"Real" proteins fold according to forces: torsion angles potential, van der Waals interactions, hydrogen bonds, hydrophobicity, secondary structure propensity, pairwise specific interactions, ... [Fujitsuka et al. 2004]

HP model of protein folding energetics: [Ken Dill 1990]

simple combinatorial model capturing an often prominant force: hydrophobic bonds (avoid surrounding water)

— protein = chain of (H) or (P) nodes
     hydrophobic   ↳ hydrophilic/polar

— folding = embedding on square grid in 2D or 3D
(also several variations, avoiding parity issue, e.g., equilateral $\triangle$ in 2D & cannonball packing in 3D)

— score = # pairs of (H) nodes adjacent in lattice
(usually ignoring edges of the chain — forced)

— optimal folding = maximum score

Finding optimal folding is NP-hard:

— in 3D [Berger & Leighton 1998]
— in 2D [Crescenz, Goldman, Papadimitriou, Piccalboni, Yannakakis 1998]
— various O(1)-approximations
[Hart & Istrail 1996; Mauri et al. 1999; Newman 2002; Agarwala, Batzoglou, Dancik, Decatur, Farach, Hannenhali, Muthukrishnan, Skiena 1997]

OPEN: $(1-\varepsilon)$-approximation?   APX-hard?

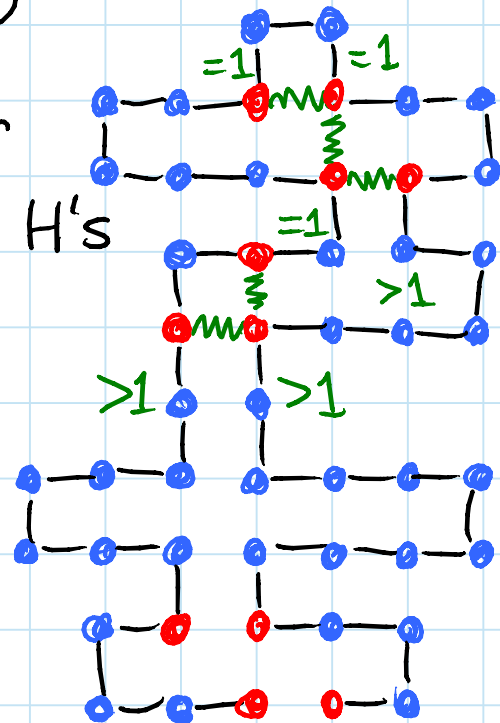# HP model approximation: (square grid)
— upper bound: $2 \cdot \min\{$# even H's, # odd H's$\}$
  — every bond is between odd & even H
  — ≤ 2 bonds per H, but double counting
  $\Rightarrow \leq 2$ bonds per even H, and ditto for odd H
— assume # even H's = # odd H's (excess → P's)
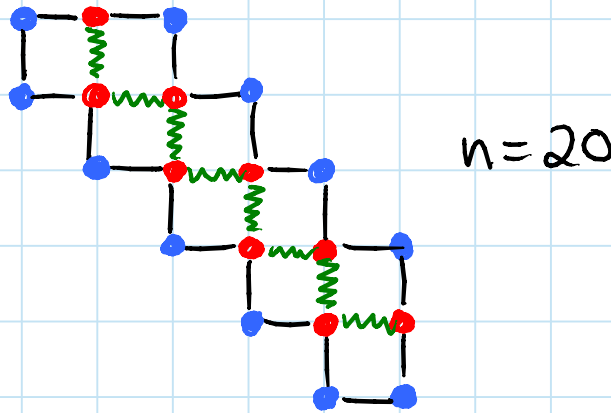
Best approximation: $\frac{1}{3}$ [Newman 2002]
  ~ works even for closed chains (harder)
  — find breakpoint such that
    — proceeding clockwise, always at least
      as many odd H's as even H's
    — proceeding counterclockwise, always at least
      as many even H's as odd H's
  — hang chain down from breakpoint edge,
    with side loops to match up
      even H's of left (ccw) half
      & odd H's of right (cw) half
  — get ≥ 2 bonds for every 3 H's
    —four cases: gap =1 or >1
       for each subchain
  — but ignoring half the H's
  $\Rightarrow \geq 1$ bond per 3 H's
  — upper bound = # H's
  $\Rightarrow \frac{1}{3}$ -approx. □

Unique optimal foldings: [Aichholzer, Bremner, Demaine, Meijer, Sacristán, Soss 2003]

exist in 2D square grid
- for all even n for closed chains
- for doubly even n for open chains

$n = 20$

OPEN : other grids than square? (presumably)

OPEN : optimal folding NP-hard among uniquely optimal proteins? (probably)

OPEN : optimal folding fixed-parameter tractable with respect to #H's? #P's?

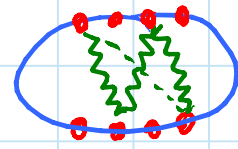OPEN : design protein with unique optimal folding with approximately optimal shape?
- potentially polynomial!

Proof of closed-chain unique optimal foldings:

— zigzag folding achieves score $h-1$

  $\underbrace{}$ # H nodes

— optimal always $\leq h$ (≤ 2 contacts per H, double counting)

— here, optimal $\leq h-1$:
  - ≥ 4 nonconsecutive edges on bounding box, one against each wall
  - ≤ 2 are P-P edges
  - $\Rightarrow \geq 2$ have H nodes
  - these H nodes lose ≥ 1 contact
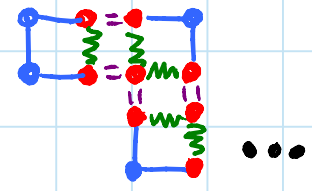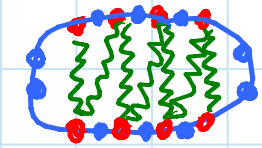 - $\Rightarrow$ lose 1 contact (double counting)

— parity: $h/2$ "even" H nodes, $h/2$ "odd" H nodes
— only bonds between even & odd H nodes
— no contacts external to closed chain
  - P-P edges must be on bounding box
  - partitions odd & even H nodes
— graph of contacts is acyclic
  - bipartite & "inside"
$\Rightarrow$ graph is a path
— path is unique up to reflection
— decomposes into squares:
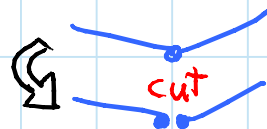
$\Rightarrow$ zigzag

Interlocked 3D chains: [Demaine, Langerman, O'Rourke, Snoeyink 2002 & 2003]

— smallest locked chain has 5 bars (knitting needles), 6 if closed [Cantarella & Johnston 1996]

— can we get away with fewer bars (per chain) if we have more than one chain?

Interlocked set of chains = cannot be separated arbitrarily far by motion avoiding intersection

Motivation: Lubiw's problem [2000]

OPEN: what is the minimum number of vertex "cuts" that suffice to "unlock" any n-bar open chain?

— certainly $\geq \lfloor \frac{n-1}{4} \rfloor$ by repeated knitting needles:

— $\lfloor \frac{n-1}{4} \rfloor$ cuts guarantee that each piece is unlocked but two or more pieces might be interlocked

— best upper bound: $\lfloor \frac{n-3}{2} \rfloor$

Main results on interlockability:

	OPEN CHAINS			CLOSED CHAINS		
OPEN CHAIN	2	3	4	3	4	5
2	sep.*	sep.*	sep.*	sep.	sep.	sep.
3	sep.*	sep.*	interlock	sep.	interlock	interlock
4	sep.*	interlock	interlock	interlock	interlock	interlock

sep.*: even with any number of 2-bar open chains

<u>Also</u>: three 3-bar open chains interlock
  ⇒ complete characterization of interlocking of
    sets of open chains <u>EXCEPT</u>:

OPEN: smallest k for which 2-bar open chain
    interlocks with a k-bar open/closed chain?
  – open 16-chain
    [Glass, Langerman, O'Rourke, Snoeyink, Zhong 2004]
  – open 11-chain, conjectured optimal
    [Glass, Lu, O'Rourke, Zhong 2006]

No finite set of 2-chains can interlock   "hairpins"

... even if 2-chains are held rigid!
- explosion motion: [Dawson 1984; de Bruijn 1954]
  - scale 3D space from origin by factor $1+t$, time $t \geq 0$ (point $p \mapsto (1+t)p$)
  - preserves non-intersection (affine transform) but not the edge lengths: all too long
  - shorten bar lengths to original lengths as time t proceeds
  - subset of scaled version $\Rightarrow$ no intersection
  - chains separate as $t \to \infty$
- key property: 2-chains are <u>starshaped</u> sets (union of line segments from one point)

Alternate view: [Jason Ku]
zoom out to keep middle vertices in place during scaling of just middle vertex coords.
$\Rightarrow$ edge lengths shrink, middle vertices static
$\Rightarrow$ each 2-chain subset of original
$\Rightarrow$ no collision

Two open 3-chains can't interlock

- ... even with finitely many 2-chains added in
- perturb chains so that no nonincident bars are coplanar & no three vertices are collinear
- find plane parallel to two middle bars such that middle bars are on opposite sides
- re-orient to make it the xy plane
- perturb chains so that all z coords. distinct
- explode just in z: $(x, y, z) \mapsto (x, y, (1+t)z)$, $t \geq 0$
- middle bar lengths preserved
- other lengths increase ~ trim appropriately
- preserves no intersection
- all chains separate vertically

4-chain + 2-chains can't interlock: similar

- re-orient to put middle bars in xy plane
- explode just in z

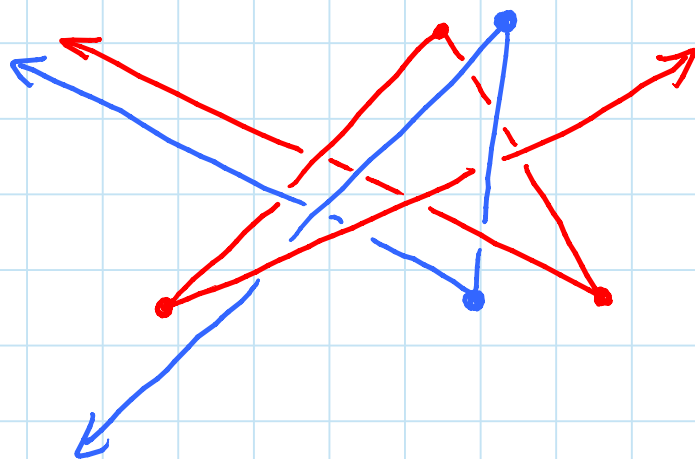$\Rightarrow \lfloor \frac{n-3}{2} \rfloor$ upper bound on Lubiw's problem
- cut 4th joint, then every other joint
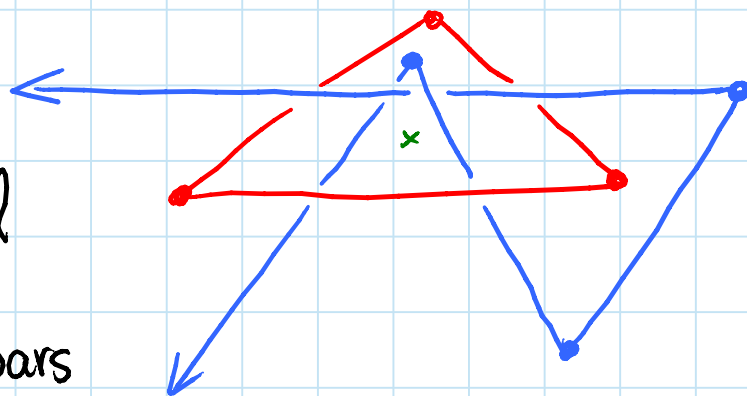
Three 3-chains can interlock:

3-chain can interlock with 4-chain:

- topological argument difficult (impossible?) for two open chains
- instead use <u>geometric argument</u> about bars striking faces of convex hull of middle joints
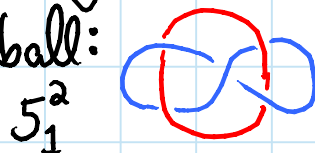
4-chain can interlock with triangle: (closed 3-chain)

- view Δ as fixed
- suppose Δ has circumcircle centered at origin, radius r
- let $R = r + \Sigma$ middle bars of 4-chain

- require end bars of 4-chain to have length $\geq 20R$
- as long as endpoints stay outside ball centered at origin & radius $15R$, and middle bars stay inside ball, can connect endpts. with rope outside ball:

5_1^2

- consider first violation: two cases
  1. if middle vertex about to exit, all middle vertices are $\geq 14R$ away from origin
  2. if endpoint about to enter, adjacent vertex is $\geq 5R$ away from origin $\Rightarrow$ others are $\geq 4R$ away
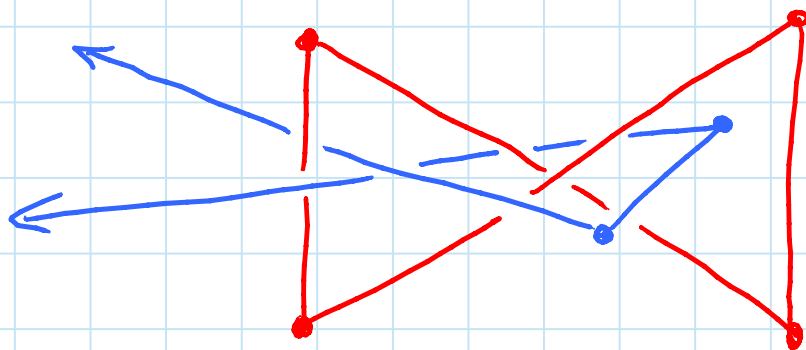- $\Rightarrow$ only end bars can strike the Δ's interior
- three cases:
  1. no end bars strike Δ: $\emptyset_1^2$ $\leftarrow$
  2. one end bar strikes Δ: 2_1^2
  3. both end bars strike Δ: 4_1^2 or $\emptyset_1^2$
     (with some argument)

- none of these match original topology (5_1^2) & topology can't change before such an event. $\square$

3-chain can interlock with quad: (closed 4-chain)

- another
  topological
  argument

Variations:
- rigid chains (like 2-chain result)
- fixed-angle ("revolute") chains

Summary of results:

		2		3			4			5
		flex.	rigid	flex.	fix.ang.	rigid	flex.	fix.ang.	rigid	rigid
2	flex.	−	−	−	−	−	−	−	−	+
	rigid	−	−	−	−	+	+	+	+	+
3	flex.	−	−	−	−	+	+	+	+	+
	fix.ang.	−	−	−	+	+	+	+	+	+
	rigid	−	+	+	+	+	+	+	+	+
4	flex.	−	+	+	+	+	+	+	+	+
	fix.ang.	−	+	+	+	+	+	+	+	+
	rigid	−	+	+	+	+	+	+	+	+
5	rigid	+	+	+	+	+	+	+	+	+

<u>Recall from Lecture 11</u>:

<u>Decision problem</u>: can you fold a linkage
from config. A to config. B?
- PSPACE-complete for 2D trees & 3D chains
[Alt, Knauer, Rote, Whitesides 2004]
- reduction uses locked 2D trees
& interlocked 3D chains (from today)
to form rigid infrastructure for gadgets
- so all locked: hence:

OPEN : characterize locked linkages
e.g. locked trees in 2D or chains in 3D
- polynomially solvable?
- special case: linear trees

6.849 Geometric Folding Algorithms: Linkages, Origami, Polyhedra
Fall 2012