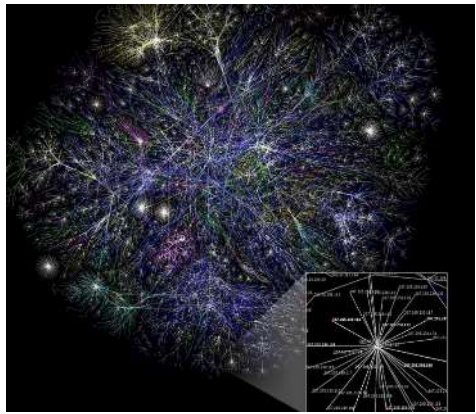


# Graph theoretical tools for DNA self-assembly

Jo Ellis-Monaghan

Korteweg-de Vries Instituut voor Wiskunde, Universiteit van Amsterdam

# New applications → New mathematics



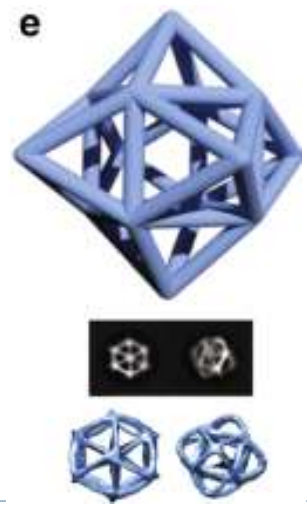
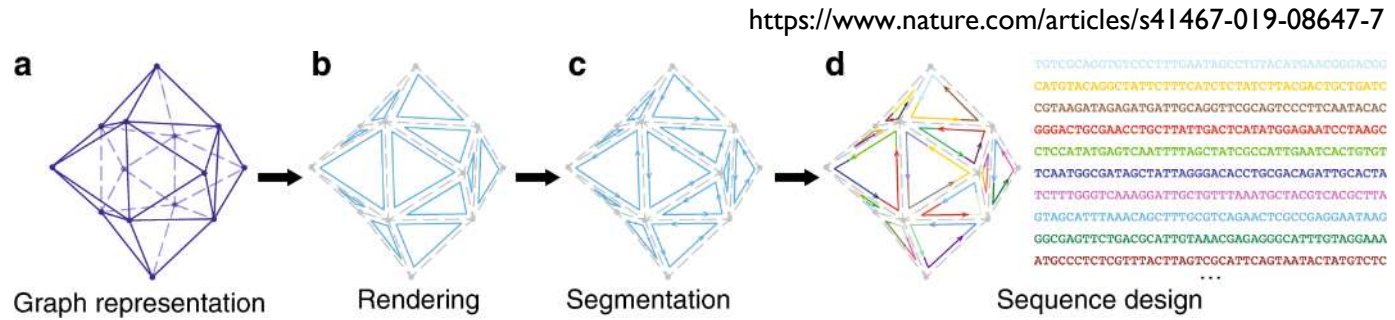
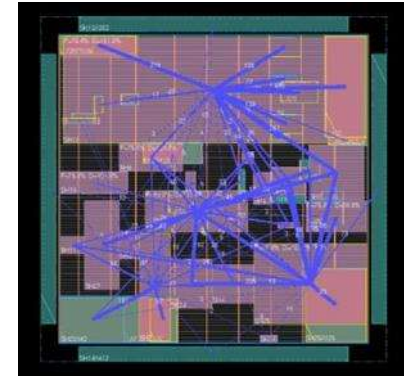
Internet →  
Random Graphs

[https://en.wikipedia.org/wiki/Random\\_graph](https://en.wikipedia.org/wiki/Random_graph)

VLSI → Graph Drawing

```

- I14A BUFPK_N + PLMBU_I COMU COMU ) FS :
- I10 NOR2_B + PLACED ( 3760 960 ) N :
- I9 AO121_B + PLACED ( 3840 480 ) N :
- I7 NOR2_A + PLACED ( 3840 1920 ) FS :
- I6 NAND2_D + PLACED ( 2320 1920 ) FS :
- I25 NOR2_A + PLACED ( 2120 2400 ) N :
- I4 NAND2_D + PLACED ( 2360 2400 ) N :
- I0 AO121_A + PLACED ( 2680 1920 ) FS :
- I13 AO121_A + PLACED ( 3080 1440 ) N :
- I17 AO121_B + PLACED ( 3180 1440 ) N :
- I11 XNOR2_A + PLACED ( 1320 0 ) FS :
- I24 NOR2_C + PLACED ( 1840 480 ) N :
- I26 XNOR2_C + PLACED ( 2040 2400 ) N :
- I27 NOR2_B + PLACED ( 1920 1440 ) N :
- I28 XNOR2_C + PLACED ( 1160 480 ) N :
- COUNT1/I_3 NAND2_A + PLACED ( 3840 1440 )
- COUNT1/I_7 NAND2_A + PLACED ( 3960 960 ) F
- COUNT1/I_8 NAND2_A + PLACED ( 1920 960 ) F
- COUNT1/I_9 NAND2_A + PLACED ( 2680 480 )
- COUNT1/I_14 NOR2_A + PLACED ( 3480 1440 )
- COUNT1/I_0 AND2_B + PLACED ( 2120 0 ) FS :
- COUNT1/I_1 NAND2_A + PLACED ( 2440 0 ) FS
- COUNT1/I_1517 NOR2_A + PLACED ( 2160 960 )
- COUNT1/I_17 NOR2_A + PLACED ( 2400 480 ) N
- COUNT1/I_16 NOR2_A + PLACED ( 2120 480 ) N
- COUNT1/I_11 AO121_A + PLACED ( 2720 0 ) FS
- COUNT1/I_10 NAND2_A + PLACED ( 2320 960
- COUNT1/counter_out_freq_0 D_F_1PH0001_LPC_E
- COUNT1/I_13 NOR2_A + PLACED ( 2200 2880 )
- COUNT1/counter_out_freq_1 D_F_1PH0001_LPC_E
- COUNT1/I_42 MUX21_B + PLACED ( 3680 1920 )
- COUNT1/counter_out_freq_3 D_F_1PH0001_LPC_E
- COUNT1/I_49 AO121_A + PLACED ( 3760 960 )
- COUNT1/I_10 AO121_A + PLACED ( 3880 0 ) FS
- COUNT1/counter_out_freq_3 D_F_1PH0001_LPC_E
- COUNT1/I_16 AO121_A + PLACED ( 2120 960 )
- COUNT1/I_17 AO121_A + PLACED ( 2720 960 )
    
```

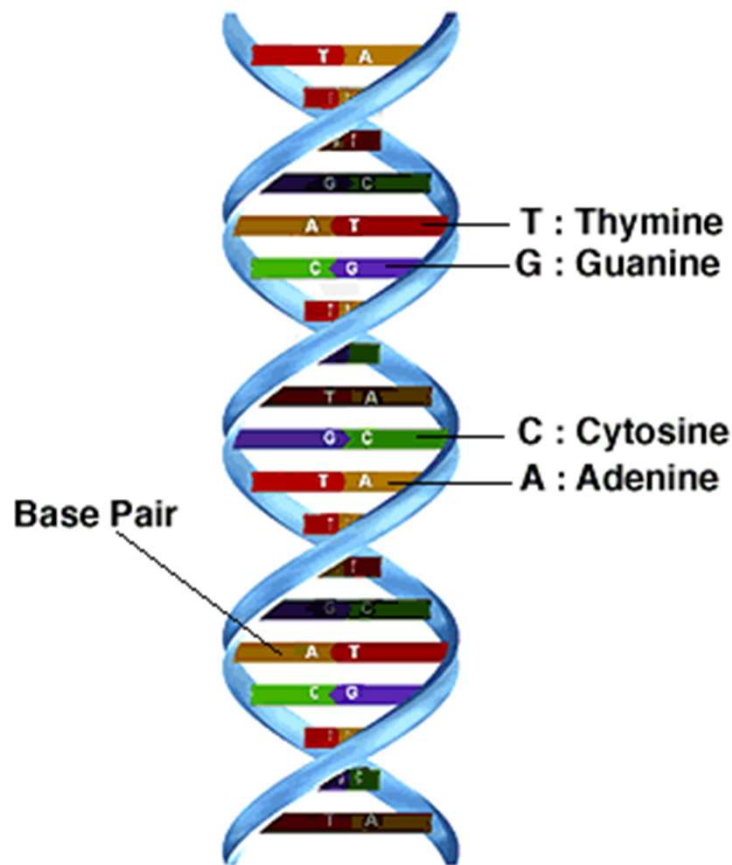


Self-assembly → Graphical Assembly Design.



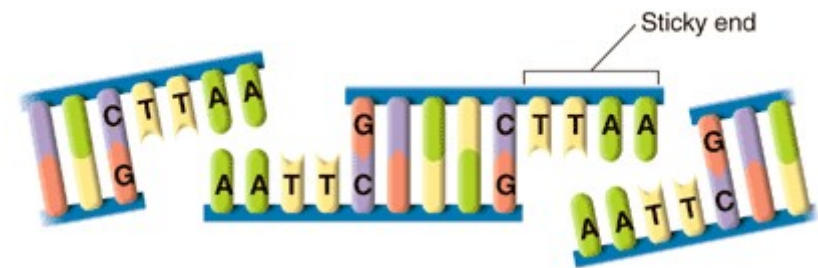
# Remember how DNA works:

## In Nature

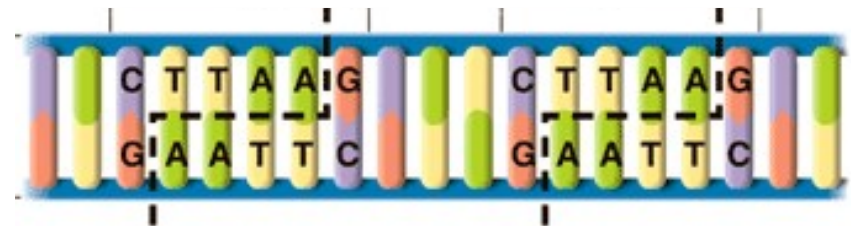


## In The Lab

Create fragments of DNA with carefully designed 'sticky ends', or unsatisfied bases.



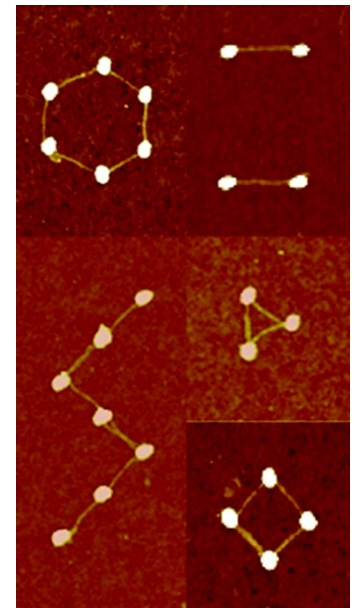
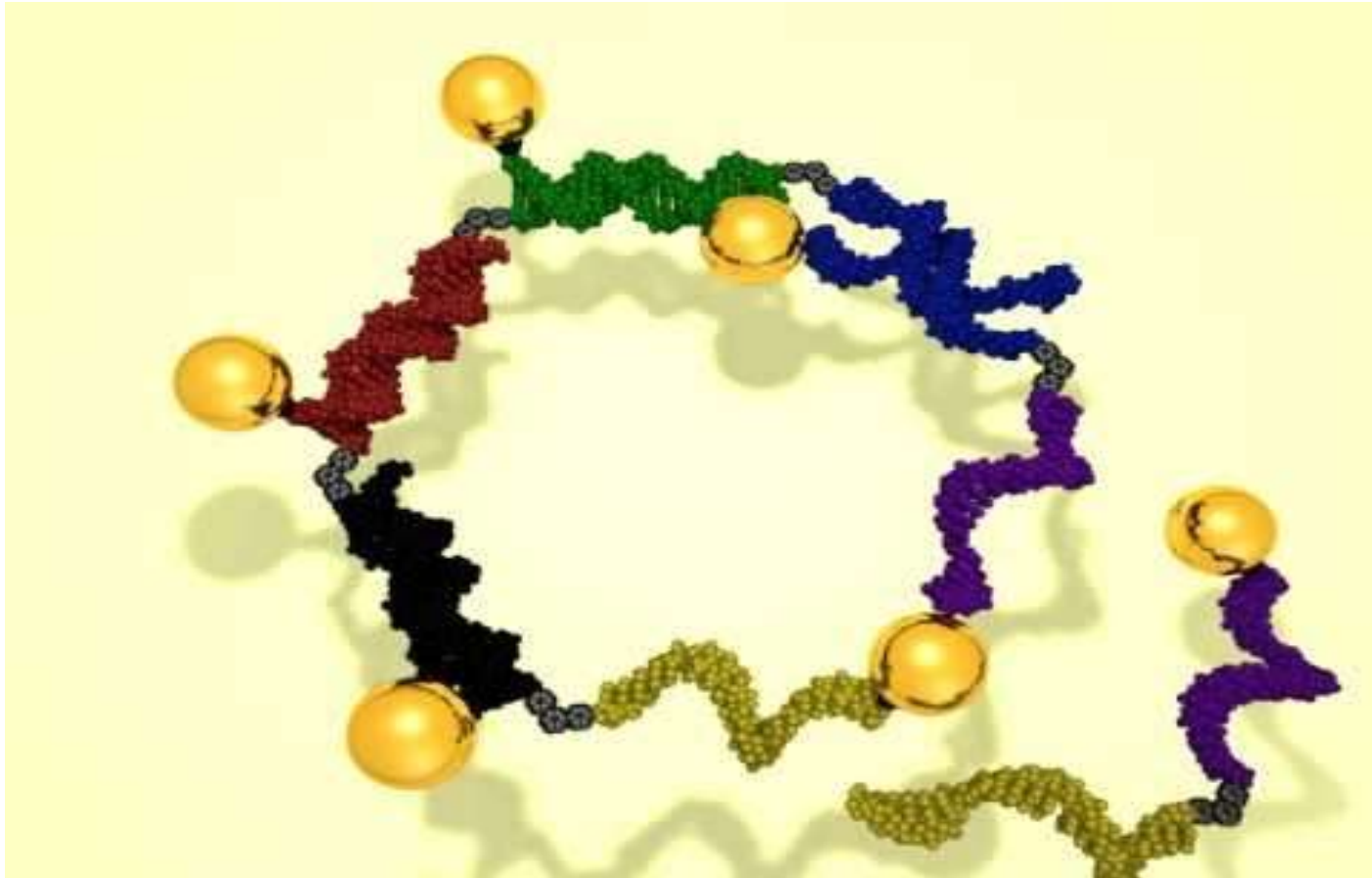
These will stick to each other, self-assembling into the target structure.



DNA self-replicates, so can create more and more of the targets.

# Controlling shapes with engineered DNA

---



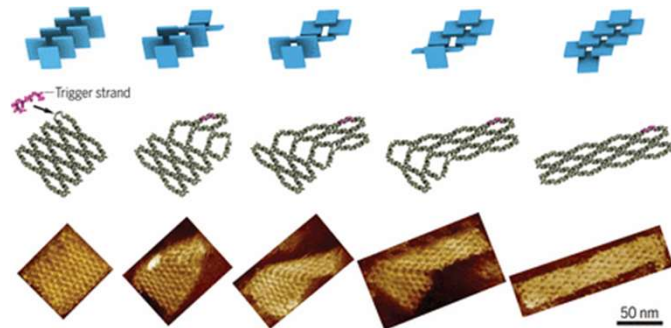
*Nano Lett.*, 2010, 10 (12), pp 5065–5069



[http://www.youtube.com/watch?v=6\\_ewP22n2mg&feature=related](http://www.youtube.com/watch?v=6_ewP22n2mg&feature=related)

# Why self-assembling nanostructures?

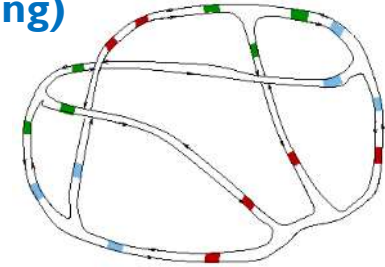
## Nano-robotics



Reconfiguration of DNA molecular arrays driven by information relay. Song, Li, Wang, Mayer, Mao, Ke  
*Science* 28 Jul 2017: Vol. 357, Issue 6349

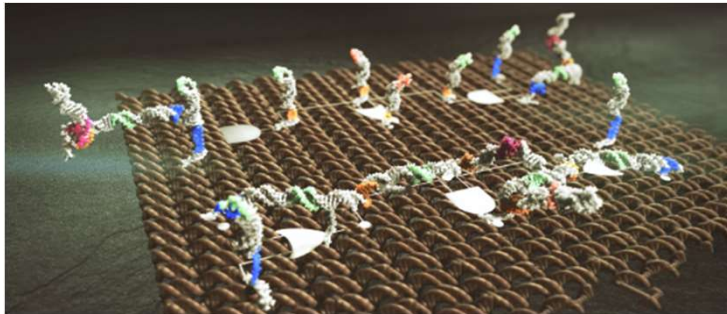
## Biomolecular computing (Hamilton Cycle/3-Sat/Graph Coloring)

<http://shell.cas.usf.edu/~jonoska/bio-comp/node4.html>

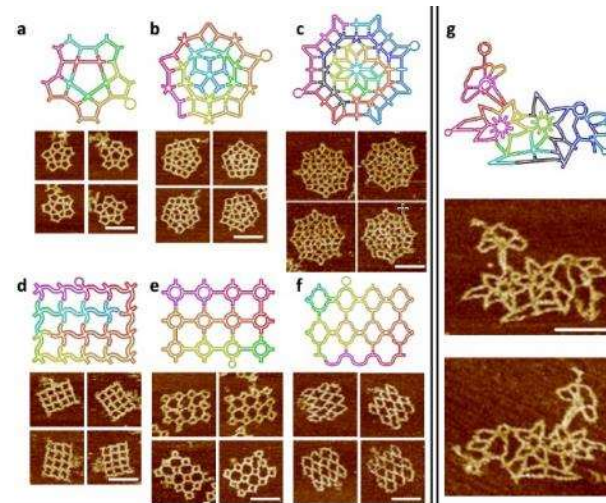


## Nano-scale meshes and filters

## Nano-circuitry



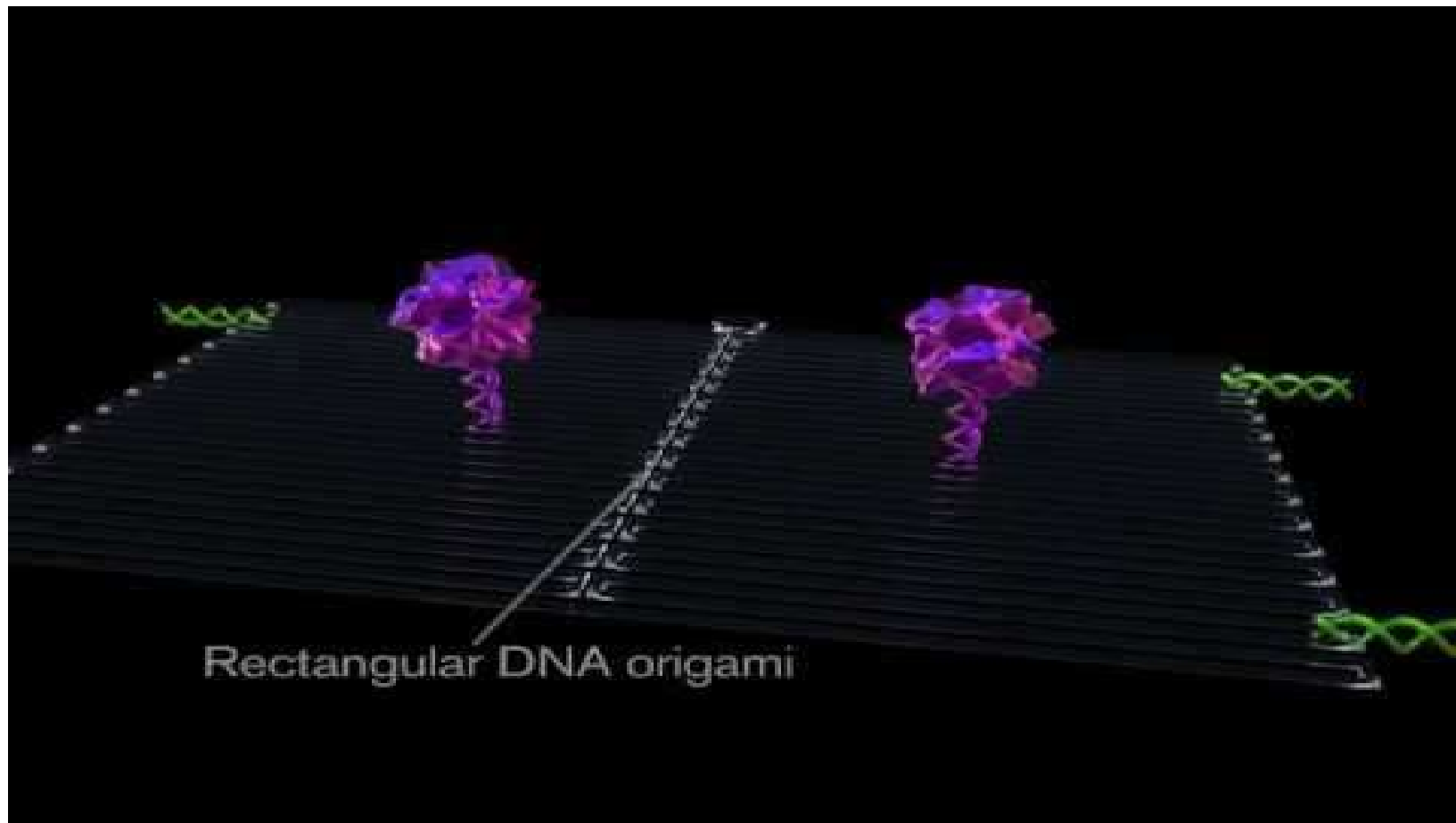
<https://www.microsoft.com/en-us/research/blog/researchers-build-nanoscale-computational-circuit-boards-dna/>



Complex wireframe DNA origami nanostructures with multi-arm junction vertices. Zhang, Jiang, Wu, Li, Mao, Liu, Yan.  
*Nature Nanotechnology* volume10, pages 779-784 (2015)

# How might this work in the body?

---

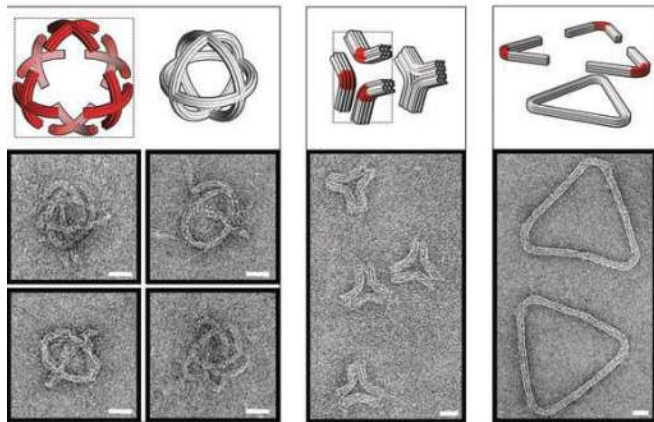


Targeted Drug Delivery

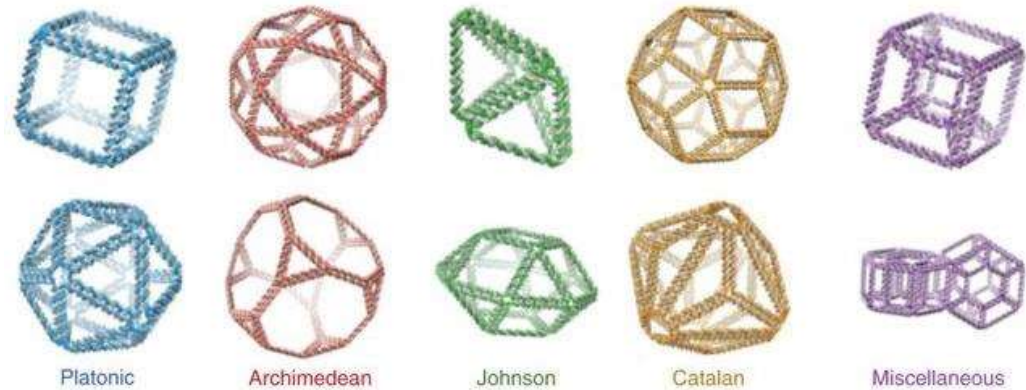
---

► <https://www.eurekalert.org/multimedia/pub/162624.php>

# Some state of the art (3D) DNA nano-objects



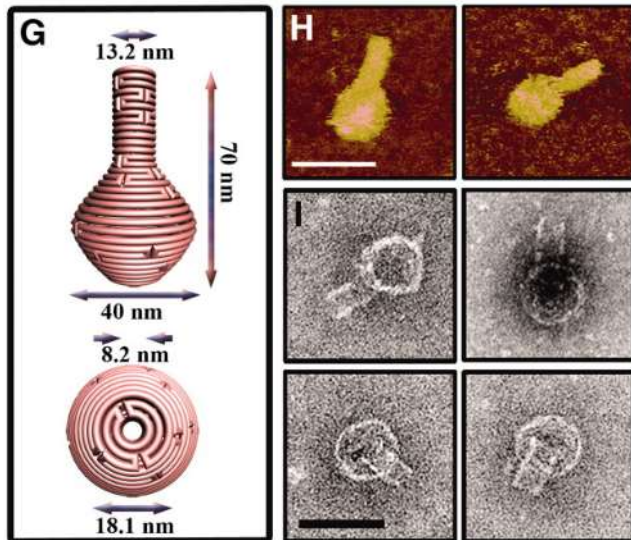
Dietz, Douglas and Shih, Science, 325, 725



Linko, Kostianen. Nature Biotechnology volume34, pages826–827 (2016)



Benson, Mohammed, Gardell, Masich, Czeizler, Orponen, Högberg. Nature 523, 441–444 (23 July 2015)



Han, Pal, Nangreave, Deng, Liu, and Yan, Science, 333, 342

# Our Objectives

---

Create new mathematical and computation tools for laboratories producing self-assembled DNA nanostructures.

- ▶ Several assembly paradigms:
  - Tile-based assembly ♦ DNA origami ♦ Reporter strands
- ▶ For each assembly method, usually we get:
  - ▶ Problem formulation and mathematical formalism
  - ▶ Proofs that design strategies are NP-Hard
  - ▶ Pragmatic approaches
  - ▶ New mathematical directions arising from the problem

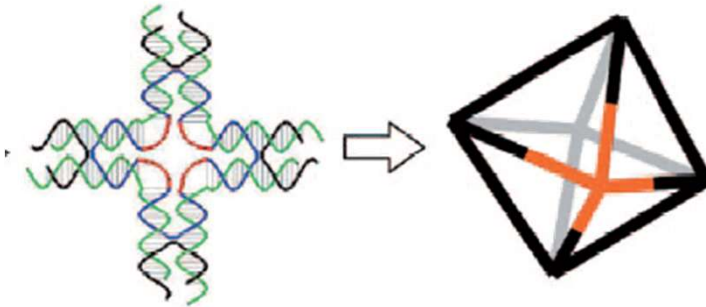




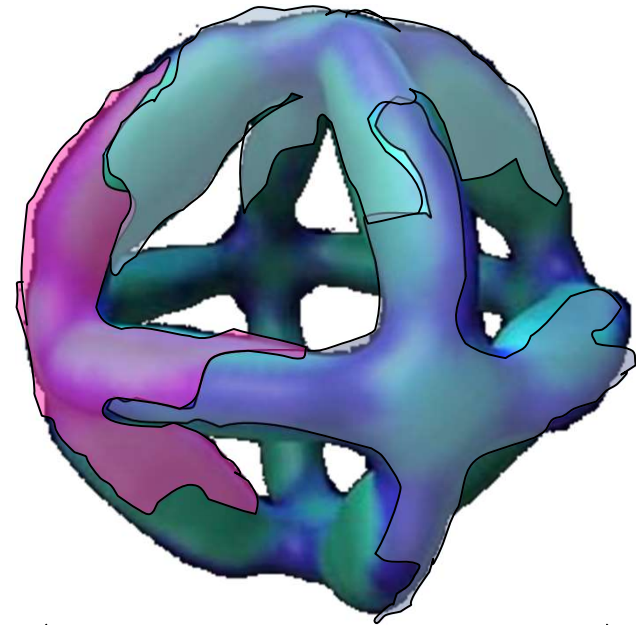
# Branched Junction Molecules

---

A self-assembled octahedron



<http://onlinelibrary.wiley.com/doi/10.1002/anie.200904513/pdf>



← 22 nanometers →

<http://www.scripps.edu/news/press/2004/021104.html>



# A visualization model

---



---

▶ <https://www.youtube.com/watch?v=X-8MP7g8XOE>

# Short and long

---



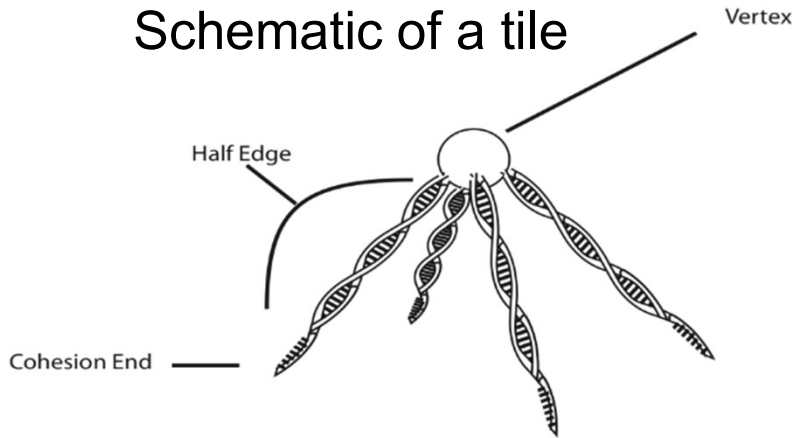
<https://gulfspecimen.org/specimen/echinochordata/brittle-stars-and-serpent-stars/>

<https://www.flickr.com/photos/edbierman/7383798192/>

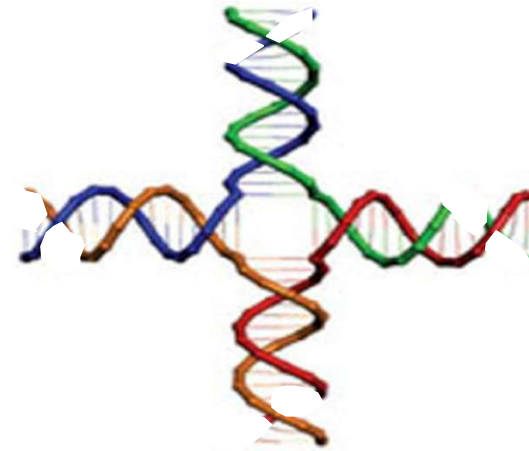


# What is a tile?

Schematic of a tile



The branched junction molecule



With the bases specified



A-T (adenine - thymine) and G-C (guanine - cytosine).

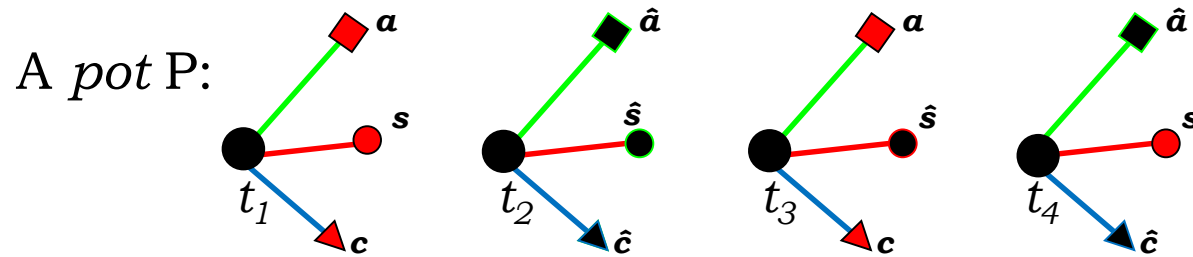
ATTCG GGTAACATTCTG  
TAAGCCATTG TAAGC

*Y-shaped DNA. Schematic diagrams of the structure (left) and sequence (middle) of Y-DNA, and dendrimer-like DNA (right).*

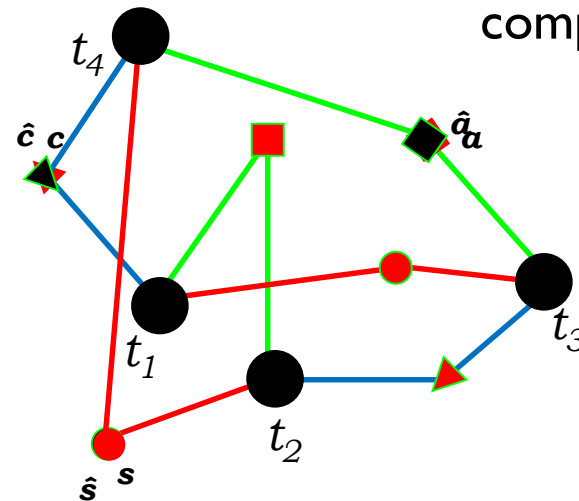
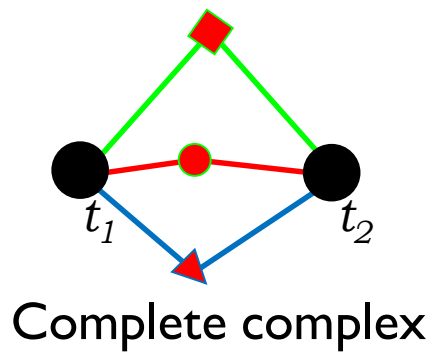
$a$                        $\hat{a}$   
 ATTCG GGTAACATTGG  
 TAAGCCCATTG TAAGC

# Combinatorial formulation

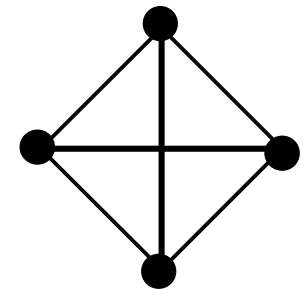
Both complete complexes and incomplete complexes can be constructed from the this pot P with 4 tiles:



The labs generally want complete complexes



Incomplete complex

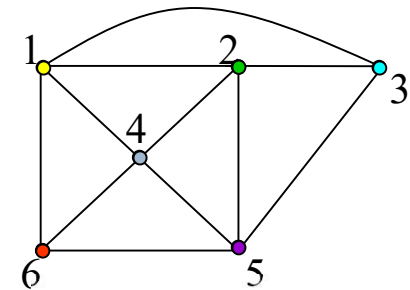


# Different paradigms:

---

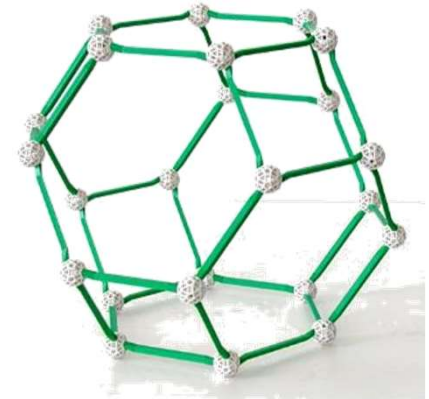
## ▶ Abstract graphs

- ▶ E.g. only care about the connectivity of nodes and edges, not physical location
- ▶ Examples: graphs for biocomputation or colorability.



## ▶ Geometric graphs

- ▶ Specific embeddings in space
- ▶ Examples: skeletons of Platonic or Archimedean solids, or crystallographic lattices.



## ▶ Constraints

- ▶ The incidental assembly of graphs smaller than the target is acceptable
- ▶ Any graph incidentally assembled must be larger than the target



# New graph invariants in each setting

---

Given a graph  $G$ ,

1. what is  $T(G)$ , the minimum number of  $k$ -armed branched junction molecules (tiles) that must be designed to create the graph?
2. What is  $B(G)$ , the minimum number of bond types (size of alphabet) needed?

For *rigid* tiles, description of the pot must also specify the geometric structures of each tile.



# Branched junction methods

---

- ▶ **Objectives:**
  - ▶ Given a graph  $G$ , find an optimal pot  $P$  that realizes  $G$ , *i.e.* minimize the number of branched junction molecules required to assemble the target structure and give their combinatorial specifications.
  - ▶ Given a pot  $P$ , find the set of graphs (up to isomorphism) realized by  $P$ .
- ▶ **In both cases need:**
  - ▶ General theory
  - ▶ Optimal solutions or algorithms for application-relevant graphs
  - ▶ Computational complexity results
- ▶ **Existing tools don't help:**
  - ▶ Coloring? No....
  - ▶ List coloring? No....
  - ▶ Automorphism group? No....

Again opens new areas of mathematical investigation

---

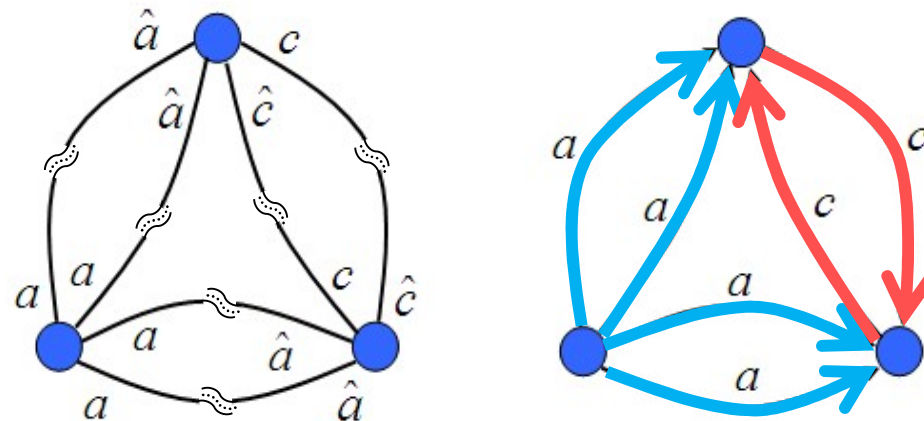




# Basic Design Strategy

---

Find an edge labeled/colored orientation of the graph, using a minimal number of colors, with as few different labelings of the half edges about the vertices as possible, and furthermore specifying the set of different vertex types (the resulting tiles).



Such a labeled/colored orientation of a graph  $G$  is called an *assembly design* of  $G$  and denoted  $\lambda$ , where  $\lambda$  maps half edges to labels.

---



# A little linear algebra

---

Let  $P$  be a pot with  $n$  tiles labeled  $t_1 \dots t_n$ , and let  $z_{ij}$  be the net number of sticky ends of type  $i$  on tile  $t_j$

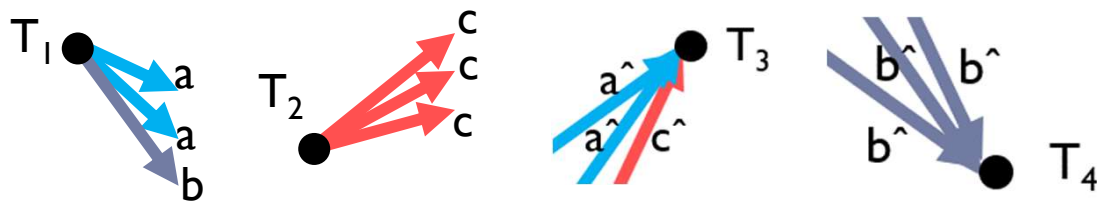
$$M(P) = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1n} & 0 \\ \vdots & \vdots & & \vdots & 0 \\ z_{m1} & z_{m2} & \dots & z_{mn} & 0 \\ 1 & 1 & \dots & 1 & 1 \end{bmatrix}$$

Theorem: Some graph  $G$  with  $m$  vertices may be constructed from the pot  $P$  if and only if  $\langle r_1 \dots r_n \rangle$  is a solution of the construction matrix  $M(P)$  and  $m \langle r_1 \dots r_n \rangle$  has integer entries.

---

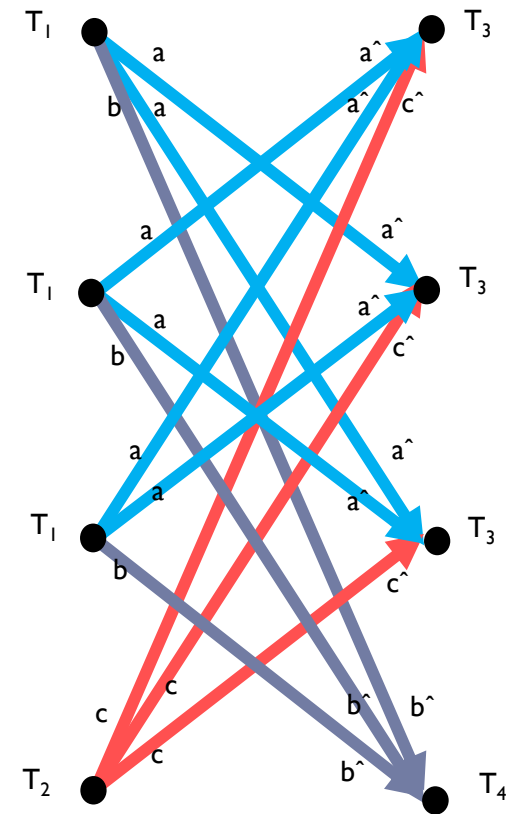


# Example



$$\begin{array}{c}
 a \\
 b \\
 c
 \end{array}
 \begin{array}{c}
 T_1 \quad T_2 \quad T_3 \quad T_4 \\
 \left[ \begin{array}{cccc|c}
 2 & 0 & -2 & 0 & 0 \\
 1 & 0 & 0 & -3 & 0 \\
 0 & 3 & -1 & 0 & 0 \\
 1 & 1 & 1 & 1 & 1
 \end{array} \right]
 \end{array}
 \xrightarrow{\text{rref}}
 \begin{array}{c}
 \left[ \begin{array}{cccc|c}
 1 & 0 & 0 & 0 & 3/8 \\
 0 & 1 & 0 & 0 & 1/8 \\
 0 & 0 & 1 & 0 & 3/8 \\
 0 & 0 & 0 & 1 & 1/8
 \end{array} \right]
 \end{array}$$

Thus, there is an 8-vertex complete complex realized by this pot with 3  $T_1$ 's, 1  $T_2$ , 3  $T_3$ 's, and 1  $T_4$ .



(Example from Harsy group)

# Integer linear programming

---

If we ask for a complete complex of a specific size  $k$ , this becomes an integer programming problem, since a solution  $\langle r_1 \dots r_n \rangle$  now is in  $\mathbf{Z}^n$

$$M(P) = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1n} & 0 \\ \vdots & \vdots & & \vdots & 0 \\ z_{m1} & z_{m2} & \dots & z_{mn} & 0 \\ 1 & 1 & \dots & 1 & \mathbf{k} \end{bmatrix}$$

All variables are restricted to be non-negative

ILP is NP hard in general, but this is a special case. Is it hard?



## The decision question

---

- ▶ Given a pot  $P$  and a positive integer  $k$ , does  $P$  realize an assembly design for a graph  $G$  with  $k$  vertices.

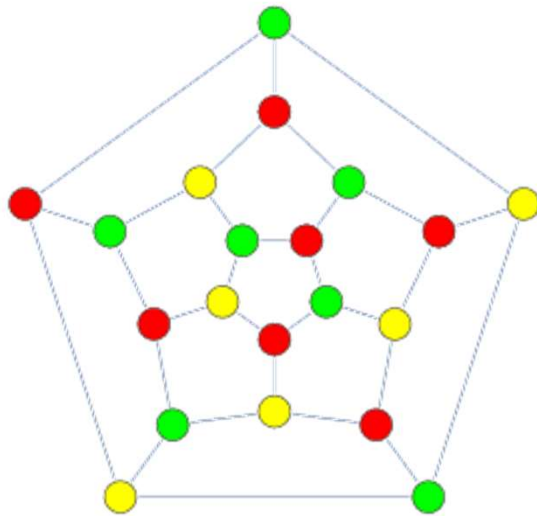
[It is fast, linear in fact, to check that an assembly design is realized by a pot  $P$ , since you only need to check that each vertex configuration appears as a tile in  $P$ .]



# Three coloring

---

- ▶ ILP is known to be NP hard in general, but this is a special case, so requires an independent proof of hardness—we do this by reduction to 3-coloring.



3-coloring a graph means coloring the vertices with 3 colors so that no pair of adjacent vertices get the same color.

3-coloring is NP-complete even for 4-regular plane graphs



## 3-coloring reduction to pot problem

---

Now assume we have a polynomial time algorithm that will tell us if a pot  $P$  will assemble a graph on  $k$  vertices, and that we are given a 4-regular plane graph  $G$  that we want to 3-color.

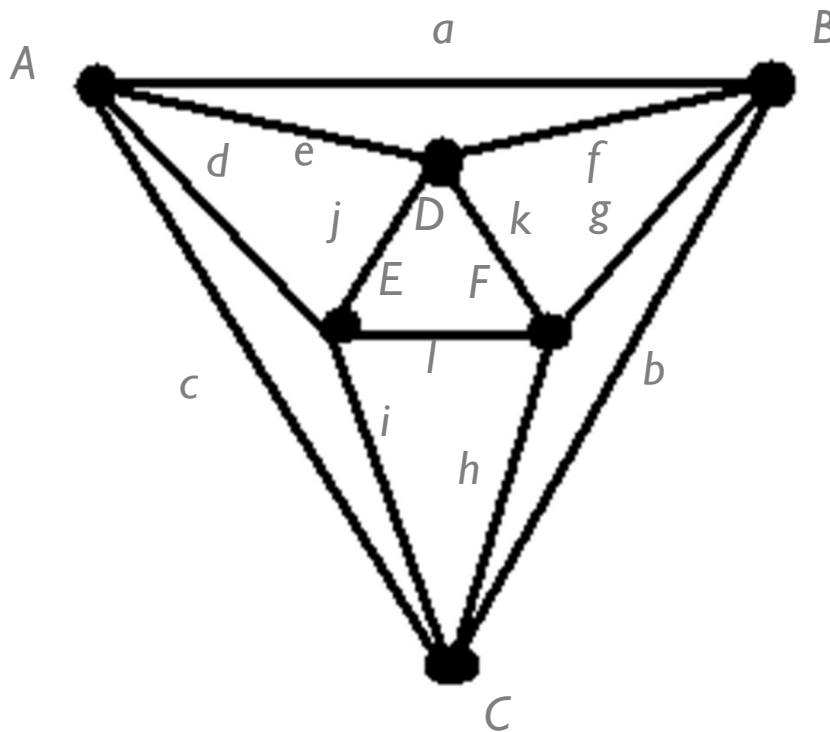
[If we are able to use this hypothetical algorithm to come up with a 3-coloring, that would mean there is a polynomial time algorithm for 3-coloring. But unless  $P=NP$ , there isn't one.]

---



# 3-coloring 4-regular plane graphs

---



$G$ , with vertices and edges labeled

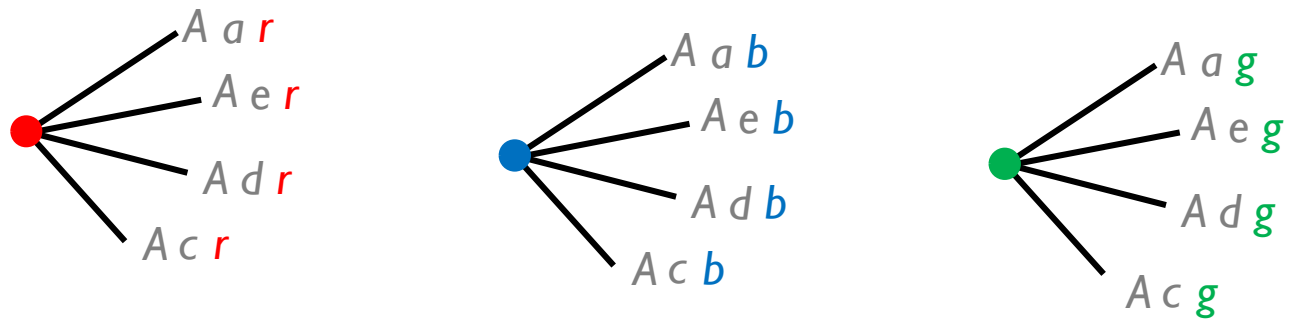
(This follows essential ideas of Jonoska, Sa-Ardyen, Seeman '03 , but adapted to our application and notation)



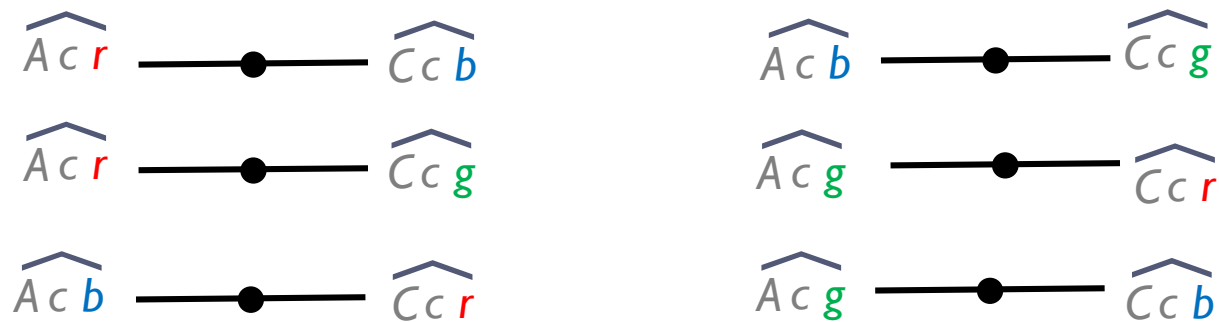


# Create a pot

- ▶ For each vertex, create three tiles, one for each color, with sticky end labels recording vertex, edge, and color, e.g. for the vertex  $A$ :



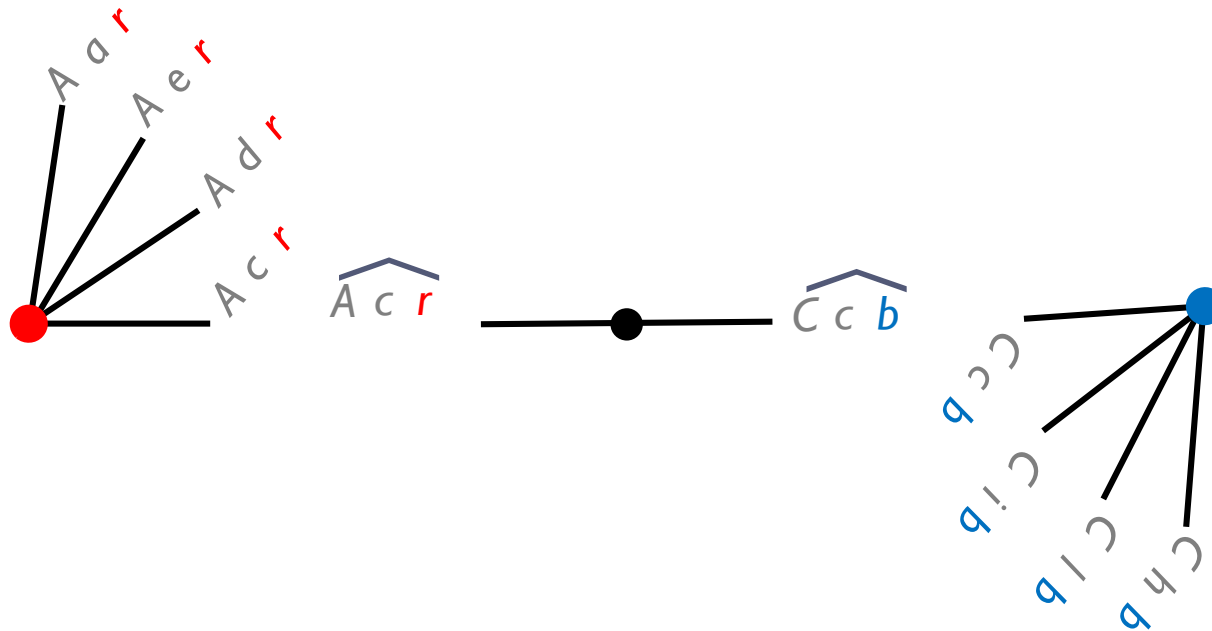
- ▶ For each edge create six tiles, one for each proper coloring of the edge's endpoints, with complementary sticky ends, e.g. for edge  $c$ :



# A unique smallest complete complex from this pot

---

- ▶ The original graph (subdivided) is the unique smallest complete complex built from this pot, and that *only* if it is three-colorable:



# Complexity summary

---

- ▶ The pot decision problem is in NP, since it is easy to check a solution.
- ▶ 3-coloring a 4-regular plane graph is NP-Complete and polynomial reduces to the pot decision problem.
- ▶ Therefore the pot-decision problem is also NP-Complete.



# Ramifications

---

- ▶ This complexity result has ramifications for both of the fundamental objectives:
  - ▶ If you have a pot  $P$ , in general you can't efficiently determine what graphs it will construct.
  - ▶ If you have a graph  $G$ , in general you can't efficiently determine a pot that will assemble  $G$ , but not anything smaller.
- ▶ This stinks for the labs, as they would really prefer a fast, general purpose algorithm.
- ▶ But it is great for mathematicians, since it means this area is a whole new playground of problems: complexity for special classes, pragmatic solutions, approximation algorithms, explicit minimal pots for high-utility graphs, etc.
- ▶ Also, exploration of the new graph parameters—what do they reveal about the structure of graphs? Girth? Tree-width?



# Pragmatic code

---

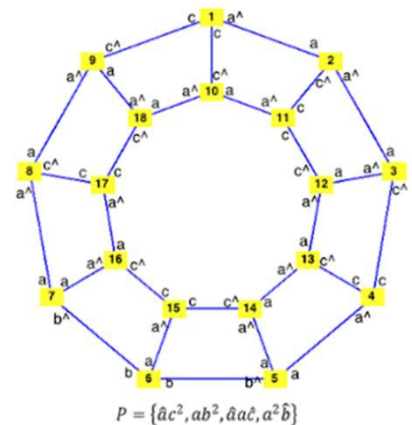
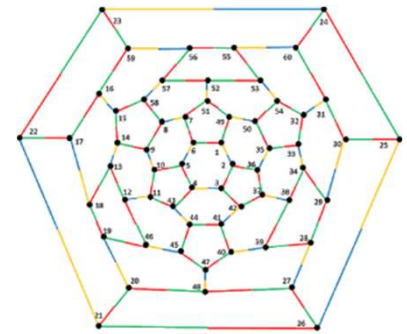
- ▶ We can use existing integer linear programming tools for small examples.
- ▶ We have specialized code that will handle 2-3 degrees of freedom and can determine if a graph is the unique smallest construct assembled by a pot.
- ▶ Lot of sporadic ad hoc cases, but quite difficult to prove optimality.
- ▶ More algorithms are needed— for special classes, in restricted settings, and for approximate solutions.



**Table A: Minimum Tile Types**

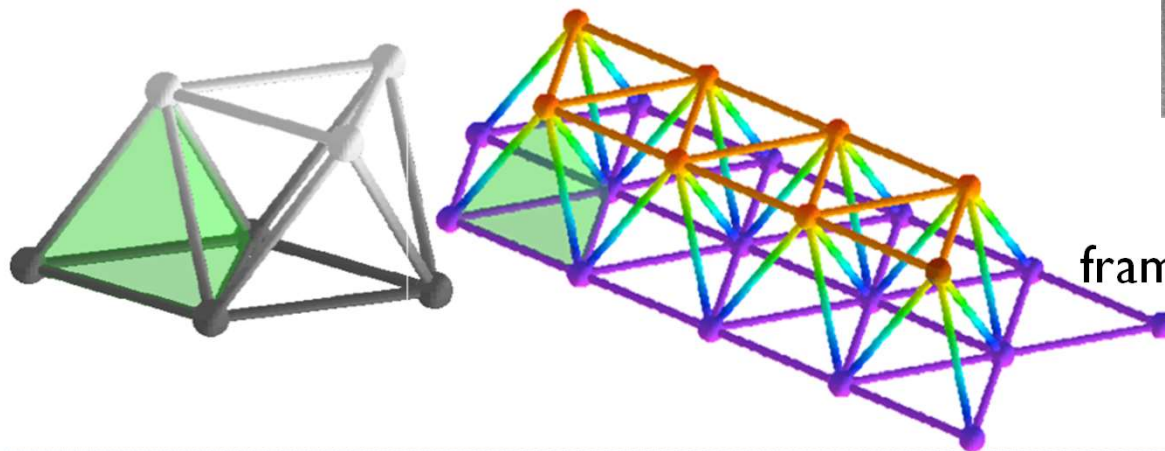
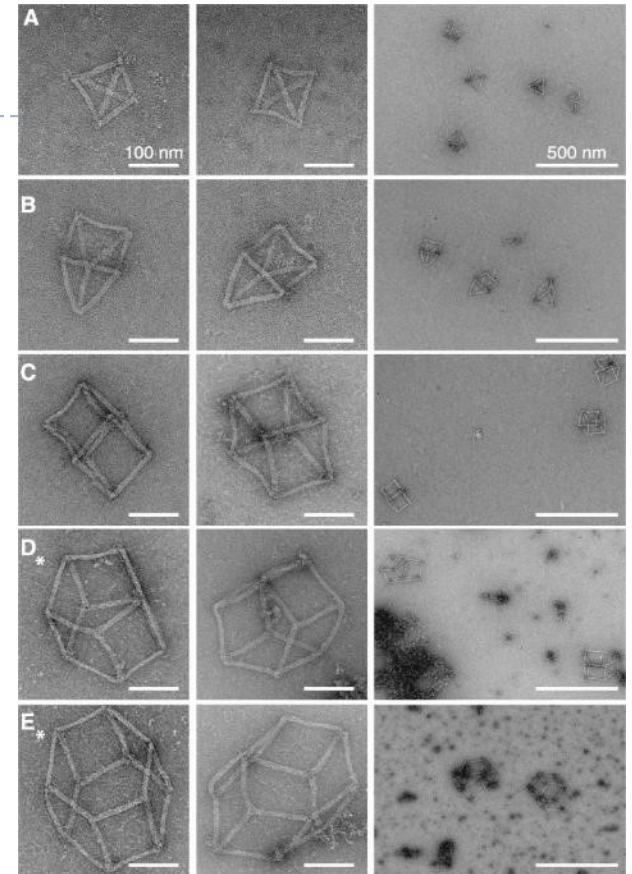
<b>Scenario 1</b>	$T_1(G)$ = Minimum number of tile types required if complexes of smaller size than the target graph are allowed
<b>General graph G</b>	The number of different vertex degrees $\leq T_1(G) \leq$ The number of different even vertex degrees + 2*(The number of different odd vertex degrees).
<b>Trees</b>	The number of different vertex degrees $\leq T_1(T) \leq$ The number of different vertex degrees + 1
$C_n$	$T_1(C_n) = 1$
$K_n$	$T_1(K_n) = 1$ if $n$ is even, and $T_1(K_n) = 2$ if $n$ is odd
$K_{n,m}$	$T_1(K_{n,m}) = 1$ if $n=m$ and even, and $T_1(K_{n,m}) = 2$ otherwise
<b>K-regular graphs</b>	$T_1(G) = 1$ if $n$ is even, and $T_1(G) = 2$ if $n$ is odd
<b>Scenario 2</b>	$T_2(G)$ = Minimum number of tile types required if allow complexes of the same size as, but not smaller than, the target graph
<b>Trees</b>	$T_2(T)$ = The number of different lesser size subtree sequences
$C_n$	$T_2(C_n) = \text{ceiling}(n/2)+1$
$K_n$	$T_2(K_n) = 2$ if $n$ is even, and $T_2(K_n) = 3$ if $n$ is odd
$K_{n,m}$	$T_2(K_{n,m}) = 2$ if $\text{gcd}(m,n)=1$ , and $T_2(K_{n,m}) = 3$ if $\text{gcd}(m,n)>1$
<b>Scenario 3</b>	$T_3(G)$ = Minimum number of tile types required if do not allow complexes of the same size as (or smaller than) the target graph
<b>Trees</b>	$T_3(T)$ = the number of induced subtree isomorphisms
$C_n$	$T_3(C_n) = \text{ceiling}(n/2)+1$
$K_n$	$T_3(K_n) = n$
$K_{n,m}$	$T_3(K_{n,m}) = \text{min}(n,m)+1$

Also the Platonic and Archimedean solids



# Rigid Tiles

1. Arms are straight and rigid
2. Arms have integer lengths
3. The positions of the arms are fixed
4. The arms do not bend or twist in order to bond.
5. No molecule has more than 12 or less than 2 arms (branched junction molecules with 2, 4, 5, 6, 8, and 12 arms have been fabricated) .
6. Final DNA structures must be complete.
7. No design may allow structures smaller than the target structure to form.



The octet truss as an ideal framework for this problem = a new setting for grid graph drawing.



# Rigid tiles are especially challenging

We have developed a novel *half-lap splice joint model*, inspired by wood-working techniques to accurately model DNA origami assembly with rigid tiles, and used it to determine provably optimal design strategies for various regular polyhedral cages. (Joint with M. Ferrari and N. Seeman, and students)

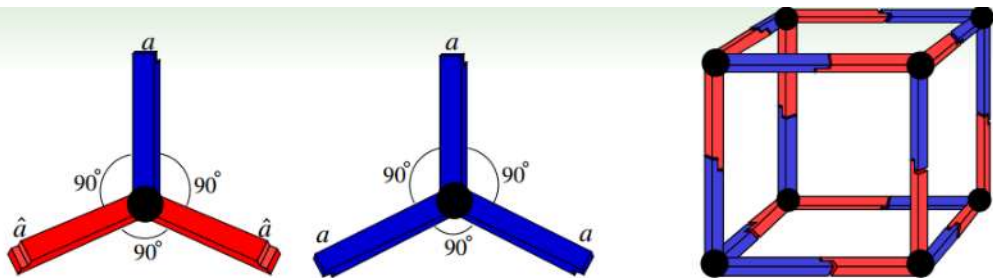


Figure: Tiling the cube

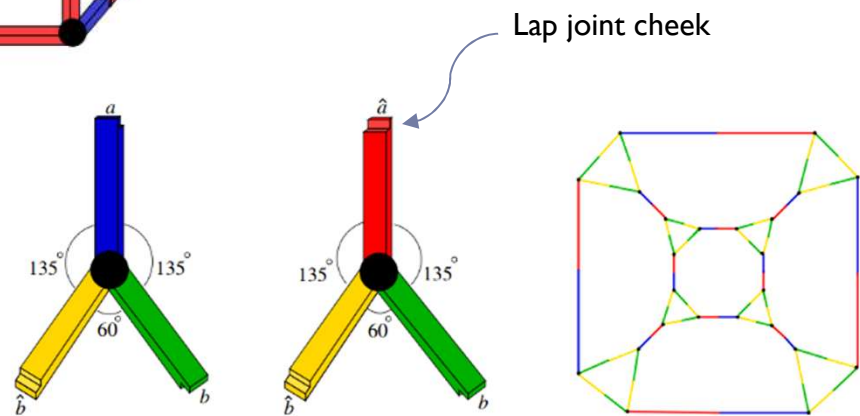


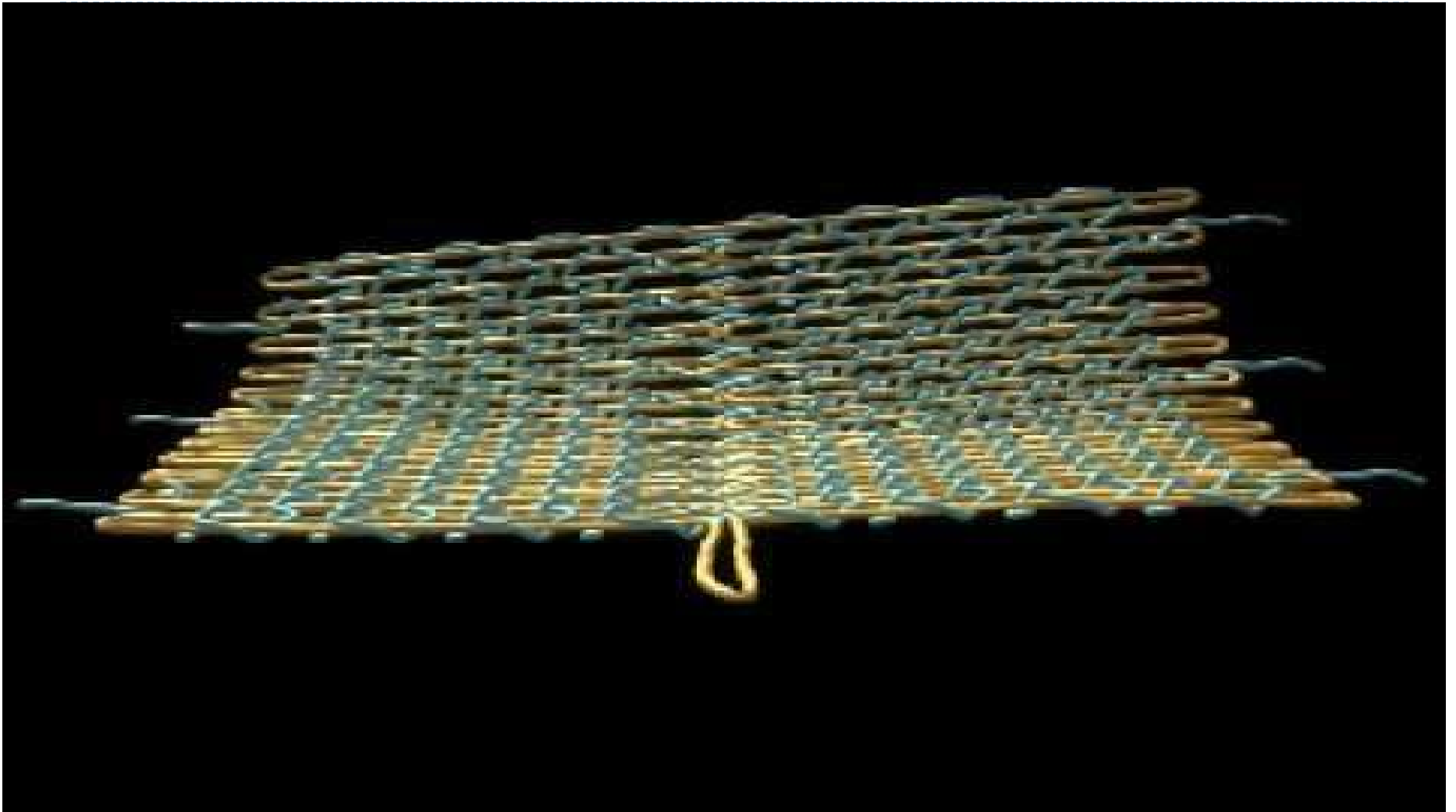
Figure: Tiling the truncated cube





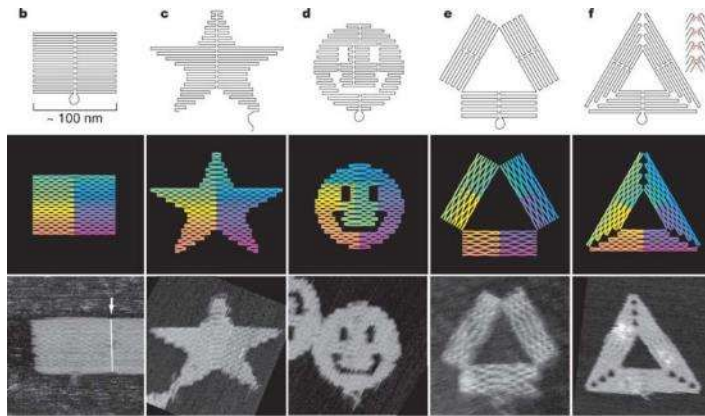
# DNA Origami

Paul W. K. Rothemund Nature 440, 297-302 (16 March 2006)

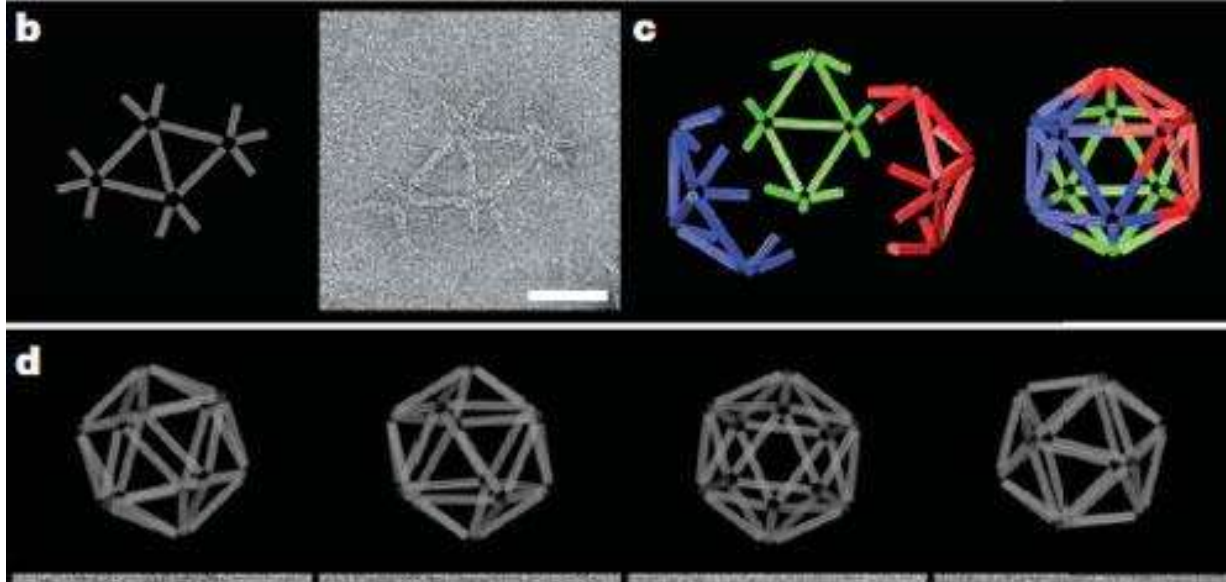


<http://www.youtube.com/watch?v=5yH5LTxFzk&feature=related>

# DNA Origami—new challenges

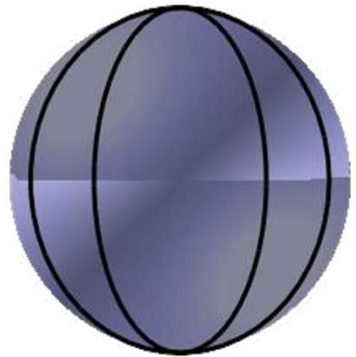


There are techniques, and even software, for flat, filled, objects (which actually come from paper origami)  
Woo and Rothmund, Nature Chemistry, 3, 620



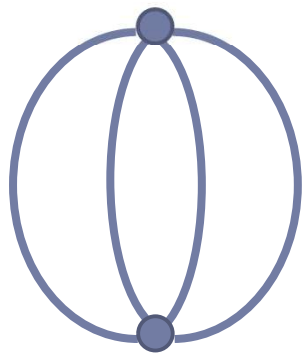
Designs for open structures, e.g. cages, graph-like objects, are much more challenging. This is where our work focuses.

# Basic design process for DNA origami

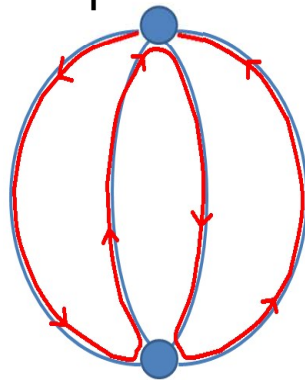


The target

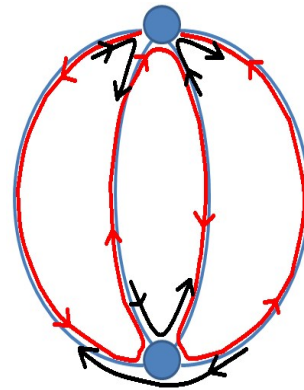
The design steps



Represent the target as a graph



Find an optimal route for the scaffolding strand



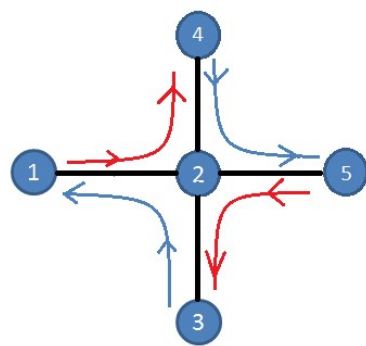
Locate the staple strands

MANY constraints:

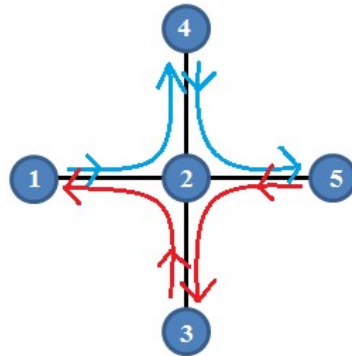
- ▶ Just an **Eulerian circuit** if the target happens to be an **Eulerian graph**.
- ▶ For non-Eulerian graphs, must adapt either the graph or the circuit to enable the construction.
- ▶ Identify structurally appropriate augmenting edges.
- ▶ **Constraints on turnings**
- ▶ **No interwoven strands**
- ▶ Symmetry preferred
- ▶ Commensurable edge lengths (full turns of DNA)
- ▶ Overall length of strand

# Scaffolding and staple constraints

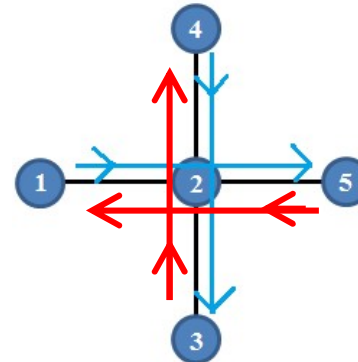
- a) Each edge is covered by a scaffolding and staple strand.
- b) Scaffolding and staples must run in opposite directions (DNA is directed).
- c) Scaffolding and staples cannot cross over each other (DNA strands don't naturally interweave).
- d) The configuration must not disconnect the vertex.



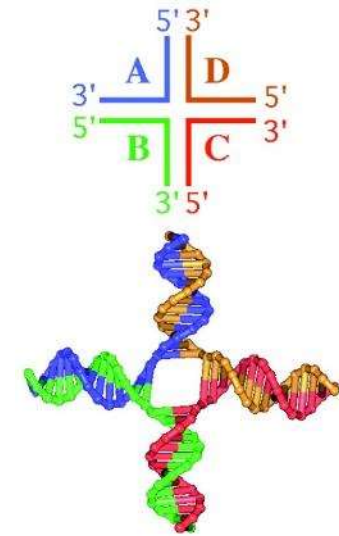
Good



Bad



Bad (crosses and disconnects)



Threading



Staple



# The generic problem from the lab

---

- ▶ We want to build a molecule with the structure of this graph in three space using DNA origami
- ▶ We want the scaffolding strand to follow faces as much as possible, but if it can't, we want it at least not to cross-over or interweave. (And, well, these other bizarre configuration are no good either...)

Please provide the best possible route for the scaffolding strand through the graph.

---



# Again, we prove the problem is NP-hard

---

Reduction to 3-SAT.

Origami methods may not be suitable for efficient biomolecular computing, since they may shift the complexity issue from the computation to the input.

Telling the lab folks that their problem is provably hard doesn't really help them conduct the next experiment.

Need pragmatic solutions.

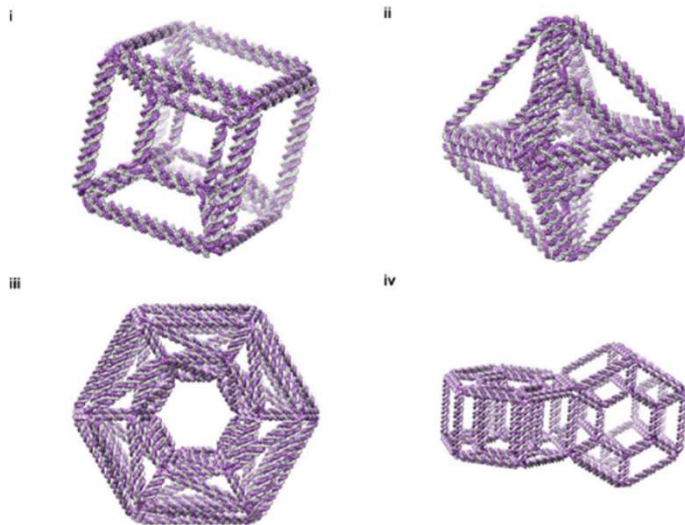
General Principle: An NP-Hard problem is often a 'place marker' for a rich field of related problems— approximation algorithms, special classes that are tractable, tractable variations of the problem, etc.

---

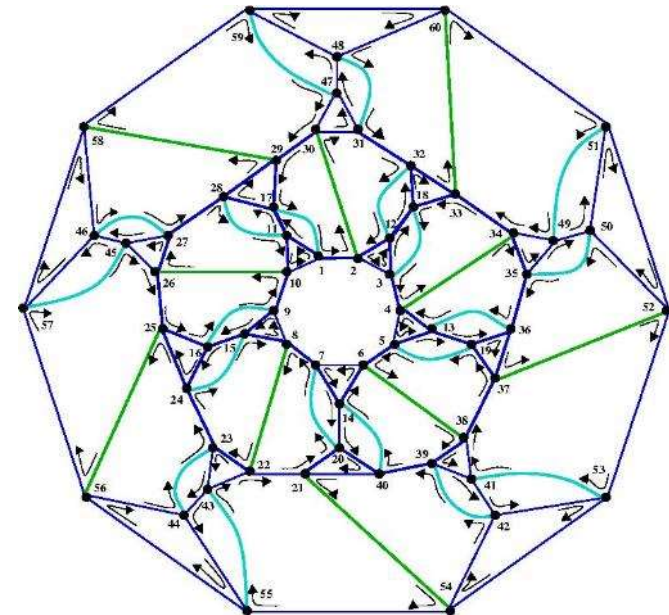


# Other complications: no knotting...

- ▶ Designs are increasingly complex
- ▶ As are topologies



Veneziano et al.  
Designer nanoscale  
DNA assemblies  
programmed from the  
top down, *Science* 2016



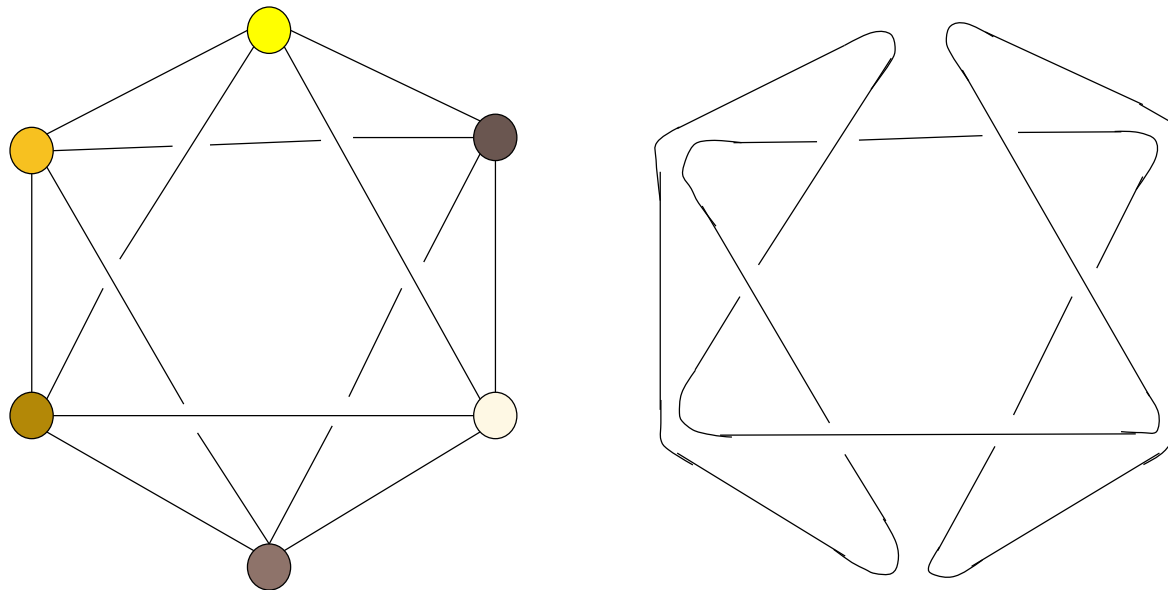
Truncated Dodecahedron

- ▶ Knots can confound assembly, so need to be sure they aren't inadvertently introduced.

## New DNA-driven area: knotted Eulerian circuits

---

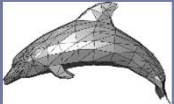


- ▶ The Eulerian circuit needed for DNA origami traces each edge exactly once, but can end up knotted.



Can we prevent this?



# Settings for new area of origami knotting

	Trail	No trail	All trails unknotted	At least one unknotted trail	At least one knotted trail	All trails knotted
<b>Cellularly embedded</b> 	A-trail	Inconstructable	Strongly origami constructable	Origami constructable	Origami surface knotted	Strongly origami surface knotted
<b>Straight-edge</b> 	O-trail	Inconstructable	Strongly origami constructable	Origami constructable	Origami stick knotted	Strongly origami stick knotted
<b>Spatially embedded</b> 	O-trail	Inconstructable	Strongly origami constructable	Origami constructable	Origami knotted	Strongly origami knotted
<b>Abstract Graph</b>	Euler Circuit	Non Eulerian	There exists an embedding with no knotted O-trails: Intrinsically strongly origami constructable	Every embedding has at least one unknotted O-trail: Intrinsically origami constructable	Every embedding has at least one knotted O-trail: Intrinsically Origami knotted	Every embedding has all O-trails knotted: Intrinsically strongly origami knotted

# The routing problem on the backend

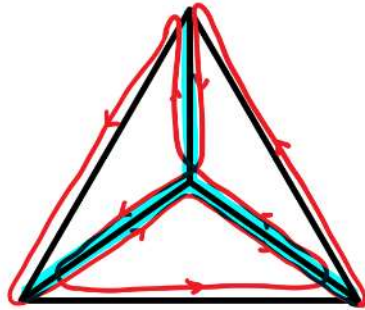
---

- ▶ The preceding were examples of ‘front end’ design problems.
- ▶ But we also need good routes, called **reporter strands** for reading the output from biomolecular computing or other experiments.
- ▶ Reporter strands may be used following other assembly methods, e.g. branched junction molecules, or tiles.
- ▶ The graph may not be Eulerian, and in this case, some edges may need to be repeated.

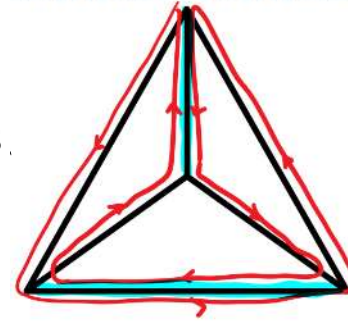
# Example

---

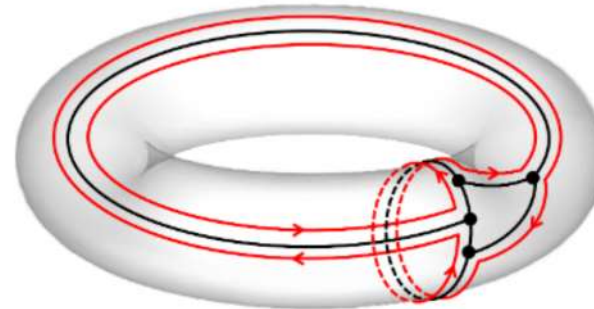
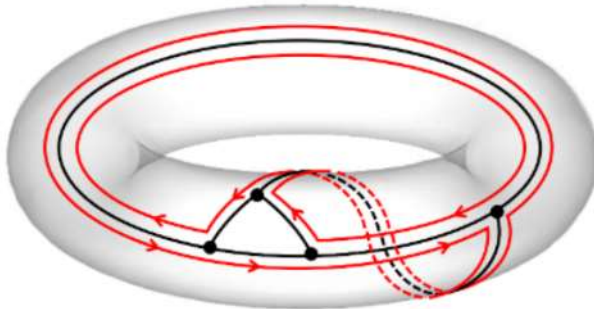
GOOD



Reporter strand walks  
in  $K_4$ : Double traced  
edges are highlighted



BEST



Corresponding edge-outer embedding in the torus, with facial walks around the outer face.

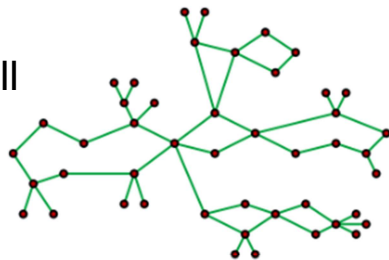
- We want the shortest possible reporter strands
- For cubic graphs, the theoretical minimum is  $|E| + \frac{1}{2} |V|$

# This problem also generated many new questions in graph theory

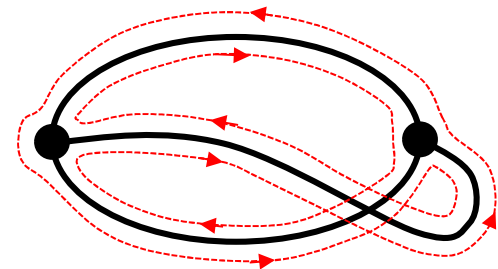
---

- ▶ For both scaffolding and reporting, we need a walk that covers every edge at least once, at most twice, and if twice, once in each direction (DNA is directed).
- ▶ Existing results in the literature are not quite what we need for the DNA application.
  - ▶ Outer planar graphs? No... (need edges, not vertices on the outer face)
  - ▶ Upper embeddable graphs? No... (covers every edge twice, and we want as few as possible double covered)
- ▶ So we need new theory and new results.

An outer planar graph: vertices all on one face



Upper embedding of the theta graph on a torus: only one face



## New directions

---

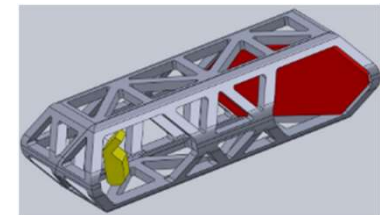
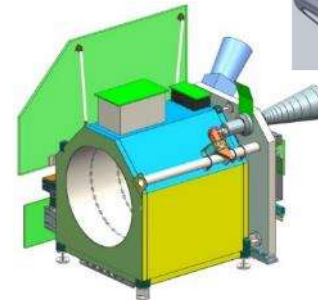
**Edge outer embeddability** is a natural area of investigation with many open questions:

1. Approximation algorithms? Is there an algorithm that will return a route that is within  $x\%$  of minimum length?
2. Special classes of graphs? Are there classes of graphs where it is polynomial time to find a minimum solution? Eulerian graphs are one such class. What others might there be?
3. What can be said about the genus range of embeddings that yield reporter strand walks, or reporter strand walks of minimum length? Are these ranges intervals? (All the usual genus questions can be reformulated for edge outer embeddings)
4. For a given graph, what is the minimum genus surface on which it is edge outer embeddable?

# Graphical self-assembly design is a rich new area

---

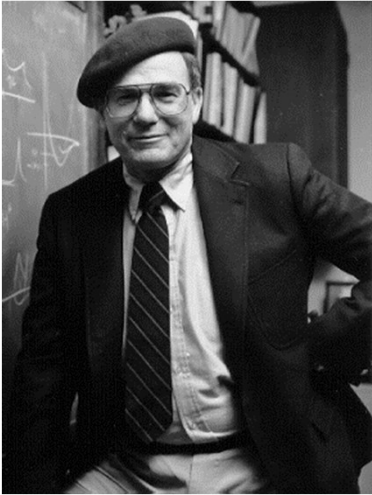
- ▶ DNA self-assembly opens whole new mathematical vistas--
  - ▶ Algorithms, approximations, tractable classes,
  - ▶ General theory (need much more of this!) in several settings
- ▶ The work involves broad synthesis
  - ▶ Mathematics (topology, combinatorics, algebra, geometry, knot theory...)
  - ▶ Operations research and computer science
  - ▶ Chemistry, biology, and physics
  - ▶ Art
  - ▶ Sheer creative ingenuity



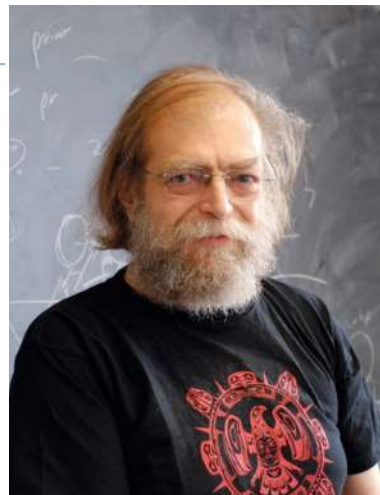
**NASA ATLAS**  
Utilizing the Octet  
Truss in the Design  
of Lateral Transfer  
Retroreflectors



# PI's & senior personnel (design problems)



Bill Goddard (PI)



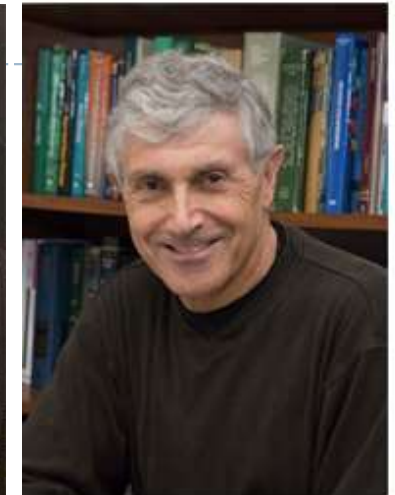
Ned Seeman



Lisa Scherer



Paul Chaikin



John Rossi

Jo Ellis-Monaghan

Greta Pangborn

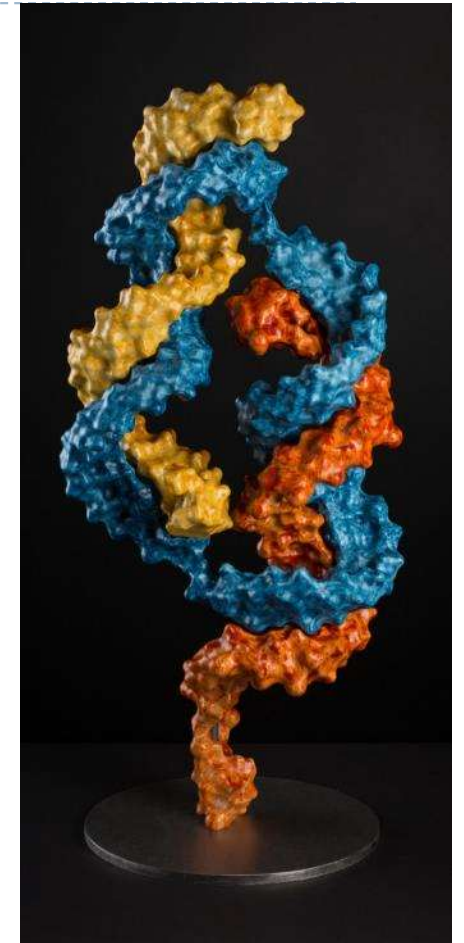
Si-Ping Han

Julian Voss-Andreae

Jim Canary



Get to collaborate with a sculptor!



<https://julianvossandreae.com/>

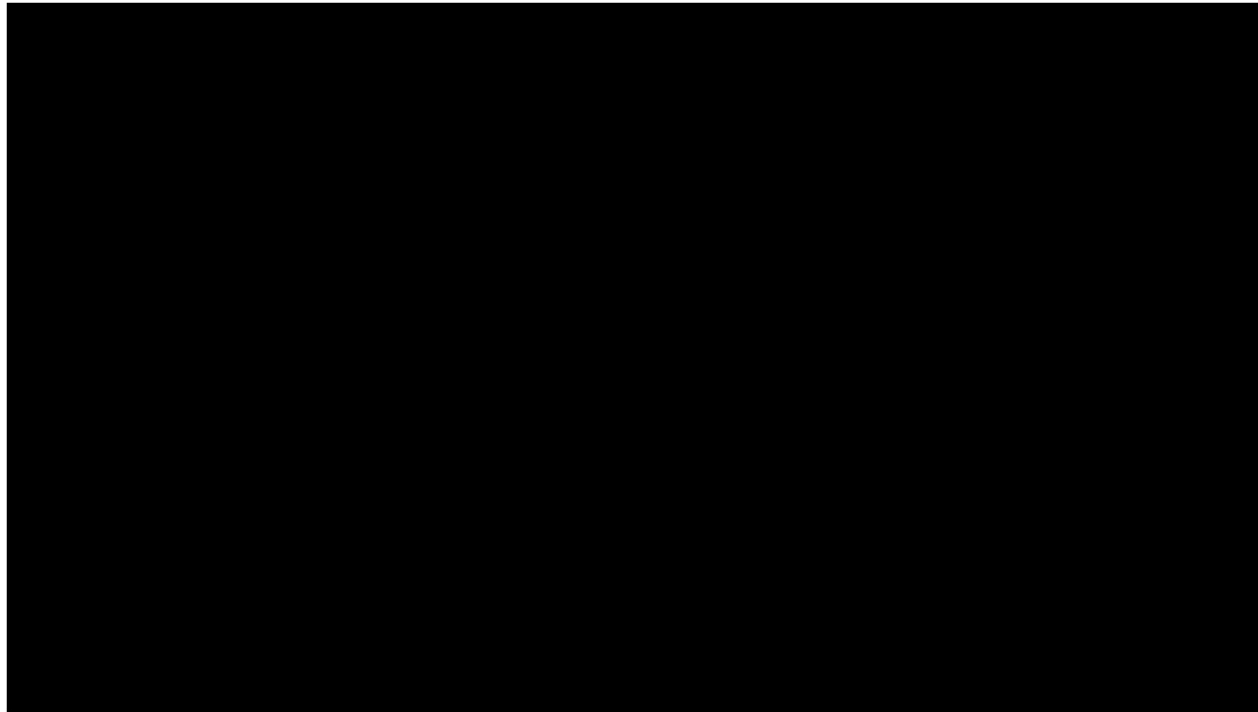


# Current collaboration

---

## Microsystems and MechanoBiology Lab

- Rebecca Taylor (Carnegie Mellon University)
- <http://www.andrew.cmu.edu/user/bex/>
- PNA assembly (a synthetic molecule that bonds more strongly than DNA).



---

▶ <https://www.meche.engineering.cmu.edu/directory/bios/taylor-rebecca.html>



## ICERM-supported REUF Junior Faculty

Jessica Williams

Cory Johnson

Leyda Almodovar Velazquez

Amanda Harsy

# Some recent graduate and undergraduate students

Ada Morse



Rebecca Rouleau,  
Margherita Ferrari



Brenna Smith, Anna Cook, David  
Perry, Jessica Greene

# Self-Assembly Design Strategies

Search this site

- Home
- Background
- Graph Theory
- Conventions
- ▼ Branched Junction Molecule
  - ▶ Flexible Tiles
  - ▶ Rigid Tiles
- Linear Strand
- Our Publications
- Project Participants
- Acknowledgements
- Bibliography
- Contact Us

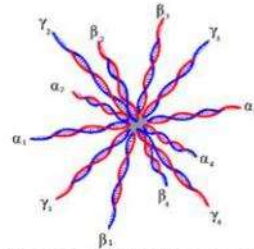
## Home

A number of exciting new laboratory techniques have been developed using the Watson-Crick complementarity properties of DNA strands to achieve the self-assembly of graphical complexes.

For all of these methods, an essential step in building the self-assembling nanostructures is designing the component molecular building blocks.

These design strategy problems fall naturally into the realm of graph theory.

There are many open questions associated with these applications, many of which are accessible to students and offer the possibility of exciting undergraduate research experiences in applied graph theory.



A 12-armed, single-vertex in the octet truss



This work has been supported by the NSF through award 1001408.

Powered by Google Sites

### Contact Us

[Edit sidebar](#)

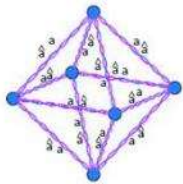
- No branched junction molecule has more than 12 arms or less than 2 arms
- Final DNA structures must be complete
- No design may allow structures smaller than the target structure to form

# Self-Assembly Design Strategies

- Home
- Background
- Graph Theory
- Conventions
- ▼ Branched Junction Molecule
  - ▶ Flexible Tiles
  - ▶ Rigid Tiles
  - Tetrahedron
  - Octahedron**
  - Cuboctahedron
  - Truncated Tetrahedron
  - Truncated Octahedron
- Linear Strand
- Our Publications
- Project Participants
- Acknowledgements
- Bibliography
- Contact Us

[Branched Junction Molecule](#) > [Rigid Tiles](#) >

## Octahedron



Summary of Results for

Octahedron: 1 tile required	
Tile Type Used in Construction	
$\alpha_1$	$\alpha_4$
$a$	$a$

**Theorem** An octahedron may be constructed with a minimum of six tiles with four arms each. We now present a construction

*Proof.* The octahedron is made up of six vertices, twelve edges, and six tiles with four arms each. We now present a construction

# Self-Assembly Design Strategies

Search this site

- Home
- Background
- Graph Theory
- Conventions
- ▼ Branched Junction Molecule
  - ▶ Flexible Tiles
  - ▶ Rigid Tiles
- Linear Strand
- Our Publications
- Project Participants
- Acknowledgements
- Bibliography
- Contact Us

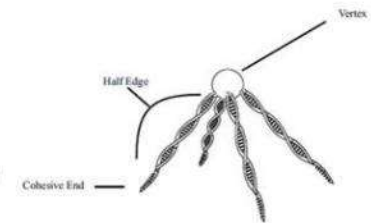
## Branched Junction Molecule

One construction method for self-assembly of DNA structures uses  $k$ -armed branched junction molecules, star-shaped molecules whose centers form the vertices of the structure, and whose arms are double strands of DNA with one strand extending beyond the other. The longer strand forms a cohesive end at the end of the arm that can bond to any other cohesive end with complementary Watson-Crick bases, as in Figure 1, thus forming the edge of the graph.



Figure 1: joining two cohesive ends

**Constructing the Graph** A target graph is constructed from these branched molecules, with a vertex of degree  $k$  realized by a  $k$ -armed branched molecule, and where two vertices in the target graph have an edge between them if and only if their two branched molecules have an arm joined between them via the complementary bases on the cohesive ends. Thus, the joined arms form the edges of the target graph.



A depiction of the typical four-armed tile.

# Some resources

---

- ▶ J. Ellis-Monaghan, N. Jonoska, G. Pangborn, “Tile-based DNA Nanostructures: Mathematical Design and Problem Encoding”, Algebraic and Combinatorial Computational Biology, R. Robeva & M. Macauley, eds. Elsevier. 2019.
- ▶ N. C. Seeman, Structural DNA Nanotechnology, Cambridge University Press, 2016.
- ▶ Y. Ke. Designer three-dimensional DNA architectures. Current opinion in structural biology, 27:122–128, 2014.
- ▶ J. Ellis-Monaghan, G. Pangborn, L. Beaudin, D. Miller, N. Bruno, A. Hashimoto, Minimal Tile and Bond-Edge Types for Self-Assembling DNA Graphs, in Discrete and Topological Models in Molecular Biology, Jonoska & Saito, Eds. . Natural Computing Series, Springer, 2013.
- ▶ J. Ellis-Monaghan, G. Pangborn, “An example of practical organization for undergraduate research experiences,” PRIMUS, 23, no. 9 (2013) 805-814.
- ▶ J. Ellis-Monaghan, G. Pangborn, “Using DNA self-assembly design strategies to motivate graph theory concepts,” Math. Model. Nat. Phenom., 6, no. 6 (2011) 96-107.
- ▶ J. Nangreave, D. Han, Y. Liu, & H. Yan. (2010). DNA origami: A history and current perspective. Current Opinion in Chemical Biology, 14(5), 608-615, 2010.
- ▶ W-Y. Qiu, Z. Wang, & G. Hu. Chemistry & Mathematics of DNA Polyhedra (DNA: Properties and Modifications, Functions and Interactions, Recombination and Applications) (2010).
- ▶ J. Pelesko. Self Assembly: The Science of Things That Put Themselves Together (2007)
- ▶ D. Luo, “The road from biology to materials,” Materials Today, 6 (2003), 38-43



---

THANK YOU!

---

