Simulating Constrained Random Walks for Applications to Polymer Models

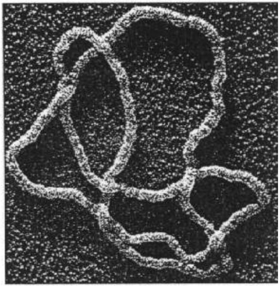
Clayton Shonkwiler, Jason Cantarella, . . .

Colorado State University and University of Georgia
shonkwiler.org and jasoncantarella.com

SIAM LS16
July 14, 2016

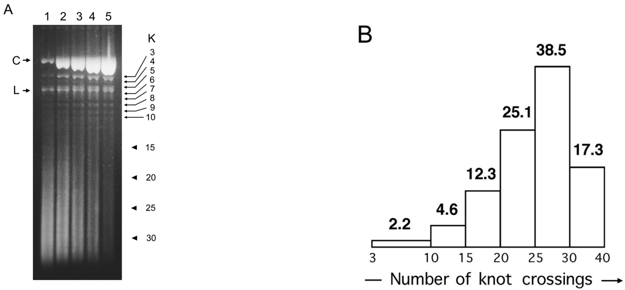
Statistical Physics Point of View

A ring polymer in solution takes on an ensemble of random shapes, with topology (knot type!) as the unique conserved quantity.



Knotted DNA
Wassermann et al.
Science **229**, 171–174

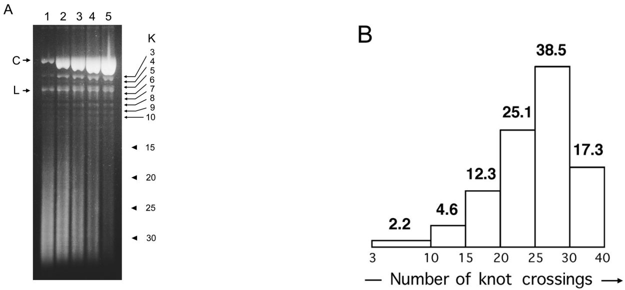
Knot complexity in DNA from P4 tailless mutants



Arsuaga et al., *PNAS* **99** (2002), 5373–5377

How Many Knots?

Knot complexity in DNA from P4 tailless mutants

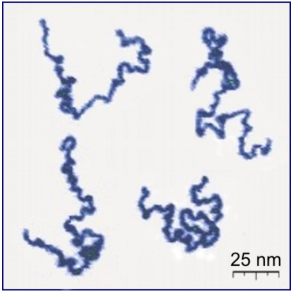


Arsuaga et al., *PNAS* **99** (2002), 5373–5377

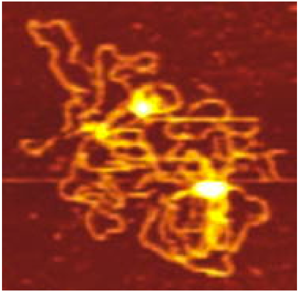
Is this surprising?

Statistical Physics Point of View

A polymer in solution takes on an ensemble of random shapes, with topology as the unique conserved quantity.



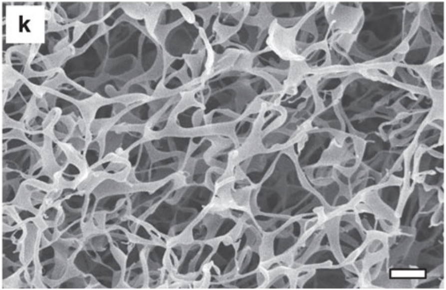
Protonated P2VP
Roiter/Minko
Clarkson University



Plasmid DNA
Alonso-Sarduy, Dietler Lab
EPF Lausanne

Statistical Physics Point of View

A polymer in solution takes on an ensemble of random shapes, with topology as the unique conserved quantity.



Nano-structured hydrogel
Xia et al., *Nat. Commun.* **4** (2013)

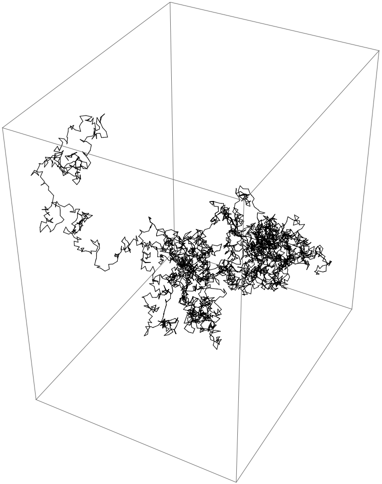
Statistical Physics Point of View

A polymer in solution takes on an ensemble of random shapes, with topology as the unique conserved quantity.

Physics Setup

> *Modern polymer physics is based on the analogy between a polymer chain and a random walk.*
>
> *—Alexander Grosberg*

Simulating Random Walks is Easy...



A random walk with 3500 steps

Sampling Algorithms for (Equilateral) Polygons:

- Markov Chain Algorithms
    - crankshaft (Vologoskii 1979, Klenin 1988)
    - polygonal fold (Millett 1994)
- Direct Sampling Algorithms
    - triangle method (Moore 2004)
    - generalized hedgehog method (Varela 2009)
    - sinc integral method (Moore 2005, Diao 2011)

Which (If Any) is Right?

Table 17 The number of distinct HOMFLY polynomials produced by each algorithm. Since the number of distinct knots sharing the same polynomial is small, this invariant is an suitable surrogate for knot type

Method	Sample size	Distinct HOMFLY
PFM	100 000 000	2 219
CRM	10 000 000	6 110
Hedgehog	10 000 000	1 111
Triangle	10 000 000	3 505

Alvarado, Calvo, Millett, *J. Stat. Phys.* **143** (2011), 102–138

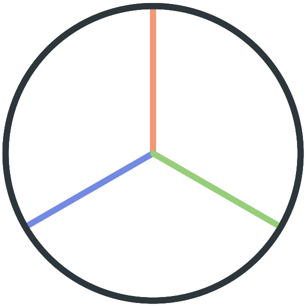
Ansatz

random walk $\iff$ *random point in some (nice!) moduli space*

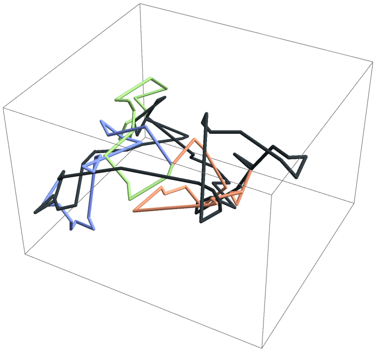
Scientific Idea

Use the (differential, symplectic, algebraic) geometry and topology of these moduli spaces to prove theorems and devise algorithms for studying random walks.

A **topologically constrained random walk** (TCRW) is a collection of random walks in $\mathbb{R}^3$ whose components are required to realize the edges of some fixed multigraph.

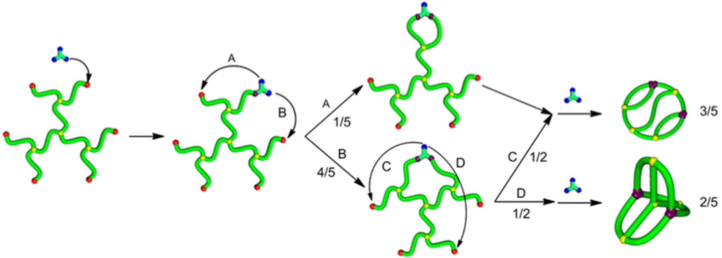


Abstract graph



TCRW

A **topologically constrained random walk** (TCRW) is a collection of random walks in $\mathbb{R}^3$ whose components are required to realize the edges of some fixed multigraph.



Tezuka Lab, Tokyo Institute of Technology

A synthetic $K_{3,3}$!

Random Walk Questions

- What is the joint distribution of steps in a TCRW?
- What can we prove about TCRWs?
  - What is the joint distribution of vertex–vertex distances?
  - What is the expectation of radius of gyration?
  - Spectrum of graph isotopy types?
- How do we sample TCRWs?

Closed Random Walks (a.k.a. Random Polygons)

The simplest multigraph with at least one edge is , which correponds to a *classical* random walk, modeling a *linear* polymer.

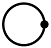
Closed Random Walks (a.k.a. Random Polygons)

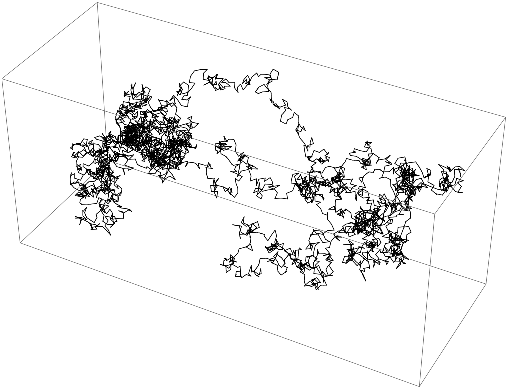
The simplest multigraph with at least one edge is , which correponds to a *classical* random walk, modeling a *linear* polymer.

The next simplest multigraph is , which yields a *closed* random walk (or *random polygon*), modeling a *ring* polymer.

Closed Random Walks (a.k.a. Random Polygons)

The simplest multigraph with at least one edge is , which correponds to a *classical* random walk, modeling a *linear* polymer.

The next simplest multigraph is , which yields a *closed* random walk (or *random polygon*), modeling a *ring* polymer.



Let Arm(n) be the moduli space of random walks in $\mathbb{R}^3$ consisting of n unit steps up to translation.

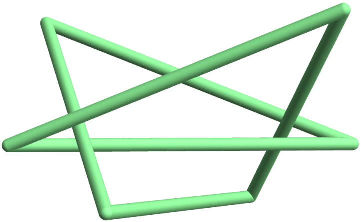
Arm(n) $\cong S^2 \times \ldots \times S^2$.

Let $\mathrm{Arm}(n)$ be the moduli space of random walks in $\mathbb{R}^3$ consisting of n unit steps up to translation.

$\mathrm{Arm}(n) \cong S^2 \times \ldots \times S^2$.

Let $\mathrm{Pol}(n) \subset \mathrm{Arm}(n)$ be the submanifold of closed random walks (or *random polygons*); i.e., those walks which satisfy

$$\sum_{i=1}^{n} \vec{e}_i = \vec{0}.$$



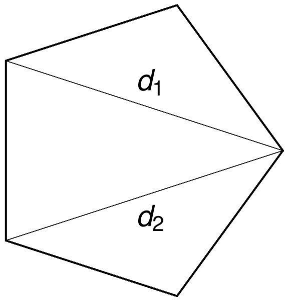
Theorem (with Cantarella and Deguchi; Millett and Zirbel)

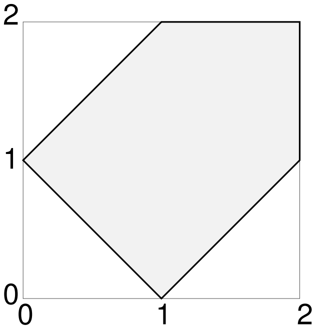
The expected value of the squared radius of gyration for an equilateral n-gon is

$$\mathbb{E}\left[\mathit{Gyradius}, \mathrm{Pol}(n)\right] = \frac{n+1}{12}.$$

Definition

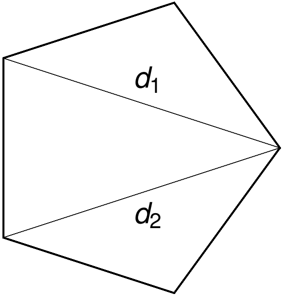
An abstract triangulation T of the n-gon picks out $n-3$ nonintersecting chords. The lengths of these chords obey triangle inequalities, so they lie in a convex polytope in $\mathbb{R}^{n-3}$ called the *triangulation polytope* $\mathcal{P}_n$.

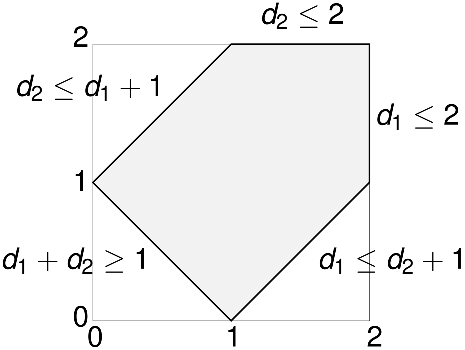




Definition

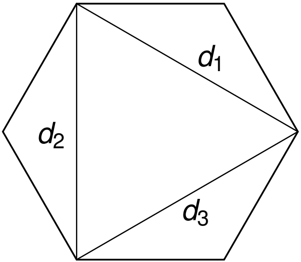
An abstract triangulation T of the n-gon picks out $n-3$ nonintersecting chords. The lengths of these chords obey triangle inequalities, so they lie in a convex polytope in $\mathbb{R}^{n-3}$ called the *triangulation polytope* $\mathcal{P}_n$.

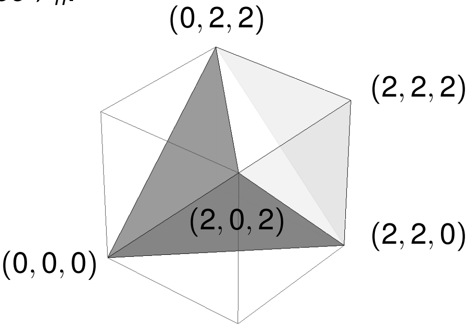




Definition

An abstract triangulation T of the n-gon picks out $n-3$ nonintersecting chords. The lengths of these chords obey triangle inequalities, so they lie in a convex polytope in $\mathbb{R}^{n-3}$ called the *triangulation polytope* $\mathcal{P}_n$.

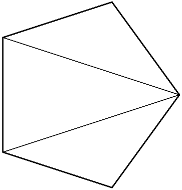


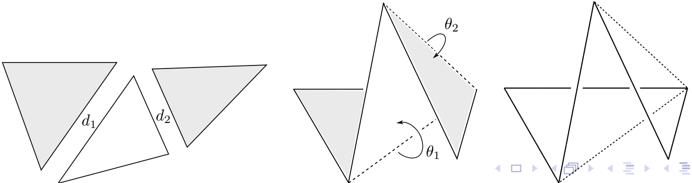


Definition

If $\mathcal{P}_n$ is the triangulation polytope and $T^{n-3} = (S^1)^{n-3}$ is the torus of $n-3$ dihedral angles, then there are *action-angle coordinates*:

$$\alpha\colon\ \mathcal{P}_n \times T^{n-3} \to \mathrm{Pol}(n)/\,\mathrm{SO}(3)$$





Theorem (with Cantarella)

α *pushes the* ***standard probability measure*** *on* $\mathcal{P}_n \times T^{n-3}$ *forward to the* ***correct probability measure*** *on* $\operatorname{Pol}(n)/\operatorname{SO}(3)$.

Theorem (with Cantarella)

*α pushes the **standard probability measure** on $\mathcal{P}_n \times T^{n-3}$ forward to the **correct probability measure** on $\mathrm{Pol}(n)/\,\mathrm{SO}(3)$.*

Corollary

Any sampling algorithm for $\mathcal{P}_n$ is a sampling algorithm for closed polygons.

Theorem (with Cantarella)

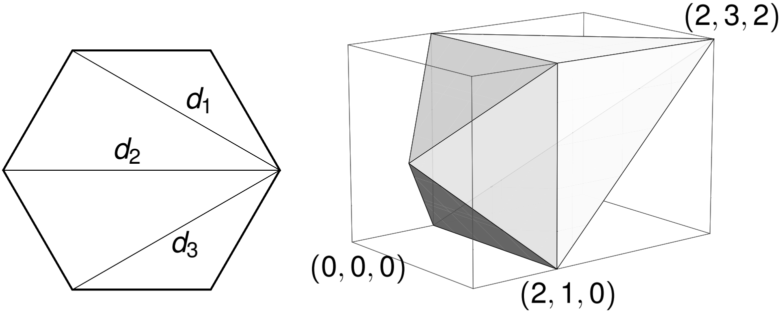
*α pushes the **standard probability measure** on $\mathcal{P}_n \times T^{n-3}$ forward to the **correct probability measure** on $\operatorname{Pol}(n)/\operatorname{SO}(3)$.*

Proof Ideas.

$\operatorname{Pol}(n)/\operatorname{SO}(3)$ is (almost) a toric symplectic manifold; chordlengths and dihedral angles give action-angle coordinates, the triangulation polytope is the moment polytope, and the Duistermaat–Heckman theorem guarantees the pushforward measure on the moment polytope is uniform. $\square$

Corollary

Any sampling algorithm for $\mathcal{P}_n$ is a sampling algorithm for closed polygons.



The polytope $\mathcal{P}_n$ corresponding to the "fan triangulation" is defined by the triangle inequalities:

$$0 \leq d_1 \leq 2 \qquad \begin{array}{c} 1 \leq d_i + d_{i+1} \\ |d_i - d_{i+1}| \leq 1 \end{array} \qquad 0 \leq d_{n-3} \leq 2$$

A Change of Coordinates

If we introduce fake chordlength $d_0 = 1 = d_{n-2}$, and make the linear transformation

$$s_i = d_i - d_{i-1}, \text{ for } 1 \leq i \leq n-2$$

then $\sum s_i = d_{n-2} - d_0 = 0$, so s_{n-2} is determined by $s_1, \ldots, s_{n-3}$

If we introduce fake chordlength $d_0 = 1 = d_{n-2}$, and make the linear transformation

$$s_i = d_i - d_{i-1}, \text{ for } 1 \leq i \leq n-2$$

then $\sum s_i = d_{n-2} - d_0 = 0$, so s_{n-2} is determined by $s_1, \ldots, s_{n-3}$ and the inequalities

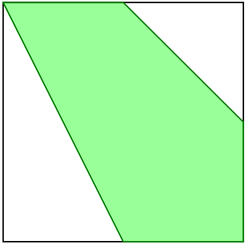
$$0 \leq d_1 \leq 2 \qquad \begin{array}{c} 1 \leq d_i + d_{i+1} \\ |d_i - d_{i+1}| \leq 1 \end{array} \qquad 0 \leq d_{n-3} \leq 2$$

become

$$\underbrace{-1 \leq s_i \leq 1, \quad -1 \leq \sum_{i=1}^{n-3} s_i \leq 1,}_{|d_i - d_{i+1}| \leq 1} \quad \underbrace{\sum_{j=1}^{i} s_j + \sum_{j=1}^{i+1} s_j \geq -1}_{d_i + d_{i+1} \geq 1}$$

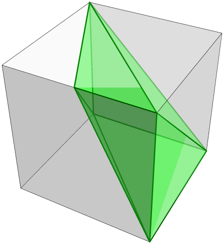
Let $\mathcal{C}_n \subset [-1, 1]^{n-3}$ be determined by the inequalities on the previous slide.

Let $\mathcal{C}_n \subset [-1,1]^{n-3}$ be determined by the inequalities on the previous slide.



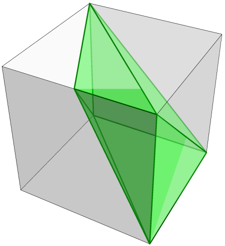
$\mathcal{C}_5$

Let $\mathcal{C}_n \subset [-1, 1]^{n-3}$ be determined by the inequalities on the previous slide.



$\mathcal{C}_6$

Let $\mathcal{C}_n \subset [-1,1]^{n-3}$ be determined by the inequalities on the previous slide.



$\mathcal{C}_6$

Theorem (with Cantarella, Duplantier, Uehara)

The probability that a point in the hypercube lies in $\mathcal{C}_n$ is asymptotic to

$$\frac{6\sqrt{6}}{\sqrt{\pi}} \frac{1}{n^{3/2}}.$$

Action-Angle Method (with Cantarella, Duplantier, Uehara)

1. *Generate* $(s_1, \ldots, s_{n-3})$ *uniformly on* $[-1, 1]^{n-3}$ $O(n)$ *time*
2. *Test whether* $(s_1, \ldots, s_{n-3}) \in \mathcal{C}_n$ *acceptance ratio* $\sim 1/n^{3/2}$
3. *Let* $s_{n-2} = -\sum s_i$ *and change coordinates to get diagonal lengths*
4. *Generate dihedral angles from* T^{n-3}
5. *Build sample polygon in action-angle coordinates*

Action-Angle Method (with Cantarella, Duplantier, Uehara)

1. *Generate* $(s_1, \ldots, s_{n-3})$ *uniformly on* $[-1, 1]^{n-3}$ $O(n)$ *time*
2. *Test whether* $(s_1, \ldots, s_{n-3}) \in \mathcal{C}_n$ *acceptance ratio* $\sim 1/n^{3/2}$
3. *Let* $s_{n-2} = -\sum s_i$ *and change coordinates to get diagonal lengths*
4. *Generate dihedral angles from* T^{n-3}
5. *Build sample polygon in action-angle coordinates*

Expected complexity $\Theta\left(n^{5/2}\right)$

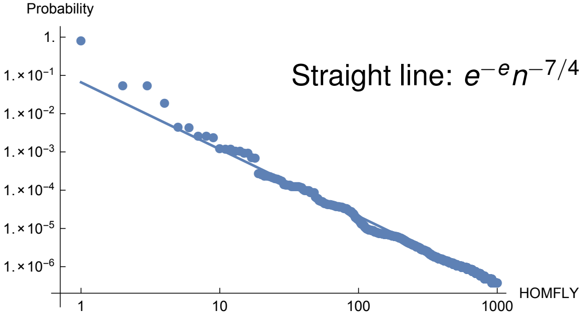
Action-Angle Method (with Cantarella, Duplantier, Uehara)

1. *Generate* $(s_1, \ldots, s_{n-3})$ *uniformly on* $[-1, 1]^{n-3}$ *O*(n) *time*
2. *Test whether* $(s_1, \ldots, s_{n-3}) \in \mathcal{C}_n$ *acceptance ratio* $\sim 1/n^{3/2}$
3. *Let* $s_{n-2} = -\sum s_i$ *and change coordinates to get diagonal lengths*
4. *Generate dihedral angles from* T^{n-3}
5. *Build sample polygon in action-angle coordinates*

Expected complexity $\Theta\left(n^{5/2}\right)$

Implemented in `plcurve`, available at
`http://www.jasoncantarella.com/wordpress/software/`

Knot Types of 10 Million 60-gons

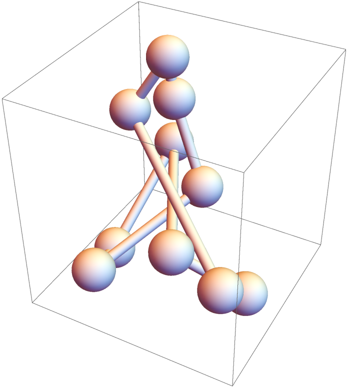


Algorithm	distinct HOMFLYs
PFM[1]	2219
CRM	6110
Hedgehog Method	1111
Triangle Method	3505
Action-Angle Method	$\geq$ 6371

[1] 100 million samples, instead of 10 million.

Future Challenges

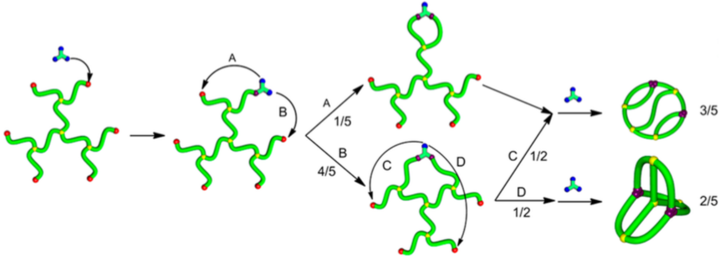
Symplectic Geometry and Excluded Volume



Question

How to incorporate excluded volume into the model?

Topologically Constrained Random Walks



Tezuka Lab, Tokyo Institute of Technology

Question

What special geometric structures exist on the moduli space of topologically-constrained random walks patterned on a given graph?

Thank you!

Thank **you** for listening!

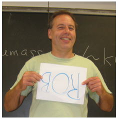
Thanks to my collaborators for being smart and hard-working

Alexander Y Grosberg

Bertrand Duplantier

Erica Uehara

Jason Cantarella



Rob Kusner

Tetsuo Deguchi

Thanks to the Simons Foundation for funding

References

- J. Cantarella, T. Deguchi, and C. Shonkwiler
  Communications on Pure and Applied Mathematics **67** (2014), no. 10, 1658–1699.

- J. Cantarella, A. Y. Grosberg, R. Kusner, and C. Shonkwiler
  American Journal of Mathematics **137** (2015), no. 2, 411–438.

- J. Cantarella and C. Shonkwiler
  Annals of Applied Probability **26** (2016), 549–596.

- J. Cantarella, B. Duplantier, C. Shonkwiler, and E. Uehara
  Journal of Physics A **49** (2016), 275202.

`http://arxiv.org/a/shonkwiler_c_1`

Theorem (Khoi, Takakura, Mandini)

The volume of $\mathcal{C}_n$ is

$$\frac{1}{2(n-3)!}\sum_{k=0}^{\lfloor n/2 \rfloor}(-1)^{k+1}\binom{n}{k}(n-2k)^{n-3}$$

Theorem (Khoi, Takakura, Mandini)

The volume of $\mathcal{C}_n$ *is*

$$\frac{1}{2(n-3)!}\sum_{k=0}^{\lfloor n/2 \rfloor}(-1)^{k+1}\binom{n}{k}(n-2k)^{n-3}$$

Observation (Edwards, 1922)

$$\operatorname{Vol}\mathcal{C}_n = \frac{2^{n-1}}{2\pi}\int_{-\infty}^{\infty}\frac{\sin^n x}{x^{n-2}}\,\mathrm{d}x$$

[T]here is in these days far too great a tendency on the part of teachers to push on their pupils so fast to the Higher Branches of Analysis or to Physical Mathematics that many have neither time nor opportunity for the cultivation of real personal proficiency, or for the acquirement of that individual manipulative skill which is essential to any real confidence of the student in his own power to conduct unaided investigation.

– Joseph Edwards, 1922

Theorem (Khoi, Takakura, Mandini)

The volume of $\mathcal{C}_n$ is

$$\frac{1}{2(n-3)!}\sum_{k=0}^{\lfloor n/2 \rfloor}(-1)^{k+1}\binom{n}{k}(n-2k)^{n-3}$$

Observation (Edwards, 1922)

$$\operatorname{Vol}\mathcal{C}_n = \frac{2^{n-1}}{2\pi}\int_{-\infty}^{\infty}\frac{\sin^n x}{x^{n-2}}\,\mathrm{d}x$$

The Volume of $\mathcal{C}_n$

Theorem (Khoi, Takakura, Mandini)

The volume of $\mathcal{C}_n$ is

$$\frac{1}{2(n-3)!}\sum_{k=0}^{\lfloor n/2 \rfloor}(-1)^{k+1}\binom{n}{k}(n-2k)^{n-3}$$

Observation (Edwards, 1922)

$$\mathrm{Vol}\,\mathcal{C}_n = \frac{2^{n-1}}{2\pi}\int_{-\infty}^{\infty}\frac{\sin^n x}{x^{n-2}}\,\mathrm{d}x$$

$$= \frac{2^{n-1}}{2\pi}\int_{-\infty}^{\infty}\mathrm{sinc}^n(x)x^2\,\mathrm{d}x$$

Making the substitution $x = y/\sqrt{n}$ gives

$$\operatorname{Vol}\mathcal{C}_n = \frac{2^{n-1}}{2\pi}\int_{-\infty}^{\infty} \operatorname{sinc}^n\left({}^{y}\!/\!\sqrt{n}\right)\frac{y^2\,\mathrm{d}y}{n^{3/2}}.$$

Making the substitution $x = y/\sqrt{n}$ gives

$$\mathrm{Vol}\,\mathcal{C}_n = \frac{2^{n-1}}{2\pi}\int_{-\infty}^{\infty} \mathrm{sinc}^n\left({}^{y}\!/\!_{\sqrt{n}}\right)\frac{y^2\,\mathrm{d}y}{n^{3/2}}.$$

But $\mathrm{sinc}\left({}^{y}\!/\!_{\sqrt{n}}\right) = 1 - {}^{y^2}\!/\!_{6n} + o\left({}^{1}\!/\!_{n}\right)$ and
$\lim_{n\to\infty}\left(1 - {}^{a}\!/\!_{n} + o\left({}^{1}\!/\!_{n}\right)\right)^n = e^{-a}$, so we can approximate by

$$\frac{2^{n-1}}{2\pi}\frac{1}{n^{3/2}}\int_{-\infty}^{\infty} e^{-y^2/6}y^2\,\mathrm{d}y = 3\sqrt{\frac{3}{\pi}}\,2^{n-\frac{3}{2}}\,\frac{1}{n^{3/2}}.$$

Making the substitution $x = y/\sqrt{n}$ gives

$$\operatorname{Vol} \mathcal{C}_n = \frac{2^{n-1}}{2\pi} \int_{-\infty}^{\infty} \operatorname{sinc}^n \left({}^{y}\!/\!\sqrt{n}\right) \frac{y^2 \, \mathrm{d}y}{n^{3/2}}.$$

But $\operatorname{sinc}\left({}^{y}\!/\!\sqrt{n}\right) = 1 - {}^{y^2}\!/\!6n + o\left({}^{1}\!/\!n\right)$ and $\lim_{n\to\infty} \left(1 - {}^{a}\!/\!n + o\left({}^{1}\!/\!n\right)\right)^n = e^{-a}$, so we can approximate by

$$\frac{2^{n-1}}{2\pi} \frac{1}{n^{3/2}} \int_{-\infty}^{\infty} e^{-y^2/6} y^2 \, \mathrm{d}y = 3\sqrt{\frac{3}{\pi}}\, 2^{n-\frac{3}{2}} \frac{1}{n^{3/2}}.$$

Therefore,

$$\frac{\operatorname{Vol} \mathcal{C}_n}{\operatorname{Vol} [-1,1]^{n-3}} \sim \frac{3\sqrt{\frac{3}{\pi}}\, 2^{n-\frac{3}{2}} \frac{1}{n^{3/2}}}{2^{n-3}} = \frac{6\sqrt{6}}{\sqrt{\pi}} \frac{1}{n^{3/2}}.$$