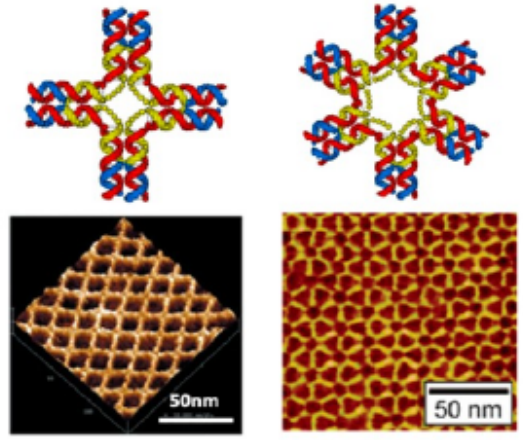
Mathematical models for describing molecular self-assembly

Margherita Maria Ferrari, and Nataša Jonoska
University of South Florida

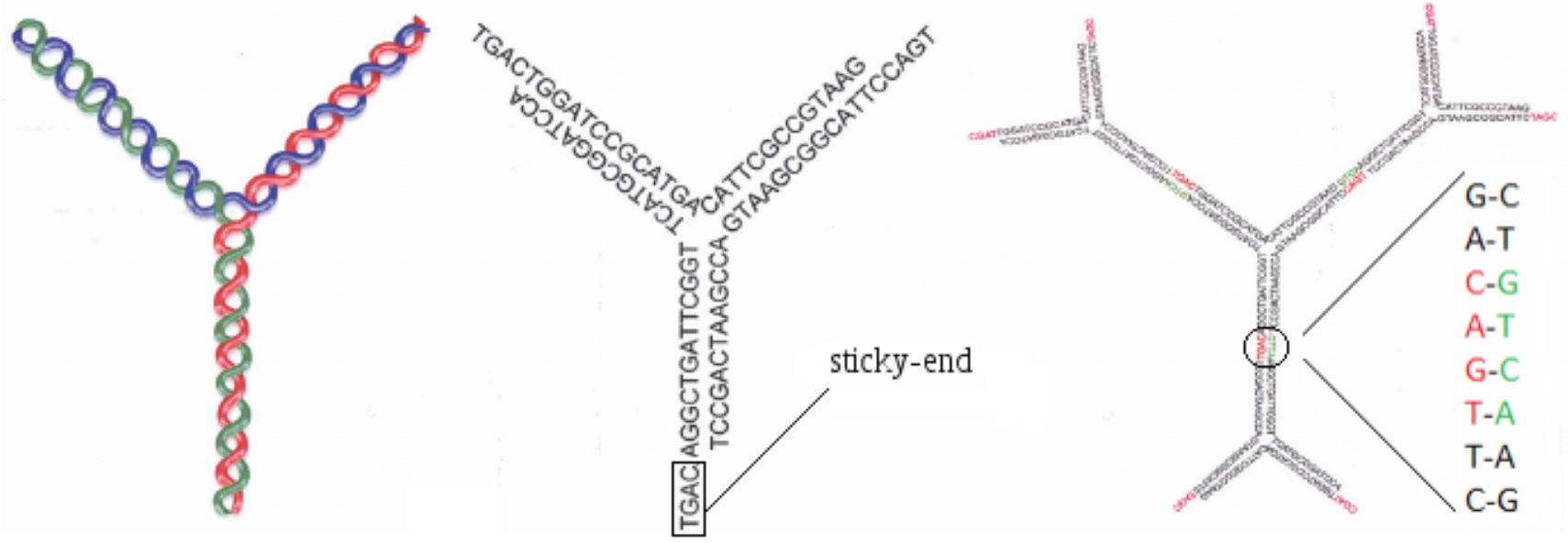
mmferrari@usf.edu
18 January, 2019

New laboratory techniques have been developed using the Watson-Crick complementarity properties of DNA strands to assemble 2D and 3D nanostructures



J. Am. Chem. Soc. **136**, 11198-11211 (2014)

The earliest self-assembly process builds target structures from DNA **branched junction molecules**



Mater. Today **6**, 38-43 (2003)

Let Σ be a finite set of symbols called an **alphabet**

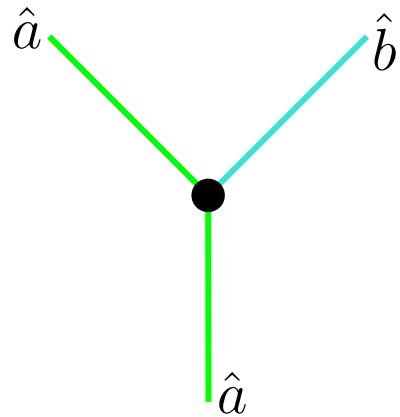
Two sticky-ends having the same sequence of bases are of the same **sticky-end type**

We denote a particular sticky-end type by $a \in \Sigma$ and its complement by $\hat{a} \in \hat{\Sigma}$

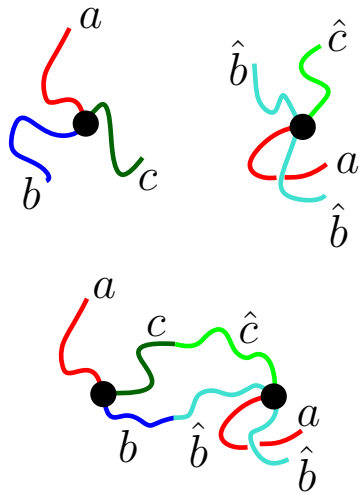
$$a = \text{ATCGTT} \leftrightarrow \hat{a} = \text{TAGCAA}$$
$$b = \text{CCATTG} \leftrightarrow \hat{b} = \text{GGTAAC}$$

$\Sigma \cup \hat{\Sigma}$ is the set of sticky-end types

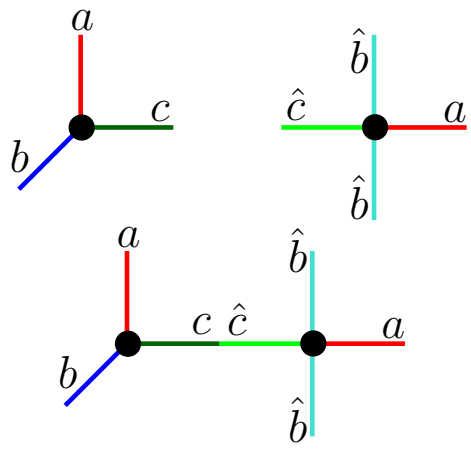
A **tile** is the combinatorial abstraction of a branched junction molecule: it consists of a vertex with labeled half-edges



Flexible Tiles VS Rigid Tiles



Flexible tile connections



Rigid tile connections

A **rigid tile** has the following properties:

- the arms are straight, rigid, and of unit length
- the arms are fixed at specified angles with respect to each other about the vertex
- the arms do not twist, compress, or elongate
- a tile has at least two arms
- two tiles connect only at the sticky-ends forming a straight edge when they connect

Flexible tiles:

- Which complexes can be assembled from a given set of tiles?
- How can we efficiently build a target graph?
- ...

Models for rigid tiles

1. Geometric model
2. Chemical model
3. Carpenter model
4. Chemical carpenter model

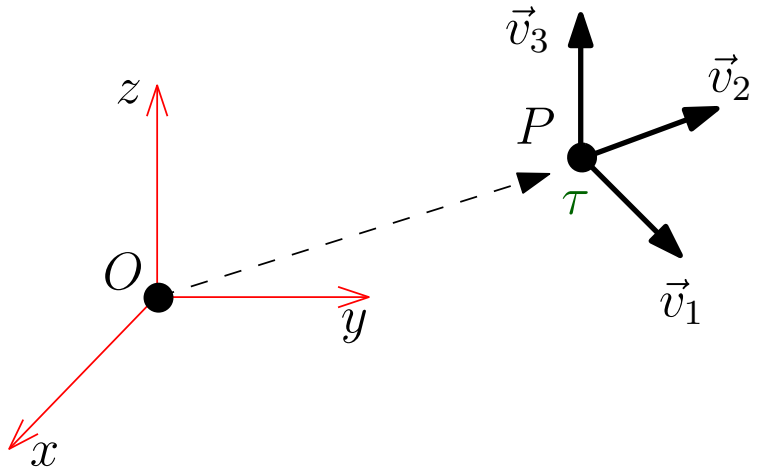
F., Cook, Houlihan, Rouleau, Seeman, Pangborn, Ellis-Monaghan, Design formalism for DNA self-assembly of polyhedral skeletons using rigid tiles. J. Math. Chem. *56*(5), 1365-1392 (2018)

Ellis-Monaghan, Jonoska, Pangborn, DNA nanostructures: mathematical design and problem encoding, in Algebraic and Combinatorial Computational Biology (2018), pp. 35-60

The geometric model

We record the geometric structure of the tile:

- the position of the arms with respect to the vertex
- the inter-arm angles



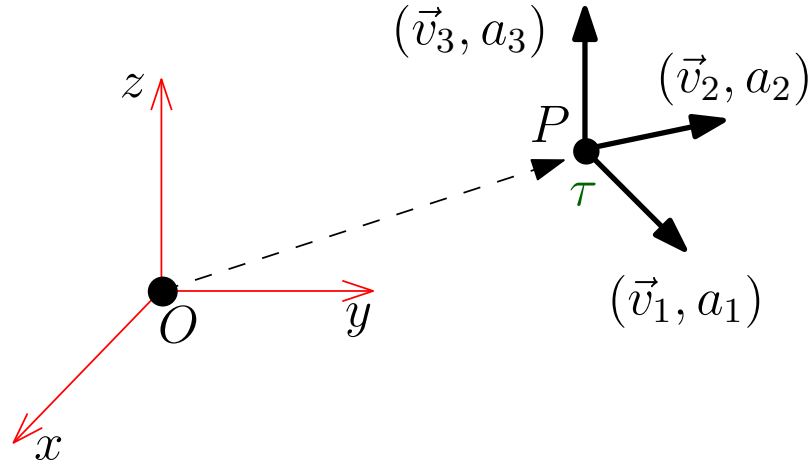
A k-armed **geometric tile** is an ordered pair $\tau = (P, \mathcal{A})$ where

- $P = (x_P, y_P, z_P)$ represents the position of the tile's center
- $\mathcal{A} = \{\vec{v}_1, \ldots, \vec{v}_k\}$ is a set of k unit vectors that are all distinct

The chemical model

We record:

- the geometric structure of the tile
- the sticky-end types for each arm



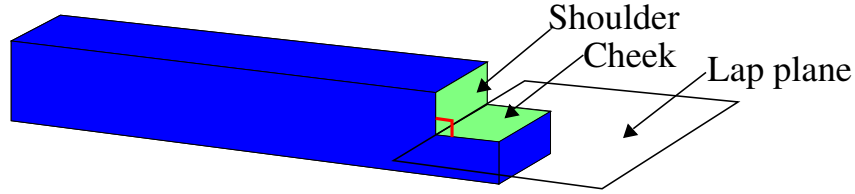
A **chemical arm type** is an ordered pair $(\vec{v}, a)$, where $\vec{v} \in \mathbb{R}^3$ is a unit vector and $a \in \Sigma \cup \hat{\Sigma}$ is a sticky-end type

A k-armed **chemical tile** is an ordered pair $\tau = (P, \mathcal{A})$ where $\mathcal{A} = \{(\vec{v}_1, a_1), \ldots, (\vec{v}_k, a_k)\}$ is a set of k chemical arm types such that $\vec{v}_i$ are all distinct

The carpenter model

We record:

- the geometric structure of the tile
- the orientation of the sticky-ends with respect to the tile



A **carpenter arm type** is an ordered pair $(\vec{v}, \vec{n})$, where $\vec{v}$ and $\vec{n}$ are unit vectors in $\mathbb{R}^3$ such that $\vec{v} \perp \vec{n}$

A k-armed **carpenter tile** is an ordered pair $\tau = (P, \mathcal{A})$ where $\mathcal{A} = \{(\vec{v}_1, \vec{n}_1), \ldots, (\vec{v}_k, \vec{n}_k)\}$ is a set of k carpenter arm types such that $\vec{v}_i$ are all distinct

The chemical carpenter model

We record:

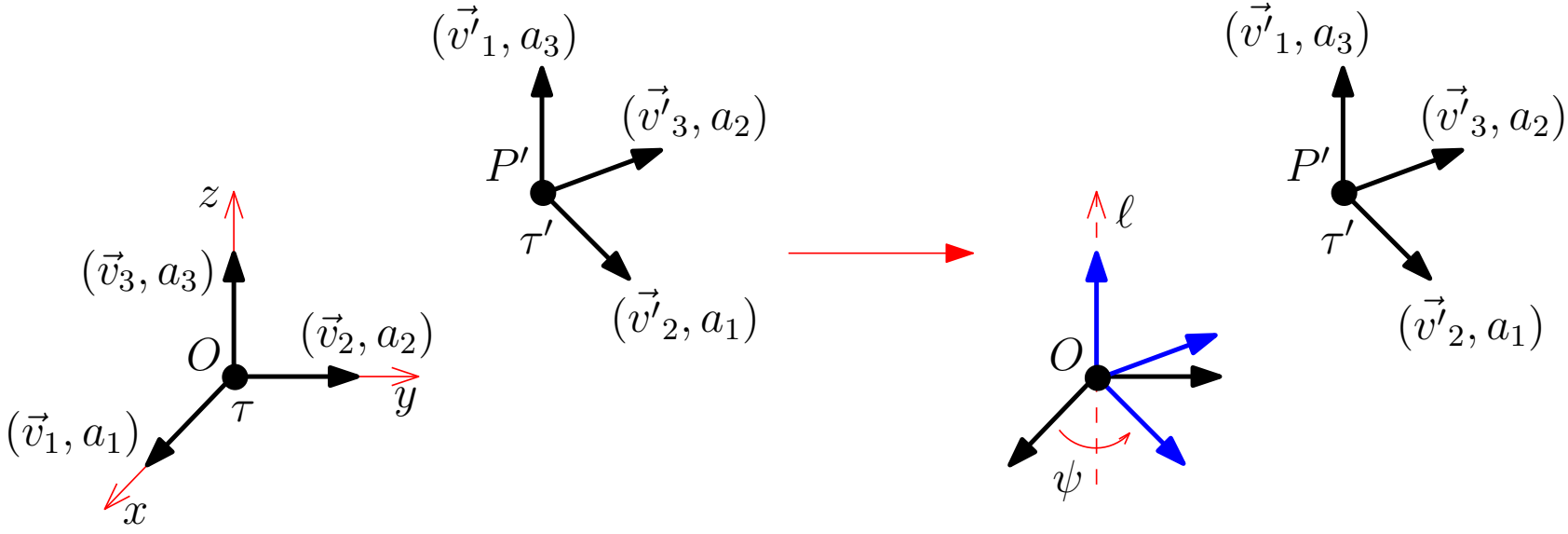
- the geometric structure of the tile
- the orientation of the sticky-ends with respect to the tile
- the sticky-end type of each arm

A **chemical carpenter arm type** is an ordered triple $(\vec{v}, \vec{n}, a)$ where $\vec{v}$ and $\vec{n}$ are orthogonal vectors in $\mathbb{R}^3$ and $a \in \Sigma \cup \hat{\Sigma}$ is a sticky-end type

A **chemical carpenter tile** with k arms is a pair $\tau = (P, \mathcal{A})$ where $\mathcal{A} = \{(\vec{v}_1, \vec{n}_1, a_1), \ldots, (\vec{v}_k, \vec{n}_k, a_k)\}$ is a set of k chemical carpenter arm types such that $\vec{v}_i$ are all distinct

Chemical tile types

When are two chemical tiles "distinct"?



Two chemical tiles $\tau = (P, \mathcal{A})$ and $\tau' = (P', \mathcal{A}')$ are **equivalent** ($\tau \sim_c \tau'$) if there exists a rotation $\mathcal{R}$ such that

$$\mathcal{A}' = \mathcal{R}(\mathcal{A}) = \{(\mathcal{R}\vec{v}_1, a_1), \ldots, (\mathcal{R}\vec{v}_k, a_k)\}$$

A **chemical tile type** is an equivalence class of chemical tiles with respect to $\sim_c$

Geometric Model

$$\mathcal{A}' = \mathcal{R}(\mathcal{A}) = \{\mathcal{R}\vec{v}_1, \mathcal{R}\vec{v}_2, \ldots, \mathcal{R}\vec{v}_k\}$$

Carpenter Model

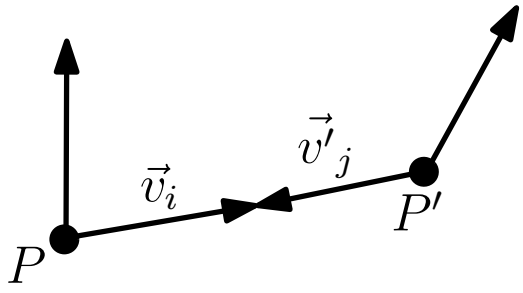
$$\mathcal{A}' = \mathcal{R}(\mathcal{A}) = \{(\mathcal{R}\vec{v}_1, \mathcal{R}\vec{n}_1), \ldots, (\mathcal{R}\vec{v}_k, \mathcal{R}\vec{n}_k)\}$$

Chemical Carpenter Model

$$\mathcal{A}' = \mathcal{R}(\mathcal{A}) = \{(\mathcal{R}\vec{v}_1, \mathcal{R}\vec{n}_1, a_1), \ldots, (\mathcal{R}\vec{v}_k, \mathcal{R}\vec{n}_k, a_k)\}$$

Joining Rigid Tiles

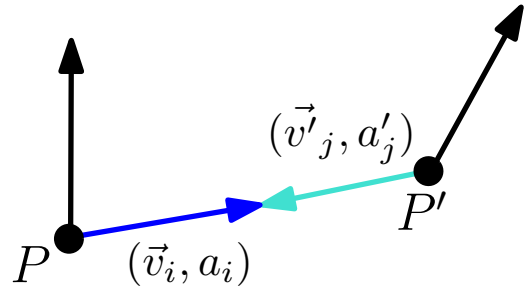
Geometric model



$\vec{v'}_j = -\vec{v}_i$

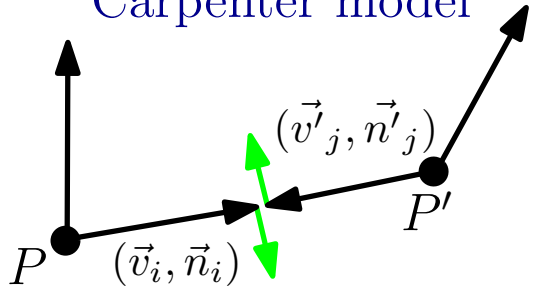
$P + \vec{v}_i - \vec{v'}_j = P'$

Chemical model



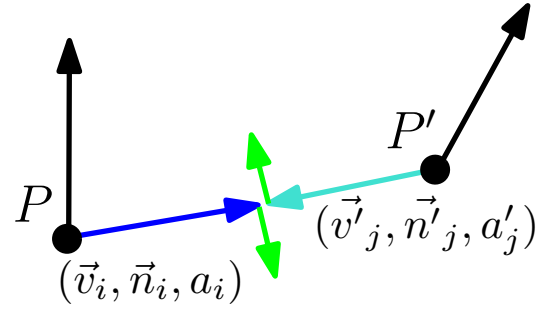
add $a'_j = \widehat{a}_i$

Carpenter model



add $\vec{n'}_j = -\vec{n}_i$

Chemical Carp. model



add both

Rigid Graphs

A **rigid graph** G is a straight edge embedding of a graph G in $\mathbb{R}^3$

Notation:

For a graph G, let μ be the map $\mu : E(G) \to V^{(2)}$ where $V^{(2)}$ is the set of unordered pairs of elements of $V(G)$

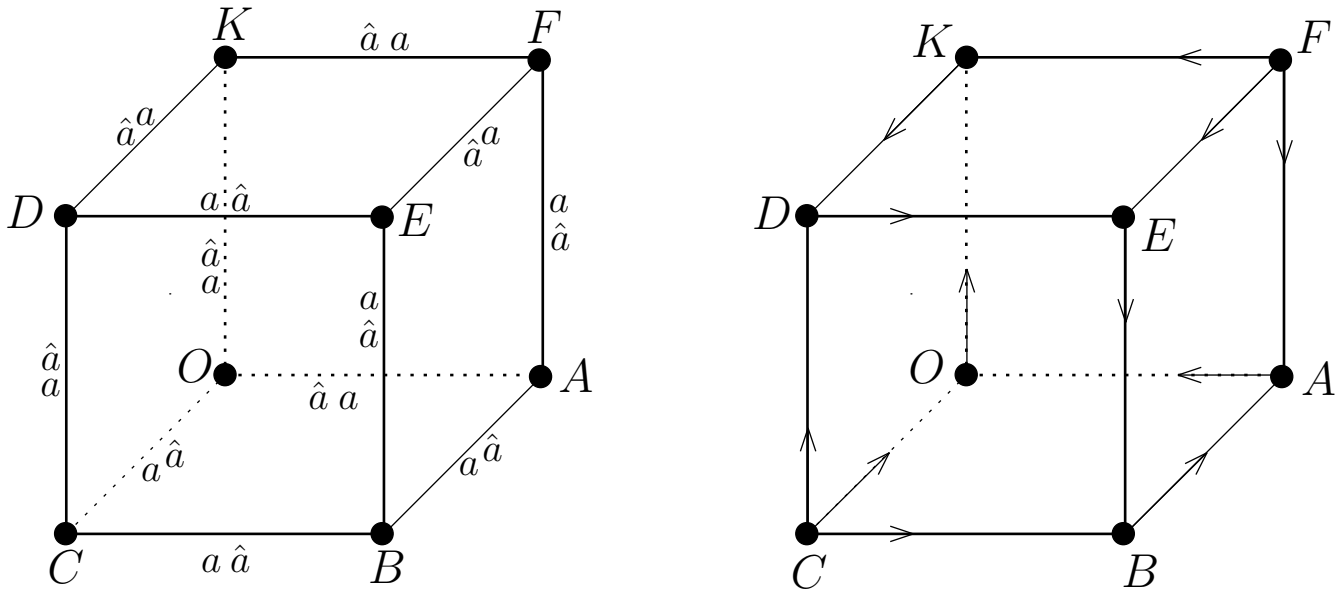
If $\mu(e) = \{u, v\}$, then u and v are the vertices incident with e

We denote by $H(G)$ the set of **half-edges** of G

For $v \in \mu(e)$, we write (v, e) to denote the half-edge of e incident to v

Assembly Design - Chemical Model

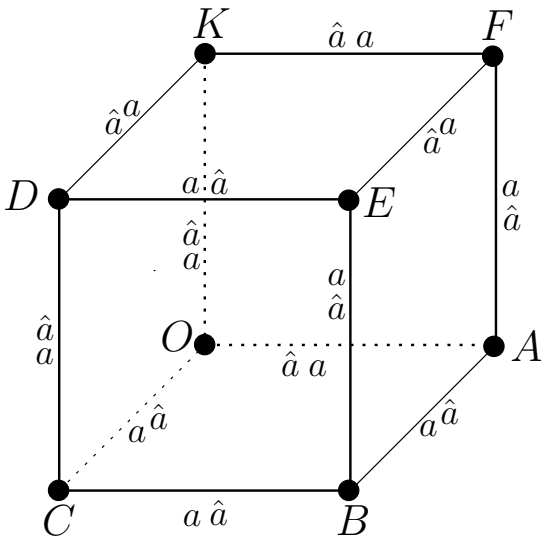
An **assembly design** of G is a labeling $\lambda : H(G) \to \Sigma \cup \hat{\Sigma}$ such that if $e \in E(G)$ and $\mu(e) = \{u, v\}$, then $\widehat{\lambda(v, e)} = \lambda(u, e)$

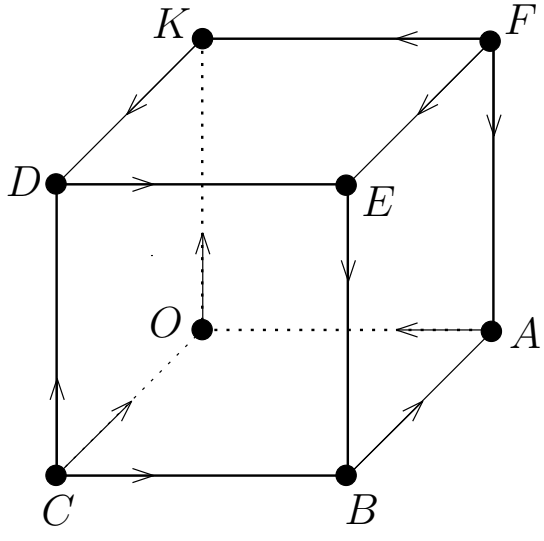


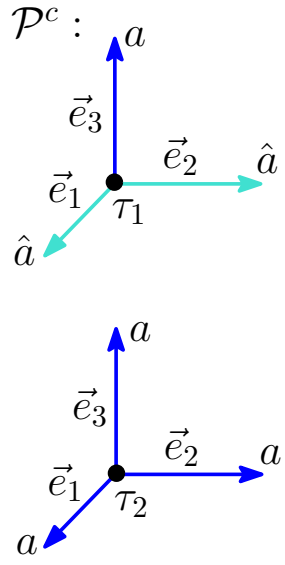
If $\lambda(v, e) = a$ and $\lambda(u, e) = \hat{a}$, we associate:

- the line segment e with $\vec{e} = u - v$
- the half-edge (v, e) with $\vec{w}_{(v,e)} = \frac{1}{2}\vec{e}$
- the half-edge (u, e) with $\vec{w}_{(u,e)} = -\frac{1}{2}\vec{e}$

A set of **chemical tiles associated with an assembly design** λ of G is a set $\mathcal{S}^c_\lambda(G) = \{\tau_v \mid v \in V(G)\}$, where $\tau_v = (v, \mathcal{A}_v)$ and $\mathcal{A}_v = \{(\vec{w}_{(v,e)}, \lambda(v,e)) \mid e \in E(G), v \in \mu(e)\}$







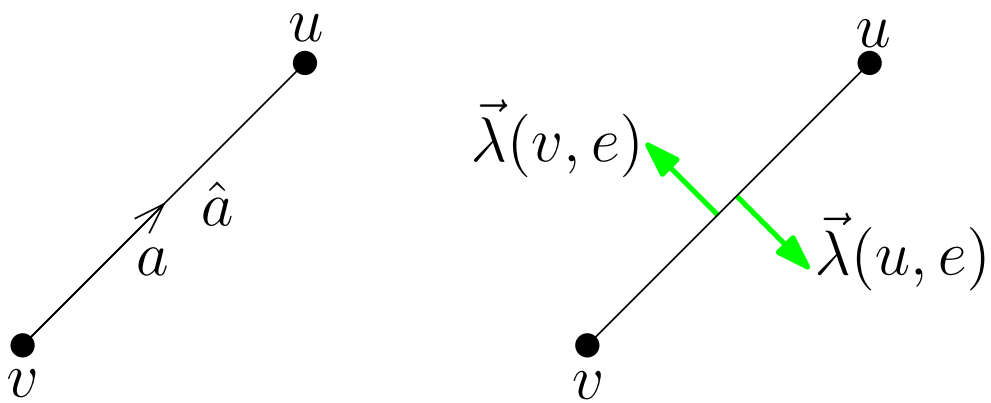
A **chemical pot** $\mathcal{P}^c$ is a set of chemical tile types

$\mathcal{P}^c$ **realizes** the rigid graph G if there exists an assembly design λ of G such that every chemical tile in $\mathcal{S}^c_\lambda(G)$ is equivalent to an element in $\mathcal{P}^c$

Carpenter Model & Chemical Carpenter Model

We consider:

- a labeling $\lambda : H(G) \to \Sigma \cup \hat{\Sigma}$ as already defined
- a labeling $\vec{\lambda}$ of the half-edges of G with the unit vectors of $\mathbb{R}^3$ such that if $e \in E(G)$ and $\mu(e) = \{u, v\}$, then $\vec{\lambda}(v, e)$ is orthogonal to the line segment e and $\vec{\lambda}(v, e) = -\vec{\lambda}(u, e)$



Distinction between the models

The set of rigid graphs realized by a pot $\mathcal{P}$ is called the **output** of $\mathcal{P}$ and is denoted by $\mathcal{O}(\mathcal{P})$

For $k \in \mathbb{N}$, let

$$\mathcal{C}_k = \{\mathcal{O}(\mathcal{P}) \mid \mathcal{P} \text{ has } k \text{ tile types}\}$$

We compare the four models by considering the following sets:

$$\mathcal{C}_k^g \qquad \mathcal{C}_k^c \qquad \mathcal{C}_k^{cp} \qquad \mathcal{C}_k^{cc}$$

Are these distinct?

Three-dimensional case

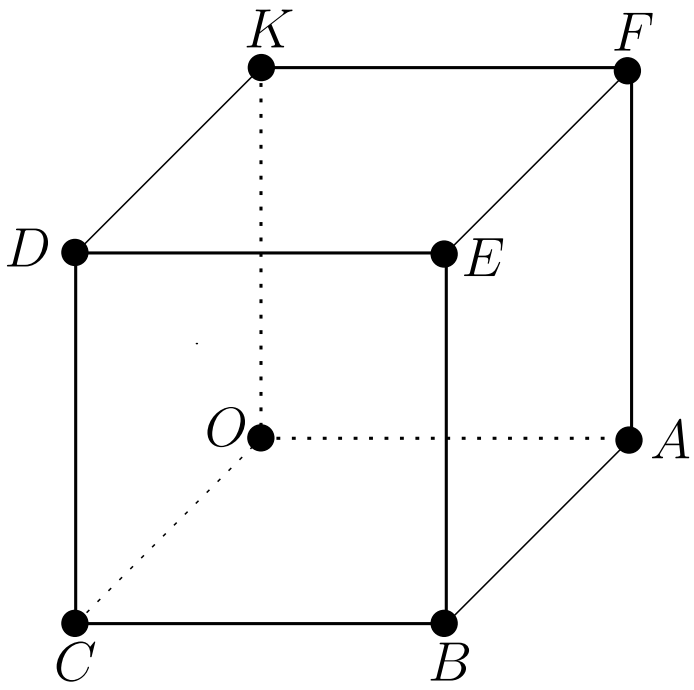
For $k = 1$, the models are distinct from each other

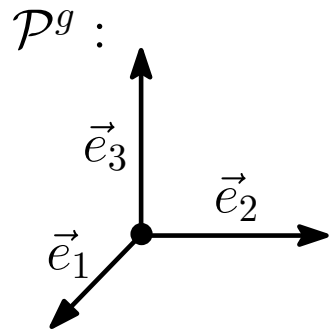
Two-dimensional case

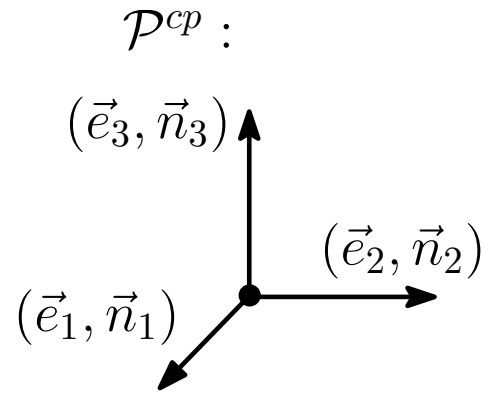
For $k = 1$, $\mathcal{C}_1^g \neq \mathcal{C}_1^c, \mathcal{C}_1^{cc}$ and $\mathcal{C}_1^{cp} \neq \mathcal{C}_1^c, \mathcal{C}_1^{cc}$

For $k \in \mathbb{N}$, we have $\mathcal{C}_k^g = \mathcal{C}_k^{cp}$ and $\mathcal{C}_k^c = \mathcal{C}_k^{cc}$

Example: the cube







$$\vec{n}_1 = \langle 0, -\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}} \rangle$$

$$\vec{n}_2 = \langle -\tfrac{1}{\sqrt{2}}, 0, -\tfrac{1}{\sqrt{2}} \rangle$$

$$\vec{n}_3 = \langle \tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}, 0 \rangle$$

Chemical (Carpenter) Model

If $\mathcal{P}^c = \{\tau\}$ realizes the cube, then the multiset of sticky-end types of τ is $\{a, a, \hat{a}\}$ or $\{a, \hat{a}, \hat{a}\}$

A cube requires at least 2 chemical tile types

Thank You for Listening!

Work supported by the NSF grants CCF-1526485 and DMS-1800443, the NIH grant R01GM109459-01, and the Southeast Center for Mathematics and Biology (an NSF-Simons Research Center for Mathematics of Complex Biological Systems) under NSF grant DMS-1764406 and Simons Foundation grant 594594