations. These results are presented in Fig. 8. We do not include the tetrahedron, since it only has four intermediates and only the one composed of two triangles is not rigid. Interestingly, the

cube and dodecahedron have a large number of rigid intermediates due to the fact that the linkage consisting of three faces (squares or pentagons) meeting at the dihedral angle at a vertex is rigid. A similar linkage consisting of four or five triangles is *not rigid*. As a consequence, the octahedron, with four triangles meeting at each vertex, has a more varied distribution of internal degrees of freedom. Moreover, the icosahedron has a very small number of rigid intermediates, and some intermediates have as many as 13 internal degrees of freedom.

We carried out the above computations in order to more carefully investigate the role of rigidity of intermediates in self-assembly. In prior work on surface tension-driven self-folding, we noted that rigidity plays an important role in the formation of Platonic and Archimedean polyhedra. In these experiments, the cube, dodecahedron, and truncated octahedron could be assembled with high yield, but icosahedra could not be formed (Pandey et al. 2011). Further, successful self-assembly was observed to proceed through a few dominant intermediates, with fewer degrees of freedom than comparable intermediates with the same number of glued edges. Why could some polyhedra be formed successfully and not others? The main distinction lies in the rigidity of corners—the polyhedral linkage consisting of three triangles, three squares, or three pentagons, meeting at the dihedral angles of the tetrahedron, cube, and dodecahedron, respectively, is rigid. In contrast, the corners of the octahedron and the icosahedron are not rigid. The dominant intermediates for the cube, dodecahedron, and truncated octahedron were composed of two or three rigid sub-linkages, whereas most other intermediates had long, floppy segments. These (informal) ideas of rigidity were used to develop a model experiment on synthetic isomers, analogous to the transformations between the ‘chair’ and ‘boat’ conformations of cyclohexane (Pandey et al. 2014).

The idea of rigidity and mobility of intermediates also appears to explain a different set of model experiments in DNA nanotechnology (Bhatia et al. 2009 and references therein). The Platonic solids were synthesized, using short, stiff DNA segments to make their edges. Unlike self-folding experiments, where the basic geometric motif is a rigid face, in DNA-based experiments, the basic geometric motif is a rigid edge. In Bhatia et al. (2009), octahedra and icosahedra could be synthesized with high yield, but not dodecahedra. Again, the rigidity of linkages suggests an explanation for these observations – the triangular linkage consisting of three rigid rods joined at hinges is rigid, but linkages consisting of more than three rods are not rigid.

In light of these experiments, the computational results in Fig. 8 are very intriguing, albeit with some qualification. The self-assembly techniques of both Pandey et al. (2011) and Bhatia et al. (2009) are modeled by different combinatorial configuration spaces, so the rigidity statistics in Fig. 8 do not apply directly to these experiments. On the other hand, there are too many unknown effects in the assembly of fullerenes and virus capsids and too little data on the intermediate pathways, to directly connect rigidity statistics such as those above with the actual process of assembly. We are unaware of *synthetic* mesoscale experiments that can be modeled by the building game that would allow a more direct interpretation of the results above in an experimental context.



**Fig. 8** Distribution of internal degrees of freedom among intermediates computed at the standard embedding. Clockwise from *top left*: cube, octahedron, icosahedron, and dodecahedron. The number of degrees of freedom is plotted on the *x*-axis and the number of intermediates on the *y*-axis

## 6 Conformational Diffusion

### 6.1 Brownian Motion on the Moduli Space $\mathcal{M}_{[x]}$

Though our work is primarily numerical, we first summarize the rigorous results that underlie our approach to conformational diffusion.

Assume that  $(\mathcal{M}_d, g)$  is a complete,  $C^\infty$  Riemannian manifold of dimension  $d$  with  $C^\infty$  metric  $g$ . As usual, let  $\Gamma_{ij}^k$  denote the Christoffel symbols of the metric. The Wiener process on  $\mathcal{M}$  is the diffusion process whose infinitesimal generator is  $\Delta/2$ , where the Laplace–Beltrami operator

$$\Delta = g_{ij} \frac{\partial^2}{\partial x_i \partial x_j} - g_{ij} \Gamma_{ij}^k \frac{\partial}{\partial x_k}. \tag{17}$$

There are several approaches to Brownian motion on manifolds, and a useful general reference is [Hsu \(1988\)](#). Our interest lies in a construction of the Wiener process that is amenable to direct numerical simulation. For this reason, we focus on a construction of Brownian motion as a limit of random walks. Basic rigorous results in this direction were established by [Jørgensen \(1975\)](#).

A random walk on  $\mathcal{M}$  is defined as follows. For each point,  $p \in \mathcal{M}$ , tangent vector  $v \in T_p \mathcal{M}$ , and time parameter  $\tau \in \mathbb{R}$ , let  $p(\tau) = \exp_p(\tau v)$  denote the (unit speed) geodesic through  $p_0$  in the direction  $v$ . In order to construct a random walk

$\{p_0, p_1, \dots, p_k, \dots\}$ , we assume given an initial condition  $p_0$ , a step size  $\tau > 0$ , and a family of probability measures  $\{\nu_p\}_{p \in \mathcal{M}}$  on  $T_p\mathcal{M}$ . Then given  $p_k, k = 0, 1, \dots$ , we define the geodesic segment on the time interval  $t \in [0, \tau]$ ,

$$p_{k+1}(t) = \exp_{p_k} \left( \frac{t}{\tau} (\tau s(p_k) + \sqrt{\tau} (V_k - s(p_k))) \right), \quad s(p) = \int_{T_p\mathcal{M}} V d\nu_p(V). \tag{18}$$

(We abuse notation by identifying  $p_k = p_{k+1}(0)$ ). The sequence  $\{p_k\}$  is a random walk taking values in  $\mathcal{M}$  that is determined by the metric  $g$  and the probability measures  $\nu_p$ . By interpolating in time as above, we also obtain a continuous path  $p^{(\tau)} : [0, 1] \rightarrow \mathcal{M}$  with  $p^{(\tau)}(k\tau) = p_k$ . This construction thus provides a discrete measure on the path space  $C([0, 1], \mathcal{M})$ . As in Donsker’s theorem, Jørgensen established the weak convergence of these measures to a limiting diffusion process, under certain assumptions on the measures  $\nu_p$  (Jørgensen 1975, §2). When  $\nu_p$  is the uniform measure on the unit sphere in  $(T_p\mathcal{M}, g)$ , the limiting diffusion is the Wiener process on  $(\mathcal{M}_d, g)$ .

In applying these ideas to the moduli space  $\mathcal{M}_{[x]}$ , certain caveats are necessary. First, to the best of our knowledge, there appear to be no systematic treatments of Brownian motion of linkages. From the physical point of view, the main issue here is that a natural notion of Brownian motion on moduli space should allow the linkage to explore distinct vibratory modes. For instance, a naive simulation of random walk for the linkage shown in Fig. 7 based on the scheme above will leave the linkage trapped in the random motion of a one-dimensional hinge (left, Fig. 7) or two-dimensional hinge (right, Fig. 7) with no possibility of transfer between these modes. It is of course possible to resolve these issues by ‘softening’ the constraints, but this is not an adequate mathematical resolution of the issue. Further, there seems to be no hope of establishing the geodesic completeness of  $\mathcal{M}_{[x]}$  as is required for a rigorous convergence proof à la Jørgensen.

Nevertheless, the theory of Brownian motion on manifolds does provide some support for our approach. For example, it immediately follows from the implicit function theorem that if  $z \in \mathcal{M}_{[x]}$  is a non-singular point there is a ball  $B_\epsilon(z) \subset \mathbb{R}^n$  with positive radius  $\epsilon > 0$  such that  $\mathcal{M}_{[x]} \cap B_\epsilon(z)$  is an analytic manifold (recall that  $n = 3N_x$  is the dimension of the space of coordinates and  $m$  denotes the number of constraints). Thus, at least locally, the notion of a Wiener process on  $\mathcal{M}_{[x]}$  is well founded and amenable to numerical experiment. In order to apply the above construction, all that is needed is a sufficiently smooth metric  $g$ . In all our numerical experiments, we simply use the uniform metric on the tangent space  $T_z\mathcal{M}_{[x]}$ . The methods presented below may be easily amended to include other metrics and general stochastic processes on  $\mathcal{M}_{[x]}$ .

### 6.2 Numerical Scheme for a Random Walk

In order to construct a random walk  $\{z_0, z_1, \dots, z_k, \dots\}$  on  $\mathcal{M}_{[x]}$  according to the above prescription, it is necessary to solve for geodesics approximately. We use a naive first-order scheme: At each point  $z_k$ , we pick a tangent vector  $v_k \in T_{z_k}\mathcal{M}_{[x]}$  at

random, shoot along the tangent space in the direction  $v_k$  for time  $\tau$ , and then project the point  $z_k + \sqrt{\tau}v_k \in \mathbb{R}^n$  orthogonally onto  $\mathcal{M}_{[x]}$ . This scheme is implemented in two steps. First, we compute the tangent space. Next, we compute the projection back to the manifold. Finally, in order to prevent self-intersection, we always begin the random walk in a conformation without self-intersection. The random walk is reflected at the boundaries corresponding to self-intersecting conformations using a rejection scheme.

At each point  $z \in \mathcal{M}_{[x]}$ , the tangent space  $\mathcal{T}_z\mathcal{M}_{[x]}$  is naturally identified with the nullspace  $\{v \in \mathbb{R}^n : C(z)v = 0\}$  where  $C : \mathbb{R}^n \rightarrow \mathbb{R}^{(n-m) \times n}$  is the Jacobian of  $c$  at  $z$ . Thus, sampling from the tangent space is the same as sampling from the null space of  $C(z)$ . We construct a basis for  $\mathcal{T}_z\mathcal{M}_{[x]}$  as follows. Let  $A = [C^T(z) B]$  be the concatenation of  $C^T(z)$  and a random matrix  $B \in \mathbb{R}^{n \times m}$  with iid  $\text{unif}(0, 1)$  entries  $B_{jk}$ . The square matrix  $A$  is of full rank with probability 1. We compute the QR decomposition

$$A = [C^T \ B] = QR = [Q^{(1)} \ Q^{(2)}]R. \tag{19}$$

The columns of  $Q^{(2)}$ , written  $Q_i^{(2)}$ ,  $i = 1, \dots, m$ , form an orthonormal basis for  $\mathcal{T}_z\mathcal{M}_{[x]}$ . By contrast, the columns of  $Q^{(1)}$  are an orthonormal basis for  $\mathcal{T}_z\mathcal{M}_{[x]}^\perp$ . The coordinates  $\alpha \in \mathbb{R}^m$  of a vector  $v \in \mathcal{T}_z\mathcal{M}_{[x]}$  are defined by  $v = Q^{(2)}\alpha$ . The metric  $g$  can be expressed in these coordinates as follows.

$$g(v, v) = \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j g(Q_i^{(2)}, Q_j^{(2)}) := \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j G_{ij}. \tag{20}$$

Our task is to sample uniformly from  $\Omega_G = \{v = Q^{(2)}\alpha : \alpha^T G \alpha = 1\}$ . Since this set corresponds to a level set of the multivariate normal distribution with mean zero and covariance matrix  $G^{-1}$ , our sampling problem reduces to sampling  $u \sim \mathcal{N}(0, G^{-1})$ . Once  $u$  is obtained, we set  $v = \frac{u}{|u|}$ . We generally choose  $G$  to be the identity, but this is not necessary.

In order to project the point  $z + \sqrt{\tau}v$  back to  $\mathcal{M}_{[x]}$ , we seek a vector  $v^\perp \in \mathcal{T}_z\mathcal{M}_{[x]}^\perp$  such that  $c(z + \sqrt{\tau}v + v^\perp) = 0$ . Since  $Q^{(1)}$  provides a basis for  $\mathcal{T}_z\mathcal{M}_{[x]}^\perp$ , we can write  $v^\perp = Q^{(1)}w$  for some  $w \in \mathbb{R}^{n-m}$  and solve  $c(z + \sqrt{\tau}v + Q^{(1)}w) = 0$ . We define an objective function  $F$  and its Jacobian  $\mathcal{J}$ ,

$$F(w) = c(z + \sqrt{\tau}v + Q^{(1)}w), \quad \mathcal{J}(w) = C(z + \sqrt{\tau}v + Q^{(1)}w)Q^{(1)}, \tag{21}$$

and use the following Newton–Raphson iteration to solve for  $w$

$$\mathcal{J}(w_k)(w_{k+1} - w_k) = -F(w_k). \tag{22}$$

The usual initial condition for the iteration is  $w_0 = 0$ .

Naive sampling from the moduli space yields self-intersecting conformations. We avoid these conformations using reflected Brownian motion. To sample from a

reflected Brownian motion on the manifold, we use a rejection scheme that first generates a proposal conformation according to the unbiased random walk scheme. We use an implicit boundary function that returns ‘true’ if a proposed configuration is valid (not self-intersecting) and ‘false’ otherwise. This process is repeated until a ‘true’ proposal is accepted—more precisely, if the system is at a position  $z$ , the system stays at  $z$ , until a valid proposed configuration, say  $\tilde{z}$ , is found. When recording the discrete steps of a trajectory, we do *not* record it as  $z, z, \dots, z, \tilde{z}$ —repeating the current configuration  $z$  until an admissible configuration  $\tilde{z}$  is found. Instead, the discrete trajectory is recorded as  $z, \tilde{z}$ . Testing for self-intersection is expensive: If there are  $k$  faces in a linkage, testing for self-intersecting requires  $\binom{k}{2}$  pairwise comparisons. It is imperative that these comparisons are carried out as efficiently as possible. We use an efficient algorithm presented by Möller (1997).

Finally, we point out that our scheme is inefficient in the sense that we do not explicitly remove the rigid body modes. Ad hoc schemes for fixing these modes lead to inconsistent results (Johnson 2015, Chapter 6). At present, our method includes diffusion of both the center of mass and the orientation of the polyhedral linkage. However, these modes do not affect the internal modes, such as those shown in Fig. 11. This issue may be avoided by using the quotient metric for  $g$  as in Holmes-Cerfon et al. (2013).

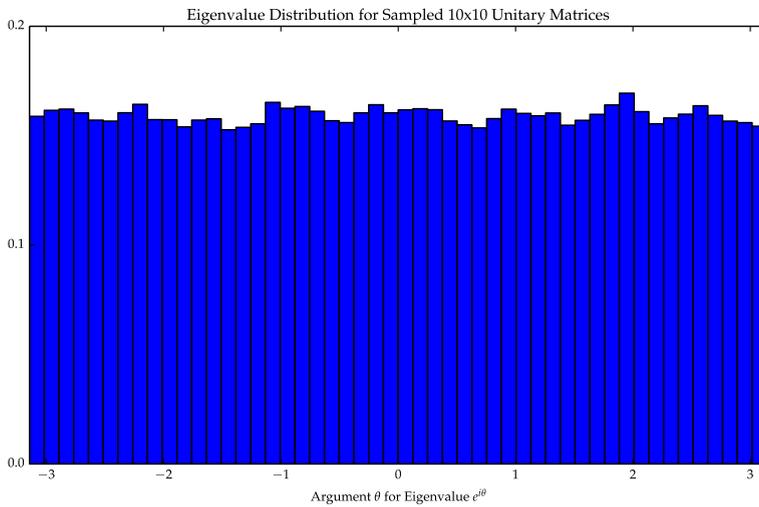
## 6.3 Test Cases

### 6.3.1 Unitary Matrices

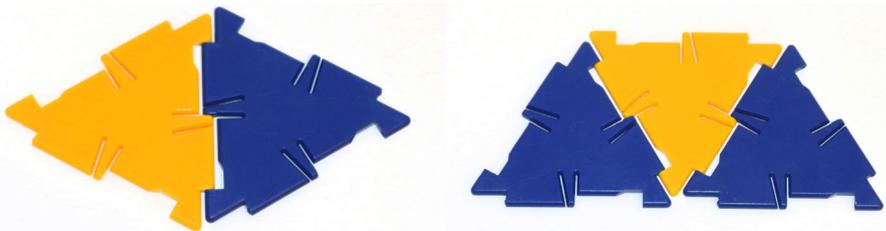
A simple consistency test for the numerical scheme is provided by Brownian motion on  $U(N)$  the group of  $N \times N$  unitary matrices. Recall that each matrix  $U \in U(N)$  satisfies the quadratic equation  $U^*U = UU^* = I$ . This matrix equation is equivalent to a system of  $N + N^2$  real quadratic polynomial constraints. As a test of our method, we simulate a random walk,  $U_0, U_1, \dots, U_k, \dots$  on  $U(N)$ , and compute the eigenvalues of the matrices  $U_k$ . The statistics of these eigenvalues are known (the joint distribution of eigenvalues is given by Weyl’s integration formula, and the 1-point distribution of eigenvalues is uniform on  $S^1$ ). Sample numerical results are shown in Fig. 9.

### 6.3.2 2- and 3-Triangle Linkages

As a more direct test of our scheme, we consider the simple linkages consisting of two and three triangles. As seen in Fig. 10, the 2-triangle linkage can be specified by the three-dimensional locations of four vertices, and lives in the ambient space  $\mathbb{R}^{12}$ . It has one internal degree of freedom corresponding to the dihedral angle at the hinged edge. Similarly, the 3-triangle linkage is parameterized by five vertices. It has two internal degrees of freedom and lives in the ambient space  $\mathbb{R}^{15}$ . Figure 11 shows histograms of the dihedral angles of these linkages as sampled using our random walk scheme. Interestingly, the distribution is not uniform (note that while we assume that the metric  $g$  is the identity in all our simulations, this does not imply that the pushforward of the equilibrium measure onto the dihedral angle must be uniform).



**Fig. 9** Empirical distribution of unitary matrices sampled by a random walk with 3,000,000 steps, each time step of size  $\Delta t = 0.05$  in  $U(10)$ . Compare with the sampling schemes for Haar measure in Mezzadri (2007) and Li and Menon (2013)

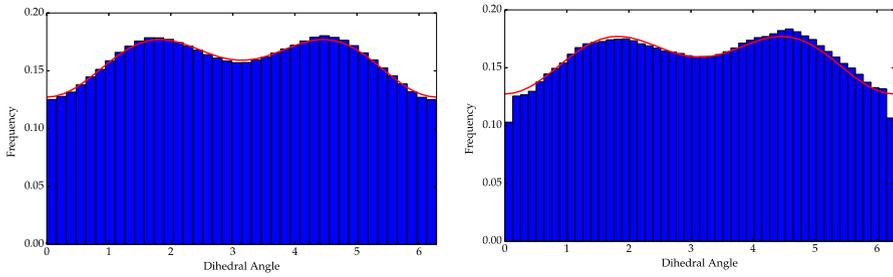


**Fig. 10** Two and three triangle linkages

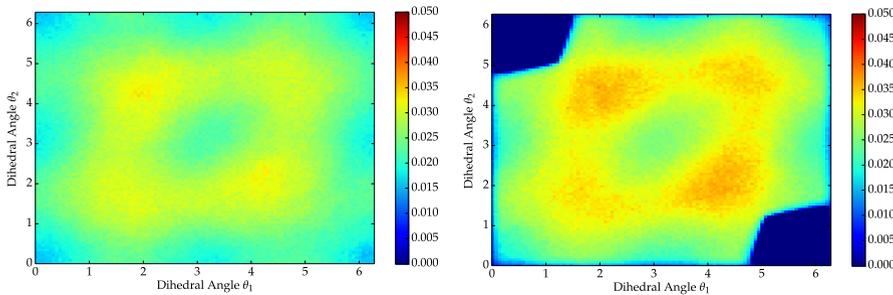
The constraint preventing self-intersection modifies the empirical distribution most at the endpoints. Similarly, we consider the 3-triangle linkage in Fig. 12. The empirical distributions agree except in the lower right and upper left corners of the plot. These regions are forbidden, since they correspond to self-intersecting conformations. A noticeable boundary layer of lower probability is seen around the periphery of the plot.

As a test of convergence for our scheme, Fig. 13 shows the Kolmogorov–Smirnov distance of the empirical distribution on the first dihedral angle of the three triangle linkage as a function of the number of samples. More precisely, let  $\{z_0, z_1, \dots, z_{n-1}\}$  be the first  $n$  steps of a random walk with  $N$  steps ( $n \leq N$ ), let  $\theta_1(z_k)$  denote the first dihedral angle of the conformation  $z_j$ ;  $\mathbf{1}_{[0, \theta]}$  denotes the indicator function for the interval  $[0, \theta]$  for any  $\theta \in [0, 2\pi]$  and defines the empirical distribution at step  $n$ ,

$$F_n(\theta) = \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{[0, \theta]}(\theta_1(z_k)). \tag{23}$$



**Fig. 11** Histogram of the dihedral angle in sampled conformations of the 2-triangle linkage. Free rotation allowing self-intersection (*left*) and no self-intersection (*right*). The *red curve* is the least squares fit to a curve of the form  $a_0 + a_1 \cos \theta + a_2 \cos 2\theta$ , where  $\theta$  denotes the dihedral angle (Color figure online)



**Fig. 12** Ramachandran plots of the two dihedral angles in the 3-triangle linkage. Free rotation allowing self-intersection (*left*) and no self-intersection (*right*). The sampled random walk on the right respects the constraint of non-intersection, as seen in the *blue zones* which are forbidden. However, a faint boundary layer is perceptible (Color figure online)

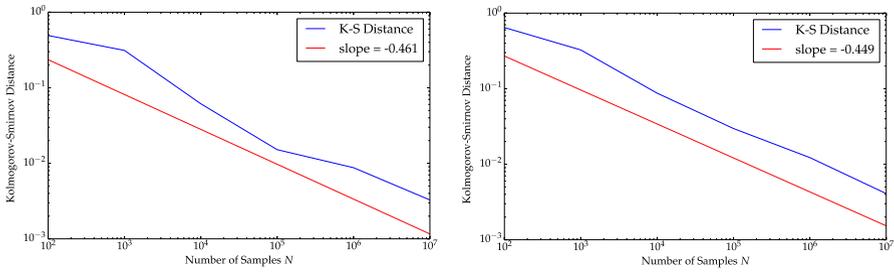
We empirically study the rate of convergence of the scheme with the Kolmogorov–Smirnov statistic

$$D_{n,N} = \sup_{\theta \in [0, 2\pi]} |F_n(\theta) - F_N(\theta)|. \tag{24}$$

The line of best fit has a slope of  $-0.461$  and  $-0.449$ , respectively. This is nearly  $-0.5$  indicating that the (naively) expected square root rate of convergence is achieved.

### 7 From Conformational Diffusion to Transition Rates

The mathematical framework of the building game allows us to separate the kinetics and geometric combinatorics of the growth process. The geometric combinatorics of the growth process is described completely by the combinatorial configuration space,  $\mathcal{C}$ . The kinetics of growth are described by a Markov process on  $\mathcal{C}$ . In order to prescribe a Markov process, we must prescribe the rate of transition  $Q_{[x][y]}$  from a configuration  $[x]$  to a configuration  $[y]$ . The transition matrix  $Q$  is of size  $|\mathcal{C}| \times |\mathcal{C}|$  and has nonnegative entries on the off-diagonal. Further, the only off-diagonal terms



**Fig. 13** Empirical convergence of the scheme. Convergence is measured with the Kolmogorov–Smirnov distance defined in Eqs. (23)–(24) for the 3-triangle linkages. We fix  $N = 10^7$  and use  $n = 10^3, 10^3, 10^4, 10^5$  and  $10^6$ . The figures show convergence for free rotation allowing self-intersection (left) and no self-intersection (right)

that are nonzero are those that correspond to neighbors. The terms on the diagonal are determined by the condition that the row sum vanishes: i.e.,  $\sum_{[y] \neq [x]} Q_{[x][y]} = 0$ .

In several biophysical applications, it is conventional to assume that the system is described by equilibrium statistical mechanics. We assume the existence of a discrete energy landscape  $E : \mathcal{C} \rightarrow \mathbb{R}$  and a thermal bath at inverse temperature  $\beta > 0$  and postulate that the equilibrium measure of the system is the Gibbs distribution

$$\pi([x]) = \frac{e^{-\beta E([x])}}{Z_\beta}, \quad Z_\beta = \sum_{[x] \in \mathcal{C}} e^{-\beta E([x])}. \tag{25}$$

For the building game, a natural choice is to define  $E([x])$  to be minus the number of edges between the faces in  $[x]$ , since each edge between faces can be viewed as a bond with unit energy. We further assume that the microscopic interactions are reversible, so that the only admissible transition matrices are those that satisfy the condition of detailed balance

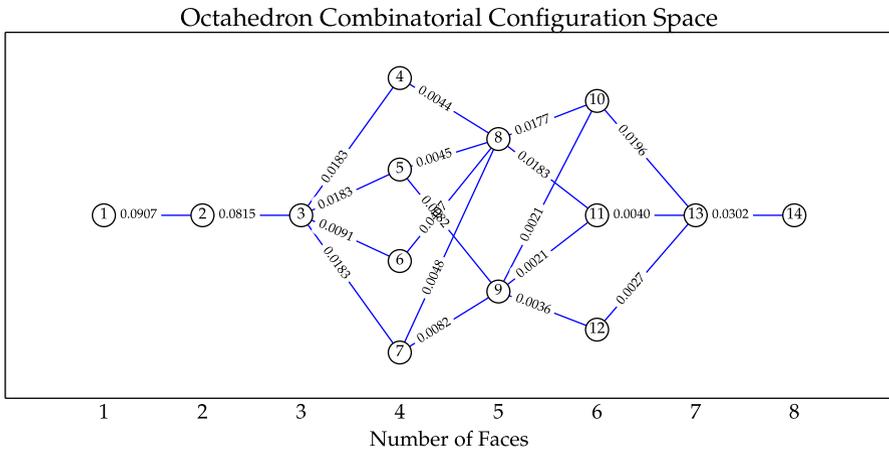
$$\pi([x]) Q_{[x][y]} = \pi([y]) Q_{[y][x]}. \tag{26}$$

These restrictions still allow many possible transition matrices. The conventional assumption is that these must be of the form

$$Q_{[x][y]} = S_{[x][y]} e^{-\beta(B([x],[y]) - E([y]))}, \tag{27}$$

where the function  $B : \mathcal{C} \times \mathcal{C} \rightarrow \mathbb{R}$  describes energy barriers between configurations, and  $S_{[x][y]} = S_{xy}$  is the degeneracy number defined in Sect. 2.

In our work, we view the energy function, equilibrium measure, and detailed balance as fundamental. Thus, any transition matrix  $Q$  must respect Eqs. (25) and (26). However, we do not introduce the notion of an energy barrier, and we focus instead on the computation of rates  $Q_{[x][y]}$  using conformational diffusion. (Of course, these results could then be used to fit energy barriers to the system, but that is unnecessary—all that the model must predict are rates of transition.)



**Fig. 14** Rates of transition for the octahedron. A rate matrix computed by conformational diffusion in the geometric configuration space,  $\mathcal{G}$  for the octahedron. The state is numbered as in Fig. 1. The phenomenological parameters in this example are  $\beta = 0.8$ ,  $\epsilon = 0.5$

The rates are determined as follows. We consider two neighboring configuration  $[x]$  and  $[y]$  such that  $[y]$  is obtained from  $[x]$  by the attachment of a face. We then simulate a conformational diffusion on  $\mathcal{M}_{[x]}$  and associate a probability of probability of transition from  $[x]$  to  $[y]$  based on the fraction of total time spent in a conformation where the dihedral angles of conformations in  $\mathcal{M}_{[x]}$  are favorably aligned for attachment of a face that corresponds to a conformation in  $\mathcal{M}_{[y]}$ . Observe that it is enough to determine the ‘attachment’ rates  $Q_{[x][y]}$  because the reverse ‘detachment’ rate  $Q_{[y][x]}$  is given by detailed balance.

To illustrate ideas, and for simplicity of implementation, we focus on the octahedron. Given a tolerance  $\epsilon$  and a conformation  $z \in \mathcal{M}_{[x]}$ , our criterion for attachment is as follows:

1. We consider all triplets of vertices in a state  $x \in [x]$  at which it is combinatorially admissible for a new face to attach to form a state  $y \in [y]$ .
2. For each such triplet, let  $v_a, v_b, v_c$  denote the coordinates of these vertices in  $\mathbb{R}^3$ . We compute the angles in the triangle formed by the vertices  $v_a, v_b$  and  $v_c$ . If each of these angles differs from  $\pi/3$  by less than  $\epsilon$ , we say that the triplet  $(v_a, v_b, v_c)$  is  $\epsilon$ -close to attachment.
3. We say that  $z$  lies in the  $\epsilon$ -exit set, denoted  $S_{[x][y]}(\epsilon)$ , if at least one of the admissible triplets of vertices is  $\epsilon$ -close to attachment.

Given the above exit criterion, the empirical transition rate  $Q_{[x][y]}$  may be obtained from the first  $N$  steps  $z_0, z_1, \dots, z_{N-1}$  of a random walk in  $\mathcal{M}_{[x]}$  as follows

$$Q_{[x],[y]}(N, \epsilon) = \frac{1}{N} \sum_{k=0}^{N-1} \mathbf{1}_{S_{[x][y]}(\epsilon)}(z_k). \tag{28}$$

In practice, we simulate and store the random walk for a large number of steps ( $10^7$ ). The rates may then be extracted for any  $\epsilon > 0$ . An example of rates computed by this procedure is shown in Fig. 14. We have not attempted to connect these computations with any physical experiment on self-assembly: They are simply chosen to illustrate the underlying mathematical structure. Various related calculations may be found in Johnson (2015).

## 8 Conclusions

In summary, this article contains a linked set of mathematical ideas that arise in the study of the self-assembly of closed-shell structures. There are four distinct aspects:

1. The combinatorics of growth-by-attachment and the combinatorial configuration space  $\mathcal{C}$ .
2. The kinematics of polyhedral linkages on the moduli spaces  $\mathcal{M}_{[x]}$ , for each  $[x] \in \mathcal{C}$ , and the geometric configuration space,  $\mathcal{G}$ .
3. A pure jump Markov process on the graph  $\mathcal{C}$  that models the kinetics of assembly pathways.
4. A continuous-time diffusion on each  $\mathcal{M}_{[x]}$  exists problems for which determine the rates of the jump process.

Informally, (4) may be thought of as a ‘blow-up’ of (3). That is, while (3) is a well-formulated problem if one knows the rates of transition between states, (4) addresses the problem of determining rates based only on the kinematics of polyhedral linkages. Mathematically, (4) is the study of diffusion on an affine variety; physically, it corresponds to perhaps the simplest microscopic model which can be used to compute rates of transition between configurations.

At present, these ideas should be viewed as a consistent framework, rather than a detailed investigation of the self-assembly of a particular polyhedron. In particular, while our results on enumeration are the most extensive to date, the work on kinematics is restricted to Platonic solids. We find simple, but interesting, results on degrees of freedom of intermediates for the dodecahedron and icosahedron (Fig. 8). The implementation of the complete framework (1)–(4) above is restricted to the octahedron, whose combinatorial configuration space has only 14 configurations. A full computational study for larger polyhedra requires a careful analysis of conformational diffusion on each algebraic variety in  $\mathcal{G}$  and may require new computational techniques because of the combinatorial explosion in the size of  $\mathcal{C}$ .

This work illustrates that conformational diffusion arises naturally in discrete geometric models of self-assembly. However, we are unaware of rigorous mathematical formulations of Brownian motion on affine varieties. In order to formulate these notions precisely, it is important to pay careful attention to the interplay between the singular locus of the variety and the diffusion coefficients (the choice of metric in our work). As illustrated by the linkage in Fig. 7, in order to obtain a Brownian motion on  $\mathcal{M}_{[x]}$  that is ergodic and explores all irreducible components of the variety  $\mathcal{M}_{[x]}$ , it is necessary to carefully resolve the Brownian motion in the neighborhood of the singular locus, in order that it passes from one irreducible component to another. This issue is of independent interest – even well-known examples of conformational changes in isomers,

such as flips between chair and boat conformations for idealized geometric models of cyclohexane, seem to yield interesting new questions on Brownian motion of linkages when one considers ‘hard’ constraint models for these isomers (Baker 1986; Pandey et al. 2014).

**Acknowledgments** The authors are grateful to the referees of this article and the related paper Russell and Menon (2016) for their detailed, constructive criticism.

## References

- Baker, J.E.: Limiting positions of a Bricard linkage and their possible relevance to the cyclohexane molecule. *Mech. Mach. Theory* **21**, 253–260 (1986)
- Bhatia, D., Mehtab, S., Krishnan, R., Indi, S.S., Basu, A., Krishnan, Y.: Icosahedral DNA nanocapsules by modular assembly. *Angew. Chem. Int. Ed.* **48**, 4134–4137 (2009)
- Caspar, D., Klug, A.: Physical principles in the construction of regular viruses. In: *Cold Spring Harbor Symposium in Quantitative Biology*, vol. 27, pp. 1–27 (1962)
- Cox, D., Little, J., O’Shea, D.: Ideals, varieties, and algorithms. In: *Undergraduate Texts in Mathematics*. Springer, New York (1992). doi:10.1007/978-1-4757-2181-2
- Crick, F.H., Watson, J.D.: Structure of small viruses. *Nature*, **177**, 473–475 (1956)
- Douglas, S.M., Dietz, H., Liedl, T., Högberg, B., Graf, F., Shih, W.M.: Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009)
- Endres, D., Miyahara, M., Moisan, P., Zlotnick, A.: A reaction landscape identifies the intermediates critical for self-assembly of virus capsids and other polyhedral structures. *Protein Sci.* **14**, 1518–1525 (2005)
- Galletti, C., Fanghella, P.: Single-loop kinematotropic mechanisms. *Mech. Mach. Theory* **36**, 743–761 (2001)
- Holmes-Cerfon, M., Gortler, S., Brenner, M.: A geometrical approach to computing free-energy landscapes from short-ranged potentials. *Proc. Natl. Acad. Sci.* **110**, E5–E14 (2013)
- Hsu, P.: Brownian motion and Riemannian geometry. *Contemp. Math.* **73**, 95–104 (1988)
- Johnson, D.C.: *Combinatorial and Geometric Structure in the Self-Assembly of Polyhedra*, PhD thesis, Brown University (May 2015)
- Johnson, D.C., Menon, G.: Supporting Data: The Building Game: From Enumerative Combinatorics to Conformational Diffusion (2016). <https://repository.library.brown.edu/studio/item/bdr:589790/>
- Jørgensen, E.: The central limit problem for geodesic random walks. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32**, 1–64 (1975)
- Kaplan, R., Klobusicky, J., Pandey, S., Gracias, D.H., Menon, G.: Building polyhedra by self-folding: theory and experiment. *Artif. Life* **20**, 409–439 (2014)
- Kroto, H., Heath, J., O’Brien, S., Curl, R., Smalley, R.:  $C_{60}$ : Buckminsterfullerene. *Nature* **318**, 162–163 (1985)
- Li, D., Zhou, W., Landskron, K., Sato, S., Kiely, C., Fujita, M., Liu, T.: Viral-capsid-type vesicle-like structures assembled from  $m_{12}l_{24}$  metalorganic hybrid nanocages. *Angew. Chem. Int. Ed.* **50**, 5182–5187 (2011)
- Li, X.H., Menon, G.: Numerical solution of dyson brownian motion and a sampling scheme for invariant matrix ensembles. *J. Stat. Phys.* **153**, 801–812 (2013)
- Liu, Y., Hu, C., Comotti, A., Ward, M.: Supramolecular Archimedean cages assembled with 72 hydrogen bonds. *Science* **333**, 436–440 (2011)
- Mezzadri, F.: How to generate random matrices from the classical compact groups. *Notices Amer. Math. Soc.* **54**, 592–604 (2007)
- Möller, T.: A fast triangle–triangle intersection test. *J. Graph. Tools* **2**, 25–30 (1997)
- Pandey, S., Ewing, M., Kunas, A., Ngyen, S., Gracias, D., Menon, G.: Algorithmic design of self-folding polyhedra. *Proc. Natl. Acad. Sci.* **108**, 19885–19890 (2011)
- Pandey, S., Johnson, D., Kaplan, R., Klobusicky, J., Menon, G., Gracias, D.H.: Self-assembly of mesoscale isomers: the role of pathways and degrees of freedom. *PLoS One* **9**, e108960 (2014)
- Rotman, J.: *An Introduction to the Theory of Groups*, 4th edn. Springer, New York (1995)
- Rotman, J.J.: *An Introduction to the Theory of Groups*, Volume 148 of Graduate Texts in Mathematic, 4th edn. Springer, New York (1995)

- Rudin, M.E.: An unshellable triangulation of a tetrahedron. *Bull. Am. Math. Soc.* **64**, 90–91 (1958)
- Russell, E.R., Menon, G.: Energy landscapes for the self-assembly of supramolecular polyhedra. *J. Nonlinear Sci.* (2016). doi:[10.1007/s00332-016-9286-9](https://doi.org/10.1007/s00332-016-9286-9)
- Sloane, N.: The On-line Encyclopedia of Integer Sequences. [arXiv:math.CO/0312448](https://arxiv.org/abs/math.CO/0312448) (2003)
- Sun, Q., Iwasa, J., Ogawa, D., Ishido, Y., Sato, S., Ozeki, T., Sei, Y., Yamaguchi, K., Fujita, M.: Self-assembled  $M_{24}L_{48}$  polyhedra and their sharp structural switch upon subtle ligand variation. *Science* **328**, 1144–1147 (2010)
- Thurston, W.P., Weeks, J.R.: The mathematics of three-dimensional manifolds. *Sci. Am.* **265**, 108–121 (1984)
- Wales, D.: Closed-shell structures and the building game. *Chem. Phys. Lett.* **141**, 478–484 (1987)
- Whiteley, W.: How to describe or design a polyhedron. *J. Intell. Robot. Syst.* **11**, 135–160 (1994)
- Wohlhart, K.: Kinematotropic linkages. In: Lenarcic, J., Parenti-Castelli, V. (eds.) *Recent Advances in Robot Kinematics*, pp. 359–368. Springer, Berlin (1996)
- Zlotnick, A.: An equilibrium model of the self assembly of polyhedral protein complexes. *J. Mol. Biol.* **241**, 59–67 (1994)