

# The spectrum of filament entanglement complexity and an entanglement phase transition

Gregory Buck and Jonathan Simon

*Proc. R. Soc. A* published online 3 October 2012

doi: 10.1098/rspa.2012.0381

---

## References

**This article cites 29 articles, 5 of which can be accessed free**

<http://rspa.royalsocietypublishing.org/content/early/2012/09/25/rspa.2012.0381.full.html#ref-list-1>

## P<P

Published online 3 October 2012 in advance of the print journal.

## Subject collections

Articles on similar topics can be found in the following collections

[applied mathematics](#) (235 articles)

[biochemistry](#) (5 articles)

[topology](#) (2 articles)

## Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

---

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

---

To subscribe to *Proc. R. Soc. A* go to: <http://rspa.royalsocietypublishing.org/subscriptions>

---

# The spectrum of filament entanglement complexity and an entanglement phase transition

BY GREGORY BUCK<sup>1,\*</sup> AND JONATHAN SIMON<sup>2</sup>

<sup>1</sup>*Department of Mathematics, St. Anselm College, Manchester, NH 03102, USA*

<sup>2</sup>*Department of Mathematics, University of Iowa, Iowa City, IA 52242, USA*

DNA, hair, shoelaces, vortex lines, rope, proteins, integral curves, thread, magnetic flux tubes, cosmic strings and extension cords; filaments come in all sizes and with diverse qualities. Filaments tangle, with profound results: DNA replication is halted, field energy is stored, polymer materials acquire their remarkable properties, textiles are created and shoes stay on feet. We classify entanglement patterns by the rate with which entanglement complexity grows with the length of the filament. We show which rates are possible and which are expected in arbitrary circumstances. We identify a fundamental phase transition between linear and nonlinear entanglement rates. We also find (perhaps surprising) relationships between total curvature, bending energy and entanglement.

**Keywords:** polymer; filament; tangling

## 1. Introduction

Topological knot theory is concerned with identifying, recognizing and classifying entanglement patterns (Rolfsen 1976; Adams 1994). The patterns are ideal objects—the filaments have no length or other physical characteristics. Polymer theory brings a different perspective, studying the properties of various models of large collections of randomly arranged filaments (De Gennes 1979; Grosberg & Khokhlov 1994; van Rensburg 2000). Complex properties of polymeric materials and fluids have been modelled, predicted and quantified with considerable accuracy. There has been little commonality between the approaches. Neither tradition has had much to say about ordered or algorithmic entanglement of long filaments or large collections of filaments.

Recently, investigators have begun to find common ground. The topology of DNA has been a fertile ground for collaborations between mathematicians and biologists (Wasserman *et al.* 1985; Sumners 1990; Vologodskii 1992; Nelson & Cox 2000). Chemists, physicists, biologists and mathematicians all use computational tools to do numerical studies of random entanglement. There is a new field of physical knot theory, studying the entanglement of filaments with various physical characteristics (Moffatt 1990; Freedman *et al.* 1994; Ghrist *et al.* 1997; Buck 1998;

\*Author for correspondence ([gbuck@anselm.edu](mailto:gbuck@anselm.edu)).

Buck & Simon 1999; Ricca 2001). In this work, we marry the two approaches to arrive at a theory that allows us to place all filamentary processes (ordered, disordered, thick, thin, compact, extended, stiff, flexible, open-ended, sealed or closed in a loop, etc.) in a single frame.

The plan of this paper is as follows. First, we discuss the measurement of entanglement. Second, we provide elementary examples of filament processes with differing entanglement rates. Third, we introduce power law density measures for filaments. Fourth, we estimate the entanglement for a filament with any given density function. Fifth, we introduce the entanglement phase diagram, place filament processes in the diagram and introduce the notion of expected entanglement. Sixth, we find bounds on the total curvature and the bending energy of the filament, depending on the arclength density, and find the connections between entanglement, total curvature and bending energy.

## 2. The measurement of entanglement

A topological string has no width or length, and can be deformed in any way as long as it is not cut and does not pass through itself. A standard way to measure topological entanglement is the minimum crossing number (MCN), the minimum number of crossings required in any planar diagram of the topological type (Rolfsen 1976; Adams 1994). The MCN can give a false negative in measuring our intuitive idea of entanglement. Tie a knot in a circular rubber band without breaking it. Then it is not topologically knotted, but might be called entangled. But the MCN cannot give a false positive—if the MCN is greater than 0, the filament is entangled under any interpretation of entanglement.

A geometrical string has a particular position in space—a conformation. A geometrical way to measure entanglement of a filament conformation is the average crossing number (ACN). The conformation is projected onto a plane, and the crossings are counted. The average of this number over all such projections is the average crossing number of the conformation. Freedman and He, following Gauss, observed that the ACN had an integral form (Freedman *et al.* 1994),

$$\text{ACN} = \frac{1}{4\pi} \iint_{K \times K} \frac{|\langle x', y', r \rangle|}{d^2}.$$


Here  $x$  and  $y$  are points on the curve  $K$ ,  $x'$  and  $y'$  are unit tangent vectors at  $x$  and  $y$ , respectively,  $d = |x - y|$ ,  $r = (1/d)(x - y)$  and  $|\langle x', y', r \rangle|$  is the norm of the triple scalar product of these three vectors.

If we manipulate a string without cutting it, we change the geometry, and the ACN changes, but the MCN does not. The ACN can be high but the MCN low. The coil of rope in figure 1 and table 1 line N (take the centre curve of the thick rope as the conformation) has high ACN, because, from most viewpoints, it appears to cross itself many times, but one could uncoil it and lay it out in a straight line, and so has  $\text{MCN} = 0$ . The ACN can in this sense give a false positive for topological entanglement, but it cannot give a false negative: the ACN bounds the MCN from above (the average over all viewpoints of a single position must be at least the minimum over all viewpoints of all positions).



Figure 1. Filamentary processes mapped to their  $(\nu, e)$  values, where  $e$  is the exponent for actual entanglement rate. In some cases (marked with asterisks (\*) in table 1), these are best estimates available, and the rest are known values. There are also graphs for the exponents of the estimated entanglement, the total curvature and the bending energy on the same axes. These graphs can be seen individually, with more detail, in the figure shown in §6. In a similar picture, grouping the processes according to total curvature or bending energy—the grouping would change, as can be seen in table 1. The phase change in the estimated entanglement at  $(0.5, 1)$  is apparent. In the shaded rectangle are the possible entanglement values for thick filaments—it is bounded above by  $4/3$  by a packing argument (see text) and below by 0 because it is always possible to have no entanglement. For more on each process depicted, and the letter code for each process (used in the text), see table 1. (Online version in colour.)

Table 1. Table of exponents for filamentary processes. Values marked with asterisks (\*) are best estimates available, the rest are proved values or have strong supporting arguments. Letters on the left are the reference letters in the text. (Online version in colour.)

table of exponents		<i>b</i> , density	<i>v</i> , end-to-end separation	estimated entanglement	actual entanglement	total curvature	total curvature bound	bending energy	bending energy bound	
A		$\infty$	0	2	2	1	1	1	1	random circles in a sphere
B		$\infty$	0	2	2*	1	1	1	1	random points on a sphere connected
C		$\infty$	0	2	2*	1*	1	1	1	Lorenz attractor
D		$0 \geq b$ $b \geq -2$	$1 \geq v$ $v \geq 1/3$	N.B.	$\frac{-v+3}{2}$	$\frac{v+1}{2}$	$-v+1$	$v$	$-2v+1$	M,N link construction
E		0	1/3	4/3	4/3	2/3	2/3	1/3	1/3	cable trefoil
F		0	1/3	4/3	4/3	2/3	2/3	1/3	1/3	iterated simple link
G		0	1/3	4/3	4/3*	1*	2/3	1*	1/3	Hamiltonian random flight, cubic lattice
H		-1.21	0.56	1	$\geq 1$	1*	0.44	1*	-0.12	self-avoiding random flight, cubic lattice
I		-2	1	1	6/5	1	0	1	-1	non-homogeneous density example
J		-2	1	1	1	1	0	1	-1	string of trefoils
K		-2	1	1	1	1	0	1	-1	simple chain
L		0*	1/3*	4/3*	1*	1*	2/3*	1*	1/3*	human chromosome
M		0	1/3	4/3	0	2/3	2/3	1/3	1/3	short braid, open ends
N		0	1/3	4/3	0	2/3	2/3	1/3	1/3	coil of rope
O		-1	1/2	$L \log L$	1	1	1/2	1	0	knit sweater
P		-1	1/2	$L \log L$	$L \log L^*$	1	1/2	1*	0	Gaussian random flight
Q		-1	1/2	$L \log L$	1	1	1/2	1	0	planar weave

(Continued.)

Table 1. (*Continued.*)

---

Notes on the processes:

- A Random circles in sphere. Entanglement rate is provably approximately  $L^2$ ; on average, each circle is linked with a percentage of all the others (see text).
- B Random points on a sphere connected cyclically.  
Each edge appears to cross a percentage of the others, giving an entanglement estimate approximately  $L^2$ . No analytic proof of any actual entanglement rate is known.
- C Lorenz attractor. The filaments are integral curves near the attractor. Our assumption is that each pass near the saddle point links with some percentage of previous passes.  
(In the processes A, B and C, the filaments are infinitely thin; they are given thickness in the figure for illustration.)
- D M, N link construction. N is the number of parallel circles in a group; M is the number of groups (figure 2). The entry for estimated entanglement (in the table denoted N.B.) would be  $-2\nu + 2$  for  $0 \leq \nu \leq 1/2$ , then 1 for  $1/2 \leq \nu \leq 1$ . This is the entanglement estimate curve graphed in figure 1 and the figure shown in §6.
- E Cable trefoil.
- F Iterated simple link. This has the maximum entanglement rate possible with a thick rope, as does the cable trefoil.
- G Hamiltonian random flight on the cubic lattice. Here the strand passes through every point in a box of the cubic lattice. If the strand has thickness, this is maximum density. So the  $ACN \sim L^{4/3}$ . The actual entanglement is an open question. This has been used as a model of a collapsed globule (see text) (Mansfield 1994).
- H Self-avoiding random flight on the cubic lattice. Here the self-avoidance makes the density lower than in the non-lattice random flight.
- I Non-homogeneous density example (see text).
- J String of trefoils—linear entanglement.
- K Simple chain. As  $L$  becomes large, this becomes a closed chain of many small links arranged in a circle. The entanglement is clearly linear in length (the number of links).
- L Human chromosome (see text). The assumption is that the strand is packed tight, but the tangling is restricted (mostly) by the compaction scheme described in the text.
- M Short braid, open ends ( $L$  increases by adding more hairs). If the braid were longer, or closed, the entanglement would not be 0 (see text).
- N Coil of rope.  $ACN \sim L^{4/3}$ , but actual entanglement is 0—as much information as possible with a thick rope.
- O Knit sweater.
- P Gaussian random flight. Thousand steps, edges have no width, so no self-avoidance. For equilateral flights in this case, it has been proven that the  $ACN \sim L \log L$  (Diao *et al.* 2003). For both sorts of random flights, it has been proven that the entanglement is at least approximately  $L$  (Sumners & Whittington 1988). Both  $K$  and  $K_2$  of this process are also at (0.5, 1); so the process has more total curvature and bending energy than is required from the packing requirements.
- Q Planar weave. The entanglement is linear, but the  $ACN \sim L \log L$ . It is perhaps interesting to note that the random flight (not self-avoiding) and the planar weave have the same entanglement estimate. The actual entanglement of the planar weave is less—it is an ordered process.

In all cases, the images are of course of the process at some finite stage, and we are considering the limit rates of the process as  $L \rightarrow \infty$ .

---

No knot in an open string, such as a shoelace or an extension cord, is topologically entangled, because it can be undone by deformation without cutting the string. However, there are often physical or temporal impediments to such a deformation. These impediments usually increase with the length of the strand. Various closing methods for considering the topology of open strands have been used—each method may introduce or bypass crossings that another might not (Millett *et al.* 2005). We will assume that as the length grows the entanglement of an open string is that of the string closed by some method. This assumption is supported by numerical evidence that several closing methods give approximately the same topological entanglement statistics (Lua & Grosberg 2006), and that knots tend to be localized in random strings (Dobay *et al.* 2001; Marcone *et al.* 2005; Millett *et al.* 2005).

A purely topological string has no natural length. But in the physical world, filaments might have thickness, a resistance to bending or be made up of sticks (and so can bend only where the sticks meet). Any of these properties, or many others, can give the filamentary process length  $L$ . Then we can define both  $\text{MCN}(L)$ , the topological entanglement rate, and  $\text{ACN}(L)$ , the apparent entanglement rate.

We can then ask two fundamental questions. Given a filament with such and such characteristics:

- What entanglement rates and patterns are possible?
- What entanglement rates and patterns are expected?

There are a few filamentary processes where we can directly compute  $\text{MCN}(L)$  and  $\text{ACN}(L)$ . In others, we can estimate the  $\text{ACN}(L)$ . With the estimate, we can answer the first question, because the  $\text{ACN}(L)$  bounds the  $\text{MCN}(L)$ . The lower bound is simply 0, because it is always possible to have no topological entanglement.

To answer the second question, we make the assumption that, for a random process, the  $\text{MCN}(L)$  is proportional to the  $\text{ACN}(L)$ ; in other words, the filament is about as tangled as it appears to be. This seems to be a reasonable assumption. Below we show that it holds for several ordered processes (entries D–F, J, K in table 1) and for randomly arranged circles in a finite volume (entry A in table 1). It is consistent with theorems proved about random walks and polygons, specifically that  $\text{MCN}(L)$  grows at least linearly for Gaussian walks and walks on the cubic lattice, and that  $\text{ACN}(L)$  grows like  $L \log L$  for random polygons (Summers & Whittington 1988; Diao *et al.* 2003). The entanglement grows more slowly in a cubic lattice restricted to a slab (Ishihara *et al.* 2012). Certainly some crossings will not be necessary—the  $\text{ACN}(L)$  bounds the  $\text{MCN}(L)$  but is not equal to it. It is the other direction that is the meat of the matter. Why would a random filament appear to be tangled at a greater order than it actually was? We do not know of a case where this happens, but failure of imagination is not a proof. This might not be the kind of thing that can be proved, unless a very precise definition of a random filament is given.

With this assumption, our estimates can give the expected entanglement of the process. If the process does not have this level of entanglement, this is evidence of information in the system.



Figure 2. Nonlinear entanglement rates. The top sequence is described in the text. The bottom two pictures are of the  $M, N$  construction. To the right is the  $M, N$  link,  $N$  parallel circles in a group,  $M$  groups successively linked. Lower left is  $M, N$  construction with knots— $M$  trefoils in a cable of  $N$  strands—this has the same properties and the  $M, N$  link construction, but can be effected with a single string. For links,  $M=2$  is given in figure 1 and table 1 line F,  $N=1$  is given in figure 1 and table 1 line K. Intermediate values allow any entanglement rate between  $4/3$  and  $1$ —all the possible rates greater than  $1$  for thick filaments. To see this, assume  $M=N^a$ , and the individual strands have circular cross-section with radius  $1$ . So if  $a=0$ , the number of ‘links’ or ‘knots’ stays finite while the number of strands in the cable tends to infinity, as in E, F. As  $a$  tends to infinity, we tend towards single strands with many links or knots as in J, K. The cross-section of the cable is approximately  $N^{1/2}$  (assume tight packing). So end-to-end distance  $\nu$  is approximately  $N^{a+1/2}$ . The entanglement is  $MN^2 = N^{a+2}$ . The length  $L \sim$  (‘links’) (number of strands in each link) (length of link)  $= MNN^{1/2} = N^{a+3/2}$ . Then  $\nu = (a + 1/2)/(a + 3/2)$ , and the entanglement exponent  $e = (a + 2)/(a + 3/2)$ ; so we have  $e = -1/2\nu + 3/2$ . This is the line depicted in figure 1 and the figure shown in §6. (Online version in colour.)

### 3. Elementary examples of filaments with differing entanglement rates

#### (a) Example 1

Tie a sequence of overhand knots in a piece of string (entry J in table 1). This is linear entanglement—the number of crossings grows linearly with the length of the string, three crossings for each two inches of string, say.

#### (b) Example 2

Wrap string around your hand several ( $N$ ) times, forming a small open coil. Now, passing through the centre of the coil, wrap around the coil an equal ( $N$ ) number of times (figure 2). This is nonlinear entanglement—each, say, three inches of string passes over several ( $N$ ) strands. If each turn through the coil



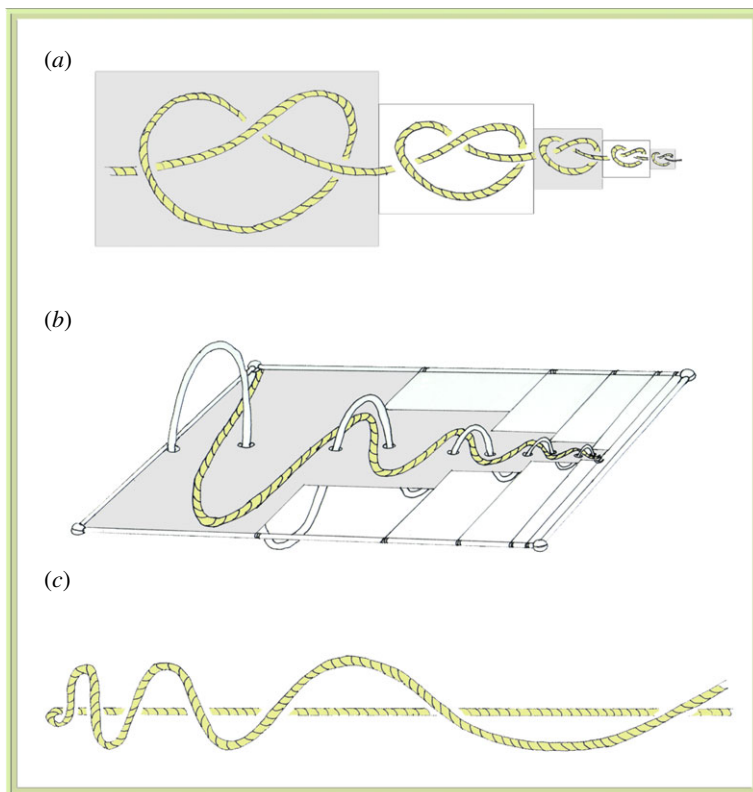


Figure 3. Exponential and zero entanglement. (a) A sequence of trefoils, each succeeding having length  $2/3$  of the preceding. Total length is a geometric series, and so have an infinite number of crossings in finite length. Of course, this cannot be achieved if the string has any thickness. (b) Entanglement of two strands. Entanglement rate is exponential as in (a). Here the curvature is also a geometric series, so this entanglement can be infinite with finite total curvature. (c) Crossings vary with  $L^{1/2}$ , becoming less frequent as  $L$  grows. The crossing strand is always above, so there is no actual entanglement. (Online version in colour.)

is a ring, then we have  $2N$  rings, and half of these rings are linked with the other half. With an infinitely thin string, we could take as many turns as we wanted and have each be less than, say, eight inches long. Then, each eight inches would contribute  $N$  crossings. If  $L$  is the total length, then  $2N = L/8$ , and the  $\text{MCN}(L)$  is of the order of  $N^2$ , but  $N^2 = (L/16)^2$ ; so the  $\text{MCN}(L)$  grows with  $L^2$ . A real string has thickness, so the turns will have to become longer to wind around the volume taken up by previous turns; so  $\text{MCN}(L)$  is of the order of  $L^{4/3}$ , which is the maximum rate possible for filaments of some thickness (Buck 1998; Buck & Simon 1999).

### (c) Example 3

With an infinitely thin filament, we can have arbitrarily large entanglement rates, by using less and less length for each additional bit of entanglement. In figure 3, the rate is exponential.

#### 4. Density measures

To approximate  $\text{ACN}(L)$ , we assume that there is a radial density function  $p(r)$ , giving the density of arclength at a distance  $r$  from an arbitrary test point along the curve. We will also assume that  $p(r)$  is reasonably approximated by a function of the form  $p(r) = cr^b$ , where  $c$  and  $b$  are constants. This power law homogeneity is known to hold for phantom or Gaussian chains—because they have no volume exclusion—no memory of earlier steps—each step is statistically like every other, and the distribution is easy to compute (Sun 1994; Grosberg & Khoklov 1997). Similarly for phantom circles in a finite volume (see below). Self-avoiding walks are a different story, and while homogeneous power law distributions are suspected, they have not been proved to hold (Sun 1994; Grosberg & Khoklov 1997). We will ignore the coefficient  $c$ , focusing only on the exponent  $b$ , thinking that the exponent will give us the most elementary categorization of the processes;  $c$  would certainly be valuable for distinguishing within categories and merits further investigation.

The end-to-end separation  $R$  of a random polymer is often given as a power of  $L$ :  $R \sim cL^\nu$  (De Gennes 1979; Grosberg & Khokhlov 1994; van Rensburg 2000). It is straightforward to see that  $\nu = 0.5$  for phantom chains (in a sense it is just three-dimensional coin flipping) (Sun 1994; Grosberg & Khoklov 1997). If this function gives the spatial separation of any two points along the filament (not just the endpoints), then we have  $b = (1/\nu) - 3$  or  $\nu = 1/(b + 3)$ . If  $L$  denotes total length, then  $L \sim \int_1^R r^b r^2 dr$ , where  $R$  is the radius of the smallest ball containing the filament. (Here and henceforth, use of the tilde means we have quantities of the same order—we drop coefficients.) If we assume  $R \sim L^\nu$ , then we have  $L \sim \int_1^{L^\nu} r^b r^2 dr \sim L^{\nu(b+3)}$ , and so  $\nu(b + 3) = 1$ .

Both  $b$  and  $\nu$  have their uses. It is easier to understand how to estimate the ACN using  $b$ , easier to connect with polymer theory and perhaps easier to present the results with  $\nu$ .

Table 1, and figure 1, have several examples of filaments with values for  $b$  and  $\nu$ . If we have a thick filament (has non-zero volume), then  $-2 \leq b \leq 0$ , and  $1/3 \leq \nu \leq 1$ . The case  $b = 0$ ,  $\nu = 1/3$  is a tight packing of the thick filament. In the case  $b = -2$ ,  $\nu = 1$ , there is only constant arclength in each shell about the test point; this is the density given by a single straight line filament, and so the density cannot be less for a conformation of a single long filament. A phantom Gaussian chain has  $b = -1$ ,  $\nu = 0.5$  (Sun 1994), an estimate for a self-avoiding random walk on the cubic lattice is  $\nu = 0.57$ ,  $b = -1.24$  (Sun 1994).

#### 5. Estimating entanglement

We approximate the ACN via the following formula:

$$\text{ACN} = \frac{1}{4\pi} \iint_{K \times K} \frac{|\langle x', y', r \rangle|}{d^2} \approx L \int_K \frac{|\langle x', y', r \rangle|}{d^2} \approx L \int_{K-\varepsilon} \frac{1}{d^2} \approx L \int_1^R r^b dr.$$

First,

$$\iint_{K \times K} \frac{|\langle x', y', r \rangle|}{d^2} \approx L \int_K \frac{|\langle x', y', r \rangle|}{d^2}.$$

If the process is homogeneous, we can just integrate once at an arbitrary test point and multiply the result by the length of the curve  $L$ . (Below we give an example of a non-homogeneous process and show that the estimate does not hold.)

Next,

$$L \int_K \frac{|\langle x', y', r \rangle|}{d^2} \approx L \int_{K-\varepsilon} \frac{1}{d^2}.$$

The notation  $K - \varepsilon$  signifies that we are not integrating in some neighbourhood of the diagonal—where the points come together along the curve. The triple product in the numerator cancels the singularity there. Because  $x'$ ,  $y'$  and  $r$  are unit vectors, the magnitude of the triple product is at most 1. This gives us left-hand side  $\leq$  right-hand side. For the other direction, we need the assumption that the tangents are, on average, skewed to one another. Without this assumption, the formula gives an upper bound on the entanglement rate. A process of thick filaments with tangents that are on average parallel would have an entanglement rate less than that given by the right-hand side.

Finally,

$$L \int_{K-\varepsilon} \frac{1}{d^2} \approx L \int_1^R r^b dr.$$

Here  $R$  is the radius of the smallest sphere containing the filament, and  $r^b$  the radial density function (at distance  $r$  from a test point, the density is  $r^b$ ). The area of a shell at distance  $r$  is approximately  $r^2$ . So the integrand becomes  $r^2 r^b / r^2 = r^b$ . The basic properties of this remaining integral, properties that are covered in an introductory calculus course, give our entanglement rate measurements and the phase transition. Let  $I = \int_{K-\varepsilon} \frac{1}{d^2} \approx \int_1^R r^b dr$ .  $I$  stands for illumination, because if each point along the curve were a point source, this integral would measure how much they illuminate the test point.

Our general estimate of topological complexity is given by:  $\text{ACN}(L) \sim I \times L$ . There are three cases:

- (i)  $-1 < b$ , or  $\nu < 0.5$ . Then  $I \sim R^{(b+1)}$ ,  $\text{ACN} \sim I \times L \sim R^{b+1} R^{b+3} = R^{2b+4} = L^{(2b+4)/(b+3)} = L^{2-2\nu}$ .
- (ii)  $b = -1$ , or  $\nu = 0.5$ . Then  $I \sim \log(R)$ ,  $\text{ACN} \sim I \times L \sim \log(L^{1/(b+3)}) \times L \sim L \log L$ .
- (iii)  $-2 \leq b < -1$ , or  $1 \geq \nu > 0.5$ . Then  $I \sim (1 - R^{b+1}) \sim 1$ . In this case, the local contribution dominates, and so the approximation of  $I$  at a point is simply a constant. Then  $\text{ACN} \sim I \times L \sim L$ .

## 6. Interpreting the entanglement estimates

The phase transition is clear (figure 1). For  $b > -1$  or  $\nu < 0.5$ , the far-field contribution dominates, and the rate of growth is nonlinear. For  $b < -1$  or  $\nu > 0.5$ , the near-field dominates, and the rate of growth is linear. Other investigators have estimated the entanglement of bond networks and polymer chains with related

approaches (Grassberger 2001; Arteca 2002; Diao *et al.* 2003). Using analysis (as we do, as opposed to numerics), nonlinear growth in entanglement was found in the bond network case, though the exponent was estimated at 1.38, differing from any example discussed here.

The maximum possible  $b$  for thick ropes, or any strand with any sort of non-zero cross-section, is  $b = 0$ . Then, case (i) gives a bound on  $\text{MCN}(L)$  of  $L^{4/3}$ ; we established the bound in this case in earlier work, and showed that it is sharp (Buck 1998; Buck & Simon 1999).

Next, an arbitrarily thin filament with a bound on entanglement through points close in arclength—for example, an equilateral random walk. To get a bound, make the density as great as possible, putting all the arclength a constant distance from the test point. That is, let  $b \rightarrow \infty$ , or  $\nu \rightarrow 0$ . This is case (i); so  $\text{ACN}(L) \sim L^2$ . We have four processes with this growth rate. (i)  $2N$  unit circles, divided into two groups, in each group they are laid upon one another, and the two groups are ‘linked’ (the rings would look somewhat like the example depicted in entry F in table 1, except that here the rings have no thickness, so we could put any number in a finite space). Then  $\text{MCN}(L) \sim N^2$ ,  $L \sim N$ , and, for any test point, the total arclength is on average some constant (approx. 1) away. This example is akin to the first example we discussed (figure 2). (ii)  $N$  circles of radius 1, distributed so that their centres are chosen randomly in a sphere of radius 1, and their orientations are also chosen randomly (entry A in table 1). There is a non-zero probability that any given pair is linked. So, on average, any circle is linked with a percentage of all the others and  $\text{MCN}(L) \sim L^2$ . (iii)  $N$  random points on a sphere of radius 1, connected cyclically (entry B in table 1). This gives a closed chain of  $N$  edges, and  $L \sim N$  (there is an average length for an edge). From an arbitrary perspective, an arbitrary edge appears to cross a percentage of all the other edges. So the number of crossings from an arbitrary perspective is proportional to  $N \times N$ , so  $\text{ACN}(L) \sim N^2 \sim L^2$ . It has not been shown that  $\text{MCN}(L) \sim L^2$  for this construction. (iv) It has been shown that there are knots of arbitrary complexity near the Lorenz attractor (Ghrist *et al.* 1997). The construction appears to have the same sort of patterns as the conformation in figure 2 (top sequence), except that the switching from one loop template to the other is irregular; so it is a reasonable conjecture that there is a family of knots with  $\text{MCN}(L) \sim L^2$ , and that this is the maximal rate.

With the M, N construction, we can create examples for any  $\nu$  value in the thick filament domain (entry D in table 1, figures 2 and 4). The analysis includes the cases usually studied in polymer theory, including the random coil and the random globule. If the coil is a Gaussian phantom chain, then  $b = -1$ , or  $\nu = 0.5$ , which is at the phase transition. The analysis predicts here that  $\text{ACN}(L) \sim L \log L$ . This has been shown, analytically, to be the case for equilateral flights (Diao *et al.* 2003). If the random flight is self-avoiding, then  $b < -1$  and  $\nu > 0.5$ , so the prediction is  $\text{ACN}(L) \sim L$ . For this case, it has been analytically shown that  $\text{MCN}(L) \geq L$  (Summers & Whittington 1988). The globule is often assumed to be a tightly packed conformation of a filament of some thickness—this is  $b = 0$ ,  $\nu = 1/3$ , and so the expected entanglement is  $L^{4/3}$ .

In the literature, the  $\theta$ -transition, from coil to globule, is discussed, and the value  $b = -1$  or  $\nu = 0.5$  is called the  $\theta$  value (Pande *et al.* 1997; Jennings *et al.* 2000; Lucas & Dill 2003; Virnau *et al.* 2005). From our viewpoint, this is either a transition from one  $b$  or  $\nu$  regime to the other, or from the transition value

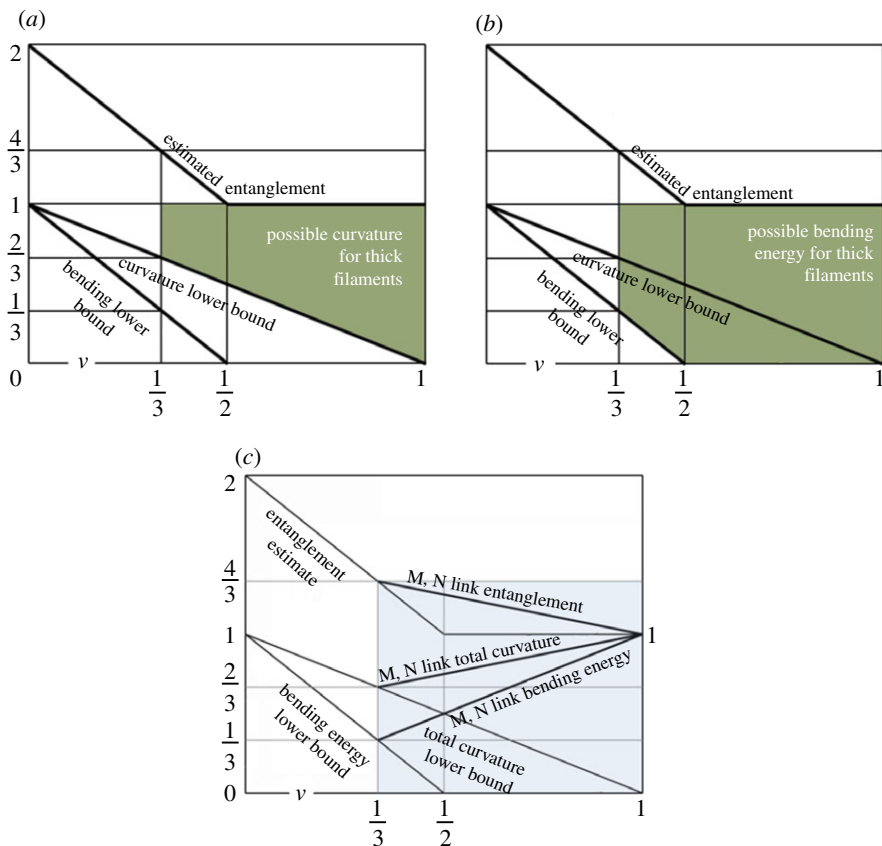


Figure 4. Bounds and possible values. (a,b) Graphs are estimated entanglement, total curvature lower bound and bending energy lower bound; these are normalized log/log plots (see text). Total curvature and bending energy are bounded above by 1 for a thick filament, because the local curvature is bounded for a thick filament by volume exclusion (the curvature cannot be greater than  $1/r$ , where  $r$  is the radius of the rope). The possible values for total curvature and bending energy are shaded. (c) Graph has entanglement, total curvature and bending energy for the M, N construction. Note that they are contained in the respective allowed regions. (Online version in colour.)

$b = -1$ ,  $\nu = 0.5$  to the second regime. This transition may not be meant literally; any single conformation may not change—the values  $\nu$  and  $b$  may be providing a descriptive spectrum for possible conformations.

On the other hand, the coil-globule transition is sometimes thought of as the coil collapse to a globule. If the filament is built in the first regime, then collapsed to the second, and the filament cannot pass through itself, then the topological complexity will be of the order of the first regime (linear) or transition ( $L \log L$ ), not of the second regime ( $L^{4/3}$  for tightly packed thick filaments)—the coil collapsed to a globule does not have the topological entanglement of a randomly built globule. The topological entanglement is less than the expected entanglement, and so there is information in the globule—a memory of the

regime where it was formed. Recent work suggesting the genome is a fractal globule, with less entanglement than the expected entanglement, is an example (Lieberman-Aiden *et al.* 2009).

In general, the phase difference between regimes can be seen from at least three related perspectives.

- (i) *Local/global*. In one regime, the local or near-field contributions (from points near one another in arclength) dominate the ACN. In the other regime, the contributions from pairs distal in arclength, the global or far-field, dominates.
- (ii) *Linear/nonlinear*. In one regime, the entanglement rate is linear (or less). In the other regime, it is greater than linear.
- (iii) *Product structure*. In one regime, the entanglement can often be viewed as the linear topological or geometric sum or product of smaller contributions. In the other regime, the entanglement cannot be decomposed thusly.

A connection between the viewpoints is as follows. If only local contributions ‘matter’, then if we assume homogeneity of the filament, each local segment contributes, on average, a constant, and so we have linear growth in entanglement. For example, if the filaments have any effective thickness, then only small ‘entanglements’ can be made with the finite local length segments. If the local contribution dominates, the total entanglement can be thought of as the sum of these contributions. But such a sum can only give linear growth. So if the growth is nonlinear in this case, it cannot come about from a local product structure—nonlinear growth requires global contributions.

DNA in the eukaryotic cell has several layers of compaction—it is wound around the histones, then packed into the 30 nm fibre, then looped about the nuclear scaffold and then coiled. In this sort of packing, sections of string that are close in space tend also to be close in arclength along the string. This can be compared with simply running the string back and forth lengthwise across the cell, which would tend to bring sections of string together that are far apart in arclength. The cell is a fairly volatile environment. DNA strands suffer breaks and are repaired regularly, and there are enzymes that can pass strands through one another (Grosberg & Khoklov 1997; Nelson & Cox 2000; Lieberman-Aiden *et al.* 2009). In the first packing strategy, crossing changes tend to induce only local or linear entanglement; in the second, crossing changes tend to give global entanglement.

## 7. Expected entanglement

In our sequence of approximations, we made several assumptions, and, at the end of the sequence, we arrived at the expected entanglement exponent, which we denote  $EEE$ . We denote the actual entanglement exponent, the exponent of  $MCN(L)$ , as  $AEE$ . The actual average crossing number exponent is the  $AACNE$ . For most of our examples, we have:  $AEE = AACNE = EEE$ . (We apologize for the inelegant plethora of acronyms, but we cannot seem to devise a better presentation.) Here are a few examples that illustrate how this chain of equalities can fail to hold.

If homogeneity fails, we can have  $AACNE > EEE$ , and  $AEE > EEE$ . Concentrate some entanglement in a small part of the string (entry I in table 1). If a section of length  $L^{5/6}$  is tangled at the  $4/3$  rate, we have  $MCN(L) \sim (L^{5/6})^{4/3} = L^{10/9}$ . Let the rest of the filament be distributed such that  $b = -2$ , or  $\nu = 1$ . Then  $EEE = 1$ , because the section of length  $L^{5/6}$  becomes an arbitrarily small percentage of the total length  $L$  as  $L$  tends to infinity; so our best approximation of a radial density function would have  $b = -2$ . So  $AEE = 10/9$  and  $EEE = 1$ . If the string in the concentrated section of length  $L^{5/6}$  only appears to tangle at the  $4/3$  rate, then we will have  $AACNE > AEE > EEE$ . In general, if homogeneity fails, there are a multitude of possibilities (see figure 3*a, b*, from the discussion of the exponential entanglement rate).

If the tangent vectors to the filament are not randomly distributed, then we can have  $EEE > AACNE, AEE$ . For example, consider figure 3*c*. Here the spacing between the crossings grows with  $L^{1/2}$ . There is no actual entanglement, so we have  $EEE = 1$ ,  $AACNE = 1/2$  and  $AEE = 0$ . Consider a self-avoiding walk on the cubic lattice with the restriction that it cannot take a step in the negative  $x$  direction. It cannot have topological entanglement because it cannot double back on itself. Here,  $ACN(L) \neq 0$ ; so if this is a random filament, it is a counterexample to our theory. But it does violate one sense of a random filament—that it should be free to move in any direction. Perhaps there is an opportunity for an additional theory of directed random filaments, such as random braids.

If homogeneity fails, or if the tangents are not randomly distributed, then it is natural to say that there is geometric information in the system. But it is possible to have information without either of these conditions. Take a long closed untangled loop of rope, and pack it tightly any which way in a box. Then the density will be homogeneous, the tangents non-organized (at least they will appear this way), yet there is information, because the loop cannot be knotted—it went into the box as a closed unknotted loop and was not cut. It seems sensible in this case to call the information topological information. Of all the possible closed loops of long length that could fit in the box, it would seem that the great majority of them would be knotted (although, curiously, this has not been proved), and so there is information. In general, it seems that for a random process the amount of topological information carried by geometric factors declines with length, though not necessarily to 0. This relationship leads to a model for the actions of the topoisomerase II enzyme, which detangles DNA, wherein the enzyme reads the local curvature at juxtapositions to determine global topology (Rybenkov *et al.* 1997; Buck & Zechiedrich 2004).

## 8. Entanglement and curvature

Next we explore the role of curvature in entanglement (Milnor 1950; Chakerian 1964; Van Rensberg & Promislow 1999). Our chief tool here is a natural result: if a long length is packed into a small volume, there must be a significant amount of curvature (Chakerian 1964). This result requires long continuous lengths—one could pack many discrete straight short lengths into a ball, having no curvature at all. Our measurement of curvature of the filament is the total curvature—that is, the integral of the local curvature over the curve. For a chain of unit linear segments, it is the sum of the angles between successive segments.

If a continuous length  $L$  is packed in a ball of radius  $R$ , the total curvature  $K$  is such that  $K \geq (L/R) - 2$ . This is a lower bound on  $K$ ; of course, the filament might curve more than it has to. An example is a random flight with unit length edges. Here the angle between successive edges is on average  $\pi/4$ . Therefore,  $K \sim L$ . But the leading term in the theorem gives  $K \sim L/R$ , and here  $R \sim L^{1/2}$ , and so the bound gives  $K \geq L^{1/2}$ . So the bound holds, but it is not achieved: one could put the same length inside the ball with less total curvature.

Can the entanglement be bounded in terms of  $K$  alone, as we have in terms of length above? This seems plausible: perhaps it takes a certain amount of turning around to entangle. But the answer depends on the  $b$  or  $\nu$  regime.

We have four cases here.

- (i)  $-1 < b$  or  $\nu < 0.5$ . Then  $\text{ACN} \sim R^{2b+4}$ . But  $K \sim R^{b+2}$ . So the bound would be of the form  $\text{ACN} \leq K^2$ .
- (ii)  $b = -1$  or  $\nu = 0.5$ . Then  $\text{ACN} \sim \log(R) \times R^{b+3} = \log(R) \times R^2$ . Here  $K \sim R^{b+2} = R$ , so the bound would have the form  $\text{ACN} \leq \log(K) \times K^2$ .
- (iii)  $-2 < b < -1$  or  $1 < \nu < 0.5$ . Then  $\text{ACN} \sim R^{b+3}$ . And  $K \sim R^{b+2}$  gives  $\text{ACN} \leq K^{(b+3)/(b+2)} = K^{1/(1-\nu)}$ . Note that, as  $b$  tends to  $-2$  or  $\nu$  tends to  $1$ ,  $(b+3)/(b+2)$  and  $1/(1-\nu)$  tend to infinity, so it is not surprising that we have case (iv):
- (iv)  $b = -2$ . Here  $K \sim R^{b+2}$  gives  $K \sim R^0 = \text{const}$ . So we see that, in this case, we can have arbitrarily great topological complexity with total curvature less than a constant; so there can be no general bound in terms of curvature alone.

It was known that something like case (iv) had to exist (Buck & Simon 2007). While it had been shown that any knot requires at least  $4\pi$  in total curvature, there are examples where arbitrarily high entanglement is achieved with finite total curvature. For example, the helix  $(\cos(t), \sin(t), n^2 t)$  for  $0 < t < n$  has total curvature that shrinks with  $1/n$  as  $n$  gets large. Consider the linking of this curve with the long section of the positive  $z$  axis that it goes about. (One can close both curves with finite curvature.) The analysis above tells us that case (iv), when the conformation is long and thin, is the only case where this can happen (Buck & Simon 2007). (By long and thin, we mean that  $b = -2$  requires that the filament increases its distance from any test point at as great a rate as possible.)

The bounds can be applied in the other cases. For a thick rope packed tightly in a sphere ( $b = 0$ ), the entanglement is bounded by  $K^2$ . For a second example, consider an arbitrarily thin wire, with bounded local curvature, packed into a finite radius sphere. We saw earlier that this is the case  $b \rightarrow \infty$ , and so again we have entanglement bounded by  $K^2$ . The principle seems clear: by case (i), as long as the packing is tight enough ( $b > -1$ ), entanglement is bounded by  $K^2$ .

Because a general bound cannot be given in terms of curvature alone, we could try for a bound that is the product of length and total curvature. Since  $L \sim R^{b+3}$  and  $K \geq R^{b+2}$ ,  $L \cdot K \geq R^{2b+5}$ , and for  $-2 \leq b \leq \infty$  this bounds or is equal to the estimates for the ACN in each of the cases. In fact, we can make this perfectly rigorous: we proved (Buck & Simon 2007) that  $\text{MCN} \leq 4 \text{ ropelength } K$ . (Ropelength is defined as  $L/D$ , where  $D$  is the greatest diameter of tube that has the space curve as a centre, which does not self-intersect.)



## 9. Entanglement and bending energy

The bending energy of a filament is the integral of the squared curvature over the filament (a sharp kink gives high bending energy; in a large radius circle, the bending energy is low). The bending energy  $K_2$  of a circle of radius  $R$  is  $1/R$ , so the growth rate necessitated by a continuous filament of length  $L$  in a ball of radius  $R$  is  $(L/R)$  ( $1/R$ ); in terms of  $R$  this is  $K_2 \sim R^{b+1}$ . So it is clear that, for  $b \leq -1$  or  $\nu \geq 0.5$ , there can be no bound for the ACN in terms of  $K_2$ , and it is possible to have finite bending energy for arbitrarily long length. But for  $b > -1$  or  $\nu < 0.5$ , bounds are possible and bending energy must increase with length (figures 1 and 4). For example, for  $b = 0$  or  $\nu = 1/3$ , a tightly packed thick filament (or  $b > 0$  or  $\nu < 1/3$ ), we have that the entanglement is bounded by  $K_2^4$ .

Expected entanglement, curvature and bending energy may each contribute to a theory of the spontaneous formation of knots (Raymer & Smith 2007) in natural systems such as proteins (Sulkowskaa *et al.* 2008).

This work was funded by grants from the National Science Foundation.

## References

- Adams, C. 1994 *The knot book*. New York, NY: W.H. Freeman and Company.
- Arteca, G. 2002 Analytical estimation of scaling behavior for the entanglement complexity of a bond network. *J. Chem. Inf. Comput. Sci.* **42**, 326–330. (doi:10.1021/ci010338s)
- Buck, G. 1998 Four-thirds power law for knots and links. *Nature* **392**, 238–239. (doi:10.1038/32561)
- Buck, G. & Simon, J. 1999 Thickness and crossing number of knots. *Topol. Appl.* **91**, 245–257. (doi:10.1016/S0166-8641(97)00211-3)
- Buck, G. R. & Simon, J. K. 2007 Total curvature and packing of knots. *Topol. Appl.* **154**, 192–204. (doi:10.1016/j.topol.2006.04.002)
- Buck, G. R. & Zechiedrich, E. L. 2004 DNA disentangling by type-2 topoisomerases. *J. Mol. Biol.* **340**, 933–939. (doi:10.1016/j.jmb.2004.05.034)
- Chakerian, G. D. 1964 On some geometric inequalities. *Proc. Amer. Math. Soc.* **15**, 886–888. (doi:10.1090/S0002-9939-1964-0170273-X)
- De Gennes, P. G. 1979 *Scaling concepts in polymer physics*. Ithaca, NY: Cornell University Press.
- Diao, Y., Dobay, A., Kusner, R. B., Millett, K. & Stasiak, A. 2003 The average crossing number of equilateral random polygons. *J. Phys. A: Math. Gen.* **36**, 11 561–11 574. (doi:10.1088/0305-4470/36/46/002)
- Dobay, A., Sottas, P.-E., Dubochet, J. & Stasiak, A. 2001 Predicting optimal lengths of random knots. *Lett. Math. Phys.* **55**, 239–247. (doi:10.1023/A:1010921318473)
- Freedman, M. H., He, Z. X. & Wang, Z. 1994 Mobius energy of knots and unknots. *Ann. Math.* **139**, 1–50. (doi:10.2307/2946626)
- Ghrist, R., Holmes, P. & Sullivan, M. 1997 *Knots and links in three-dimensional flows*. Springer Lecture Notes in Mathematics, vol. 1654. Berlin, Germany: Springer.
- Grassberger, P. 2001 Opacity and entanglement of polymer chains. *J. Phys. A: Math. Gen.* **34**, 9959–9963. (doi:10.1088/0305-4470/34/47/303)
- Grosberg, A. Y. & Khokhlov, A. R. 1994 *Statistical physics of macromolecules*. New York, NY: AIP Press.
- Grosberg, A. Y. & Khokhlov, A. R. 1997 *Giant molecules*. San Diego, CA: Academic Press.
- Ishihara, K. *et al.* 2012 Bounds for the minimum step number of knots confined to slabs in the simple cubic lattice. *J. Phys. A: Math. Theor.* **45**, 065003. (doi:10.1088/1751-8113/45/6/065003)
- Jennings, D. E., Kuznetsov, Y. A., Timoshenko, E. G. & Dawson, K. A. 2000 A lattice model Monte Carlo study of coil-to-globule and other conformational transitions of polymer, amphiphile, and solvent. *J. Chem. Phys.* **112**, 7711–7722. (doi:10.1063/1.481363)

- Lieberman-Aiden, E. *et al.* 2009 Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293. (doi:10.1126/science.1181369)
- Lua, R. C. & Grosberg, A. Y. 2006 Statistics of knots, geometry of conformations, and evolution of proteins. *PLoS Comput. Biol.* **2**, e45. (doi:10.1371/journal.pcbi.0020045)
- Lucas, A. & Dill, K. A. 2003 Statistical mechanics of pseudoknot polymers. *J. Chem. Phys.* **119**, 2414–2421. (doi:10.1063/1.1587129)
- Mansfield, M. 1994 Knots in Hamiltonian cycles. *Macromolecules* **27**, 5924–5926. (doi:10.1021/ma00098a057)
- Marcone, B., Orlandini, E., Stella, A. L. & Zonta, F. 2005 What is the length of a knot in a polymer? *J. Phys. A: Math. Gen.* **38**, L15–L21. (doi:10.1088/0305-4470/38/1/L03)
- Millett, K., Dobay, A. & Stasiak, A. 2005 Linear random knots and their scaling behavior. *Macromolecules* **38**, 601–606. (doi:10.1021/ma048779a)
- Milnor, J. W. 1950 On the total curvature of knots. *Ann. Math.* **52**, 248–257. (doi:10.2307/1969467)
- Moffatt, H. K. 1990 The energy spectrum of knots and links. *Nature* **347**, 367–369. (doi:10.1038/347367a0)
- Nelson, D. & Cox, M. 2000 *Lehninger principles of biochemistry*, 3rd edn. New York, NY: Worth Publishers.
- Pande, V. S., Grosberg, A. Y. & Tanaka, T. 1997 Thermodynamics of the coil to frozen globule transition in heteropolymers. *J. Chem. Phys.* **107**, 5118–5124. (doi:10.1063/1.474875)
- Raymer, D. M. & Smith, D. E. 2007 Spontaneous knotting of an agitated string. *Proc. Natl Acad. Sci. USA* **104**, 16 432–16 437. (doi:10.1073/pnas.0611320104)
- Ricca, R. (ed.) 2001 *An introduction to the geometry and topology of fluid flows*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Rolfsen, D. 1976 *Knots and links*. Berkeley, CA: Publish or Perish Press.
- Rybenkov, V. V., Ullsperger, C., Vologodskii, A. V. & Cozzarelli, N. R. 1997 Simplification of DNA topology below equilibrium values by type II topoisomerases. *Science* **277**, 690–693. (doi:10.1126/science.277.5326.690)
- Scharein, R. 2006 KnotPlot, Software for drawing, visualizing, manipulating, and energy minimizing knots. See <http://www.cs.ubc.ca/nest/imager/contributions/scharein/KnotPlot.html>.
- Sulkowska, J. I., Sukowskic, P., Szymczake, P. & Cieplak, M. 2008 Stabilizing effect of knots on proteins. *Proc. Natl Acad. Sci. USA* **105**, 19 714–19 719. (doi:10.1073/pnas.0805468105)
- Summers, D. W. 1990 Untangling DNA. *Math. Intelligencer* **12**, 71–80. (doi:10.1007/BF03024022)
- Summers, D. W. & Whittington, S. G. 1988 Knots in self-avoiding walks. *J. Phys. A: Math. Gen.* **21**, 1689–1694. (doi:10.1088/0305-4470/21/7/030)
- Sun, S. F. 1994 *Physical chemistry of macromolecules*. New York, NY: John Wiley and Sons.
- van Rensburg, E. J. J. 2000 *Statistical mechanics of interacting walks, polygons, animals and vesicles*. Oxford, UK: Oxford University Press.
- van Rensburg, E. J. J. & Promislow, S. D. 1999 The curvature of lattice knots. *J. Knot Theory Ramifications* **8**, 463–490. (doi:10.1142/S0218216599000328)
- Virnau, P., Mirny, L. A. & Kardar, M. 2005 Knots in globule and coil phases of a medal polyethylene. *J. Am. Chem. Soc.* **27**, 15 102–15 106. (doi:10.1021/ja052438a)
- Vologodskii, A. V. 1992 *Topology and physics of circular DNA*. Boca Raton, FL: CRC Press.
- Wasserman, S., Dungan, J. & Cozzarelli, N. 1985 Discovery of a predicted DNA knot substantiates a model for site specific combination. *Science* **229**, 171–174. (doi:10.1126/science.2990045)