Self-interactions of strands and sheets

Jayanth R. Banavar[1], Oscar Gonzalez[2],

John H. Maddocks[3], Amos Maritan[4]

[1]Department of Physics, 104 Davey Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802 USA

[2]Department of Mathematics, The University of Texas, Austin, Texas 78712 USA

[3] Institut Bernoulli, École Polytechnique Fédérale de Lausanne, CH-1015 Switzerland

[4]International School for Advanced Studies (SISSA), Via Beirut 2–4, 34014 Trieste, Instituto Nazionale per la di Fisica della Materia (INFM) and the Abdus Salam International Center for Theoretical Physics, Trieste, Italy

Keywords: polymer, membrane, surface, protein, self-avoidance

Abstract

Physical strands or sheets that can be modelled as curves or surfaces embedded in three dimensions are ubiquitous in nature, and are of fundamental importance in mathematics, physics, biology and engineering. Often the physical interpretation dictates that self-avoidance should be enforced in the continuum model, i.e. finite energy configurations should not self-intersect. Current continuum models with self-avoidance frequently employ pairwise repulsive potentials, which are of necessity singular. Moreover the potentials do not have an intrinsic length scale appropriate for modelling the finite thickness of the physical systems. Here we develop a framework for modelling self-avoiding strands and sheets which avoids singularities, and which provides a way to introduce a thickness length scale. In our approach pairwise interaction potentials are replaced by many-body potentials involving three or more points, and the radii of certain associated circles or spheres. Self-interaction energies based on these many-body potentials can be used to describe the statistical mechanics of self-interacting strands and sheets of finite thickness.

1 Introduction

The physical world consists of interacting matter. Many systems arising in science can be modelled effectively with a finite number of distinct constituent particles or point masses, $\boldsymbol{q}_i \in \mathbb{R}^3$ $(i = 1, \ldots, N)$. A rich variety of natural phenomena – such as the existence of distinct phases (i.e. solid, liquid, and gas), and transitions between them – may be understood merely on the basis of pairwise interactions in such systems, with the phenomenological pairwise interaction potential being induced by more fundamental interactions at the atomic level. Many-body interactions between triplets or quadruplets of particles are usually only treated as a higher-order correction. When the particles are all identical, it suffices to introduce a single potential energy function $V(\alpha)$ that is defined for all scalar arguments $\alpha > 0$. Then the potential between distinct particles with labels i and j is given by $V(|\boldsymbol{q}_i - \boldsymbol{q}_j|)$, where $|\cdot|$ denotes the usual Euclidean norm. We will be particularly concerned with systems where the potential has a qualitative form akin to that depicted in Figure 1, i.e. strongly repulsive or infinite at short distances, and with a well at some finite distance α_w. Provided that the potential is sufficiently repulsive in the sense that $V(\alpha) \to \infty$ as $\alpha \to 0$, all finite energy configurations have distinct locations for all particles, i.e. $\boldsymbol{q}_i \neq \boldsymbol{q}_j$ for all distinct $i, j = 1, \ldots, N$. Consequently, in any finite energy configuration, the minimum over all pairwise distances $|\boldsymbol{q}_i - \boldsymbol{q}_j|$ yields a distance of closest approach for the configuration that is positive. And it can be anticipated that the location α_w of the potential well provides a characteristic length scale for this distance of closest approach.

We shall consider here the case in which the interacting particles are not unconstrained, but are instead restricted to lie on, or close to, a D-dimensional manifold embedded in $\mathbb{R}^3$, such as a curve ($D = 1$) or a surface ($D = 2$). Such systems are widely studied in many branches of science; examples include polymers [1]–[4], and random surfaces or membranes [5], [6]. The potential that gives rise to the internal stresses that maintain the system close to a continuous manifold is not our concern here – there are many such tethering potentials for curves and surfaces [2]–[6] that are entirely satisfactory. Rather our focus is on the self-interactions that can arise when the curve or surface is sufficiently deformed in $\mathbb{R}^3$ so that very different parts of the manifold come together to form, or are close to forming, a self-intersection. Such self-interactions are not captured by the standard local tethering potentials, and an additional non-local self-interaction potential must be introduced. Our key result is that in order to have a singularity-free description in the continuum limit, the self-interaction potential must depend on $D+2$ or more points. That is, for curves a non-singular self-interaction potential must depend on three or more points (and such non-singular interaction potentials exist), while for surfaces non-singular self-interaction potentials must depend on at least a four-point interaction. Notice that within this framework the standard case of a discrete number of unconstrained particles corresponds to $D=0$, so that pairwise interactions suffice.

The usual descriptions of a self-avoiding curve or surface employ pairwise potentials that are singular and which lack a length scale that can be identified with the physical thickness of the system. For example, a single uniform strand of cooked spaghetti of length L and thickness h would typically be modelled by a curve $\mathcal{C}$, which may be interpreted as the centerline of the spaghetto, together with an effective potential energy [3], [4]

$$E_2[\mathcal{C}] = \int_0^L U(\boldsymbol{q}(s), \boldsymbol{q}'(s), \ldots)\, \mathrm{d}\mathcal{C}_s + \nu \int_0^L \int_0^L \delta(\boldsymbol{q}(s) - \boldsymbol{q}(\sigma))\, \mathrm{d}\mathcal{C}_s\, \mathrm{d}\mathcal{C}_\sigma \quad (1)$$

where U is a specified function, $\boldsymbol{q}(s)$ is a parameterization of the curve, $\mathrm{d}\mathcal{C}$ is an element of arclength, δ is the Dirac delta function, and ν is a model parameter. The first term arises from tethering effects at the microscopic level, and yields effective internal stresses in the strand. The second term, with its singular pairwise potential, models self-avoidance of the strand. But this model does not capture effects of the physical thickness parameter h, which indeed does not even appear. The analysis and numerical treatment of such singular models is plagued by divergences that can only be handled using sophisticated mathematical techniques [7], [8].

In this article we develop models of self-avoiding curves and surfaces that are nonsingular, and which include an explicit thickness length scale. In particular, we replace the singular delta-function potential above by a smooth potential $V(\alpha)$ of the qualitative form shown in Figure 1 that is dependent upon a single scalar variable α. The only difference between our models for discrete sets of points, for curves, and for surfaces, will be our choice for the independent variable α. Moreover, our continuum models retain the following two desirable features of the unconstrained discrete case described above: i) all finite-energy configurations of the manifold are non-self-intersecting, with a positive distance of closest approach, and ii) the location α_w of the well in the potential $V(\alpha)$ provides a characteristic length scale for the distance of closest approach of the manifold. The difficulty in constructing such models is that the standard choice of taking the argument α of the potential to be the Euclidean pairwise distance cannot satisfy our two desiderata when the underlying system is continuous. The reasons are discussed more precisely in

Section 2 below, but the basic idea is simple. For the curve depicted in Figure 2(a), one wishes to penalize true points of closest approach between distinct parts of the curve, as, for example, between points 1 and 4, *without* penalizing adjacent points from the same part of the curve, as for example points 1 and 2. The pairwise Euclidean distance simply cannot distinguish between these two cases; in other words, it cannot distinguish between proximity of points that is forced by continuity of the manifold in *any* configuration, and the real phenomenon of interest, namely proximity of points due to large scale geometrical deformation in *some* configurations.

Our principal result is that the above two desiderata can be achieved simply by taking the argument α to be a quantity other than the Euclidean distance between two points. Specifically, as discussed in Section 2, for the case of curves α can be taken to be the radius of the circle defined by three points. When the argument of $V(\alpha)$ depends upon more than simply two points, for example triplets or quadruplets of points, we shall refer to $V(\alpha)$ as a many-body or multi-point potential. The use of many-body potentials is an essential ingredient in the models of continuous systems that we propose, and they should not be viewed as a higher-order correction to two-body or pairwise potentials, as is the case in discrete models. Indeed our proposal for continuous models is to *replace* pairwise self-interaction potentials, which must be singular, with many-body self-interaction potentials, which need not be singular. For example, for the spaghetto problem, we suggest an effective energy of the form

$$E_3[\mathcal{C}] = \int_0^L U(\boldsymbol{q}(s), \boldsymbol{q}'(s), \ldots)\, \mathrm{d}\mathcal{C}_s + \int_0^L \int_0^L \int_0^L V(r(s,\sigma,\tau))\, \mathrm{d}\mathcal{C}_s\, \mathrm{d}\mathcal{C}_\sigma\, \mathrm{d}\mathcal{C}_\tau \tag{2}$$

where the last term in the more standard energy (1) has been replaced by a three-body potential $V(r(s,\sigma,\tau))$ with V of the form described in Figure 1. Here $r(s,\sigma,\tau)$ is the radius of the circle defined by the three points s, σ and τ as discussed in Section 2. The repulsive potential in (2) can be finite because in the limit of three points coalescing along the curve, for example points 1, 2 and 3 in Figure 2(a), the radius r tends to the local radius of curvature, which is well-defined and positive for twice continuously differentiable curves. (A more detailed discussion of smoothness assumptions is given in Section 2.) However, whenever different parts of a curve come together to form a self-intersection, there are triplets of points, such as 1, 2, and 4 in Figure 2(a), for which the corresponding circle radius r approaches zero. The self-avoidance of the uniform spaghetto of thickness h is modelled by a hard-core potential V which is infinitely large when its argument is less than h and zero otherwise.

Likewise, as described in Section 3, a certain generalized four-body potential can be used to model self-avoiding surfaces of finite thickness.

It should be stressed that the difficulties we address simply do not arise in a many-body system with a discrete index for the particles. On the other hand, analytic treatments of interacting systems with a very large number of particles are often facilitated by making a continuum approximation, in which the discretely indexed particles $\boldsymbol{q}_i$ are replaced by a field $\boldsymbol{q}(x)$ that is dependent upon a continuously varying independent variable x. The appropriate phenomenological interaction potential for the continuum description is to be derived from the microscopic ones. In the case of continuous phase transitions, this is a powerful procedure because the critical behavior is unaffected by the precise microscopic interactions [9], [10], and the critical exponents can be derived using field-theoretic techniques [7], [8]. But it is the passage to the continuum limit which implies that any two-body self-interaction potential must be singular. In order to avoid such singularities we suggest that in the continuum limit the effective potentials modelling non-local self-interactions should be many-body ones.

2 Self-interactions of curves

We consider various distances, other than the usual Euclidean pairwise one, that can be associated with points on a curve. Here and throughout a curve $\mathcal{C}$ will mean a function $\boldsymbol{q}(s) \in \mathbb{R}^3$ of a variable (arc-length) $s \in [0, L]$. We shall consider only sufficiently smooth curves, specifically those that are twice continuously differentiable. This stands in contrast to some models in field theory where polymers are sometimes represented by curves that are continuous but not smooth, for example piecewise linear. In point of fact there is an emerging body of literature [11]–[14] which suggests that our ideas could usefully be applied to a slightly larger class of curves, namely those with only a Lipschitz continuous first derivative, but we do not pursue such questions here. A curve will be called *simple* if it has no self-intersections; that is, if $\boldsymbol{q}(s) = \boldsymbol{q}(\sigma)$ only when $s = \sigma$.

In [15] it was shown that certain *ideal* shapes of knots are related to various physical properties of knotted DNA. Intuitively these ideal configurations can be described as having the property that for a given knot type and prescribed length they are as far as possible from self-intersection. The idea of a three-point distance based on the radius of the associated circle was introduced in [16] as one way to make the notion of distance from self-intersection mathematically precise. In [17] the same three-point circular distance was used as an ingredient in the numerical study of the optimal shapes of compact strings. The properties and relations between all possible circular and spherical distance functions defined on curves are discussed at length in [18].

In the present article we argue that these generalized circular and spherical distances also provide natural means for defining singularity free self-interaction energies of curves through a potential function $V(\alpha)$ with the qualitative form of Figure 1 that takes a multi-point distance as argument. For our purposes, a self-interaction energy will mean a functional $E[\mathcal{C}]$ that is finite for any simple curve $\mathcal{C}$, and which tends to infinity as $\mathcal{C}$ tends to a non-simple curve. We remark that within the specific context of knot theory the construction of simple geometric self-interaction energies for curves has already received much attention; the case of a pure inverse power of a circular three-point distance was proposed in [16], and surveys of alternative approaches can be found in several Chapters of [19].

2.1 Two-point distance for curves

Given an arbitrary simple curve $\mathcal{C}$, the most intuitive approach to the construction of a scalar argument α for a self-interaction energy $V(\alpha)$ is to take the usual pairwise or two-point distance function

$$\beta(s,\sigma) = |\boldsymbol{q}(s) - \boldsymbol{q}(\sigma)|.$$

In particular, a candidate energy functional $E[\mathcal{C}]$ would then be the double integral

$$E[\mathcal{C}] = \iint V(\beta(s,\sigma))\, \mathrm{d}\mathcal{C}_s\, \mathrm{d}\mathcal{C}_\sigma$$

with for example $V(\alpha) = \alpha^{-m}$. The basic idea is that, for $m \geq 2$, the integral tends to infinity as $\boldsymbol{q}(s)$ tends to $\boldsymbol{q}(\sigma)$ with $s \neq \sigma$, thus meeting the infinite-energy condition associated with self-intersections. However, such an integral is *always* divergent due to nearest-neighbor effects since $\beta(s,\sigma) = 0$ when $s = \sigma$, so that the energy $E[\mathcal{C}]$ is infinite for *any* curve $\mathcal{C}$.

To cure the above divergence problem one may consider regularizing the integrand by subtracting something equally divergent as $s \to \sigma$, or mollifying the integrand using a multiplicative factor that tends to zero at an appropriate rate as $s \to \sigma$. (See for example [19] and references therein.) In essence, these procedures introduce a length scale to compensate for the fact that there is no inherent small-distance cutoff for the pairwise distance between nearest-neighbors along a curve. Renormalization group techniques may then be used to extract critical behavior that is independent of this artificial cutoff length scale [4], [8].

2.2 Three-point distance for curves

An alternative approach to defining the argument α of the self-interaction energy is based on triples of points. To begin, consider any three distinct points $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ on a simple curve $\mathcal{C}$. When these points are not collinear they define a triangle with sides of lengths $|\boldsymbol{x}-\boldsymbol{y}|$, $|\boldsymbol{x}-\boldsymbol{z}|$ and $|\boldsymbol{y}-\boldsymbol{z}|$, perimeter $\mathcal{P}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$ and area $\mathcal{A}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$. Each of these quantities vanishes in any limit in which all the points coalesce into one, so they do not individually yield an appropriate length scale for self-interaction. On the other hand, certain combinations provide quantities that remain positive in coalescent limits. One natural combination is

$$r(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) = \frac{|\boldsymbol{x}-\boldsymbol{y}||\boldsymbol{x}-\boldsymbol{z}||\boldsymbol{y}-\boldsymbol{z}|}{4\mathcal{A}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})}$$

which can be identified as the radius of the circumcircle, i.e. the unique circle passing through $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$. By convention, we take this radius to be infinite when the points are collinear.

Various properties of the three-point circumradius function $r(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$ were studied in [16]. For our purposes we merely note that the domain of the function $r(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$ can be extended by continuous limits to all triples of points on $\mathcal{C}$, distinct or not. For example, if $\boldsymbol{x} = \boldsymbol{q}(s)$, $\boldsymbol{y} = \boldsymbol{q}(\sigma)$ and $\boldsymbol{z} = \boldsymbol{q}(\tau)$ are three distinct points on $\mathcal{C}$, then it is straightforward to show that

$$\lim_{\sigma,\tau \to s} r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \rho(\boldsymbol{x}) \tag{3}$$

where $\rho(\boldsymbol{x})$ is the standard local radius of curvature of $\mathcal{C}$ at $\boldsymbol{x}$. (Because we consider only curves that are twice continuously differentiable this coalescent limit exists at each point; see [13] for further details when the underlying curve is not smooth.) From its geometrical interpretation we may also deduce that $r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ and its limits are invariant under translations and rotations of a curve. Moreover, whenever different parts of a curve come close to forming a self-intersection, there are points $\boldsymbol{x}$ and $\boldsymbol{y}$ for which the limits $r(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y})$ and $r(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{x})$ are both equal to half of the distance of closest approach.

A general class of self-interaction energies involving the circumradius function can now be defined. In particular, one may consider the energy

$$E[\mathcal{C}] = \iiint V(r(s, \sigma, \tau)) \, \mathrm{d}\mathcal{C}_s \, \mathrm{d}\mathcal{C}_\sigma \, \mathrm{d}\mathcal{C}_\tau$$

where

$$r(s, \sigma, \tau) = r(\boldsymbol{q}(s), \boldsymbol{q}(\sigma), \boldsymbol{q}(\tau))$$

and V is of the form introduced in Figure 1, or more simply $V(\alpha) = \alpha^{-m}$ with m an appropriately large exponent. In contrast to the pairwise distance function $\beta(s, \sigma)$, the circumradius function $r(s, \sigma, \tau)$ does not suffer nearest-neighbor effects due to curve continuity. In particular, $E[\mathcal{C}]$ is well-defined for any simple curve $\mathcal{C}$.

The condition $V(\alpha) \to \infty$ as $\alpha \to 0$ simultaneously provides control over a curve at both a local and global level. For example, at a point s on the curve, the limit $r(s, s, s)$ defined in (3) is just the local radius of curvature at that location and the potential $V(r(s, s, s))$ is finite as long as this radius is non-zero (or equivalently, as long as the local curvature is finite). In this case, the three-body potential V plays the role of a local curvature energy that encourages curve smoothness. On the other hand, whenever different neighborhoods of a curve come together to form a self-intersection, there are points of closest-approach (s, σ) for which the limits $r(s, \sigma, \sigma)$ and $r(\sigma, s, s)$ tend to zero, leading to infinite values for $V(r(s, \sigma, \sigma))$ and $V(r(\sigma, s, s))$. Thus, the three-body potential is also a global self-interaction energy that acts to discourage self-intersections.

Characteristic length scales for curve self-interactions can be identified depending on the functional form of V. For example, potentials with the general form introduced in Figure 1 provide a natural scale for modelling the steric self-interactions of material filaments with non-zero thickness. A single generic potential (repulsive at short r, attractive for intermediate r) based on a three-body argument $r(s, \sigma, \tau)$ suffices to obtain both a swollen phase at high temperatures and a dense phase at low temperatures, along with a phase transition between them in analogy with the fluid-solid transition for unconstrained particles with a pairwise potential. In the more standard continuum approach [2]–[4] a similar transition is obtained by introducing both attractive singular two-body (in order to encourage collapse) and repulsive singular three-body (in order to account for self-avoidance) δ-function types of potentials.

We remark that the forces derived from a three-body potential have a geometrical interpretation analogous to those for a two-body potential. In the two-body case, the force on each particle is directed along the line that contains the two particles. In the three-body case, the force on each of the three distinct particles is directed along a radial line through the particle and the center of the circle that contains all three of the particles (cf. Figure 3).

2.3 Other distances for curves

The three-point circumradius function leads to a notion of distance between pairs of points that contains geometrical information in addition to the standard distance. For example, if $\boldsymbol{x} = \boldsymbol{q}(s)$, $\boldsymbol{y} = \boldsymbol{q}(\sigma)$ and $\boldsymbol{z} = \boldsymbol{q}(\tau)$ are three distinct points on $\mathcal{C}$, then one can associate a distance to $\boldsymbol{x}$ and $\boldsymbol{y}$ according to

$$\mathsf{pt}(\boldsymbol{x},\boldsymbol{y}) = \lim_{\tau\to\sigma} r(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) = \frac{|\boldsymbol{x}-\boldsymbol{y}|}{2|\sin\theta_{\boldsymbol{x}\boldsymbol{y}'}|}$$

where $\theta_{\boldsymbol{x}\boldsymbol{y}'}$ is the angle between the vector $\boldsymbol{x}-\boldsymbol{y}\neq\mathbf{0}$ and the tangent vector to $\mathcal{C}$ at $\boldsymbol{y}$. The function $\mathsf{pt}(\boldsymbol{x},\boldsymbol{y})$ can be identified as the radius of the unique circle through $\boldsymbol{x}$ that is tangent to $\mathcal{C}$ at $\boldsymbol{y}$. Notice that the function $\mathsf{pt}(\boldsymbol{x},\boldsymbol{y})$ will typically not be symmetric since the circle through $\boldsymbol{y}$ and tangent at $\boldsymbol{x}$ need not have the same radius as the circle through $\boldsymbol{x}$ and tangent at $\boldsymbol{y}$. We will refer to $\mathsf{pt}(\boldsymbol{x},\boldsymbol{y})$ as the point-tangent function. As a matter of convention, we consider it to be a particular three-point function, because it corresponds to a limit of the three-point circumradius function.

The point-tangent function $\mathsf{pt}(\boldsymbol{x},\boldsymbol{y})$ shares two important properties with the circumradius function $r(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$. First, the coalescent limit $\sigma\to s$ is just the local radius of curvature, namely

$$\lim_{\sigma\to s}\mathsf{pt}(\boldsymbol{x},\boldsymbol{y}) = \rho(\boldsymbol{x}).$$

Second, whenever different neighborhoods of a curve come together to form a self-intersection, there are points $\boldsymbol{x}$ and $\boldsymbol{y}$ for which $\mathsf{pt}(\boldsymbol{x},\boldsymbol{y})$ and $\mathsf{pt}(\boldsymbol{y},\boldsymbol{x})$ are both equal to half of the distance of closest approach. As a consequence of these two properties, a general class of self-interaction energies involving the point-tangent function as argument can be defined as

$$E[\mathcal{C}] = \iint V(\mathsf{pt}(s,\sigma))\, \mathrm{d}\mathcal{C}_s\, \mathrm{d}\mathcal{C}_\sigma$$

where $\mathsf{pt}(s,\sigma) = \mathsf{pt}(\boldsymbol{q}(s), \boldsymbol{q}(\sigma))$ and V is a general potential function as before. The energy $E[\mathcal{C}]$ is well-defined for any simple curve $\mathcal{C}$.

Other distances for curves based on four points, or limits thereof, can also be considered. For example, to any four distinct points $\boldsymbol{w}$, $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ on a simple curve $\mathcal{C}$ one can associate the four-point distance $R(\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ defined as the radius of the smallest sphere that contains all four points. Usually there will be such a unique sphere, but if all four points happen to be co-circular, there are many spheres passing through them with the smallest having the radius of the circle through the four points. Any limit of the spherical radius function as the four points coalesce into one is always positive because it is greater than or equal to the radius of curvature of the curve at the limit point. If the coalescence point has non-zero torsion, one obtains the radius of the *osculating sphere* in the limit, see for example [20, p. 25]. Moreover, just as for the three-point functions r and pt, whenever different parts of a curve come together to form a self-intersection, there are points for which R is equal to half of the distance of closest approach.

(But now the possibility of end-point effects must be explicitly excluded, as for example if the curve is closed.) Similar results hold for the symmetric tangent-tangent distance function $\mathtt{tt}(\boldsymbol{x}, \boldsymbol{y})$ defined to be the radius of the smallest sphere containing $\boldsymbol{x}$ and $\boldsymbol{y}$, that is also tangent to $\mathcal{C}$ at both these points.

3 Self-interactions of surfaces

Self-avoiding surfaces have been a subject of much interest and study in diverse disciplines ranging from mathematics to biology (see for example [5], [6], [21]–[24]). Our objective is therefore to extend to the case of surfaces, our construction of multi-point distances that can lead to non-singular repulsive energies that preclude self-intersection. Our first conclusion is that for surfaces the simplest two-point (Euclidean), three-point (circular), and four-point (spherical) distances all share a similar problem: they cannot distinguish between proximity due to continuity and proximity due to geometry. We will demonstrate, however, that the notion of point-tangent distance for a curve has a natural counterpart for surfaces that can make this distinction, and which provides a suitable argument α for a self-interaction potential. Throughout our developments a surface $\mathcal{S}$ will mean a (twice continuously differentiable) function $\boldsymbol{p}(u) \in \mathbb{R}^3$ of a variable $u \in A \subset \mathbb{R}^2$, and a surface will be called simple if it has no self-intersections; that is, $\boldsymbol{p}(u) = \boldsymbol{p}(v)$ only when $u = v$.

3.1 n-point distances for surfaces ($n = 2, 3, 4$)

Just as for the case of curves, an intuitive approach to the construction of a self-interaction energy for a simple surface $\mathcal{S}$ is to consider a repulsive potential dependent upon the pairwise or two-point distance function $\beta(u, v) = |\boldsymbol{p}(u) - \boldsymbol{p}(v)|$ as argument. As before, while $\beta(u, v)$ has the desirable feature that it tends to zero as different neighborhoods of $\mathcal{S}$ approach a self-intersection, it also tends to zero as $u \to v$ by continuity, and thus leads to singular interaction potentials.

In contrast to the case of curves, one may show that for surfaces both the three-point function $r(u, v, w)$ and the four-point function $R(t, u, v, w)$ introduced earlier, also lead to singular self-interaction potentials. In particular, these functions tend to zero by continuity in the coalescent limit. This conclusion may be established as follows. Consider any fixed point $\boldsymbol{x}$ of $\mathcal{S}$ and let $\Sigma_\varepsilon(\boldsymbol{x})$ be the sphere of radius ε centered at $\boldsymbol{x}$. For each sufficiently small $\varepsilon > 0$ the intersection $\Sigma_\varepsilon(\boldsymbol{x}) \cap \mathcal{S}$ of the sphere and surface is a curve. The radius r of the circle through any three points on this curve satisfies $r \leq \varepsilon$

(because the circle lies on the sphere of radius ε), and similarly the radius R of the smallest sphere through any four points on this curve satisfies $R \leq \varepsilon$. Thus, by considering the limit $\varepsilon \to 0$, we can find sequences of three distinct points in the surface for which $r \to 0$, and sequences of four distinct points for which $R \to 0$.

3.2 Tangent-point distance for surfaces

For surfaces, the argument of a non-singular self-interaction potential can be obtained by passing directly to tangent-point distances. In particular, to any two distinct points $\boldsymbol{x}$ and $\boldsymbol{y}$ of a simple surface $\mathcal{S}$ we may associate the distance

$$\mathsf{tp}(\boldsymbol{x}, \boldsymbol{y}) = \frac{|\boldsymbol{y} - \boldsymbol{x}|^2}{2|\boldsymbol{n}_{\boldsymbol{x}} \cdot (\boldsymbol{y} - \boldsymbol{x})|}$$

where the vector $\boldsymbol{n}_{\boldsymbol{x}}$ is either of the two unit normals to the surface $\mathcal{S}$ at the point $\boldsymbol{x}$. The function $\mathsf{tp}(\boldsymbol{x}, \boldsymbol{y})$ can be identified as the radius of the unique sphere through $\boldsymbol{y}$ and tangent to $\mathcal{S}$ at $\boldsymbol{x}$. When $\boldsymbol{y}$ happens to be in the tangent plane to $\mathcal{S}$ at $\boldsymbol{x}$, the sphere itself degenerates into a plane and $\mathsf{tp}(\boldsymbol{x}, \boldsymbol{y})$ becomes infinite. Notice that $\mathsf{tp}(\boldsymbol{x}, \boldsymbol{y})$ need not be symmetric since the sphere through $\boldsymbol{x}$ and tangent at $\boldsymbol{y}$ need not have the same radius. We refer to $\mathsf{tp}(\boldsymbol{x}, \boldsymbol{y})$ as the tangent-point function for surfaces. Moreover, we consider it to be a particular four-point function since the tangent plane to $\mathcal{S}$ at $\boldsymbol{x}$ may be constructed through a coalescent limit of three points.

The tangent-point function $\mathsf{tp}(\boldsymbol{x}, \boldsymbol{y})$ enjoys various properties analogous to those of its counterpart defined for curves. For example, consider any curve $\boldsymbol{q}(s)$ in $\mathcal{S}$ such that $\boldsymbol{q}(0) = \boldsymbol{x}$ and let $\boldsymbol{t_x} = \boldsymbol{q}'(0)$. Then

$$\lim_{s \to 0} \mathsf{tp}(\boldsymbol{x}, \boldsymbol{q}(s)) = \frac{1}{|\boldsymbol{n_x} \cdot \boldsymbol{q}''(0)|} = \rho(\boldsymbol{x}, \boldsymbol{t_x})$$

where $\rho(\boldsymbol{x}, \boldsymbol{t_x})$ is the absolute value of the local normal radius of curvature to $\mathcal{S}$ at $\boldsymbol{x}$ in the direction $\boldsymbol{t_x}$. Thus, in coalescent limits $\boldsymbol{y} \to \boldsymbol{x}$, $\mathsf{tp}(\boldsymbol{x}, \boldsymbol{y})$ may assume limiting values between the maximum and minimum of $\rho(\boldsymbol{x}, \boldsymbol{t_x})$ over directions $\boldsymbol{t_x}$ with $\boldsymbol{x}$ fixed. When different parts of $\mathcal{S}$ come together to form a self-intersection, there are points $\boldsymbol{x}$ and $\boldsymbol{y}$ for which $\mathsf{tp}(\boldsymbol{x}, \boldsymbol{y})$ and $\mathsf{tp}(\boldsymbol{y}, \boldsymbol{x})$ are both equal to half of the distance of closest approach of the surface to itself.

While $\mathsf{tp}(\boldsymbol{x}, \boldsymbol{y})$ need not be continuous when $\boldsymbol{y} = \boldsymbol{x}$, one may unambiguously consider self-interaction energies for surfaces of the form

$$E[\mathcal{S}] = \iint V(\mathsf{tp}(u, v))\, \mathrm{d}\mathcal{S}_u\, \mathrm{d}\mathcal{S}_v$$

where $\mathsf{tp}(u, v) = \mathsf{tp}(\boldsymbol{p}(u), \boldsymbol{p}(v))$ and V is a general potential function. As in the case of a spaghetto, the choice of a potential of the form introduced in Figure 1 could be used to model a surface of finite thickness. Similar ideas could be used for the description of triangulated or discretized random surfaces. In particular, one may discretize the energy introduced above to obtain a self-interaction energy for the triangulated surface. Here three vertices of each triangle could be used to define a tangent plane, which could then be used to evaluate the tangent-point function tp.

We remark that Möbius-invariant energies for self-avoiding surfaces involving pairs of spheres, each tangent to the surface, are discussed in [25]. However, energies based on the radius of a single such sphere are not discussed.

4 Summary

Curves and surfaces with an associated finite thickness can be used as continuum models of strands (such as a spaghetto) and sheets (such as this page). We find that non-singular self-interaction energies for such self-avoiding continuous systems cannot depend on the usual pairwise distance. However, by using certain many-body potentials, we are able to construct self-interaction energies which are non-singular and which include a mesoscopic length scale for the physical thickness. Each many-body potential is a function of the radius of a suitably chosen circle or sphere. Our energies provide simple continuum models of strand and sheet systems that can be used to study the equilibrium and non-equilibrium statistical mechanics of distinct phases at different temperatures, along with phase transitions.

We conclude by highlighting an application of our work in the context of the protein problem. Small globular proteins are linear chains of amino acids which, under physiological conditions, fold rapidly and reproducibly in a cooperative manner into their native state conformations [26]. These conformations are somewhat compact structures corresponding to the minima of an effective energy. Furthermore, for proteins, form determines function and, yet, the total number of distinct native state folds is believed to be only of the order of a few thousand [27] and are made up of helices, hairpins or sheets. An important issue in the protein field is to elucidate the bare essentials that determine the novel phase adopted by biopolymers such as proteins. Unfortunately, polymer science, which is a mature and technologically important field, does not provide an answer to this question. The standard, simple model of a polymer chain is one of tethered hard spheres – in the continuum limit, self-avoidance in this model is captured by a singular pairwise potential as in the second term on the right hand side of equation

(1). Such a model with an additional effective attraction arising from the hydrophobicity or mutual aversion of certain amino acid residues to the solvent (water), fails to account for the protein native structure phase on several counts. First, a generic compact polymer phase has many conformations which neither provide for specificity nor are kinetically readily accessible. Proteins have a limited number of folds to choose from for their native state structure and the energy landscape is vastly simpler. Second, the structures in the polymer phase are not especially sensitive to perturbations and are thus not flexible and versatile as protein native state structures are in order to accommodate the dizzying array of functions that proteins perform.

Recent work [28], built on the ideas presented here, has pointed out a crucial missing ingredient in the simple model of a polymer chain of tethered hard spheres. Strikingly, the simple physical idea of a chain viewed, instead, as a tube leads to several dramatic consequences. We are conditioned to think of objects as spheres and the effective interactions between them as being pair-wise in nature. This bias in our thinking arises from our everyday experience with unconstrained objects. However, when one deals with objects tethered together in a chain, some of the old notions need to be discarded. For example, a pairwise interaction only provides information regarding the mutual distance between two interacting particles but does not have any contextual information regarding how far apart the two particles are along the chain. Our work here shows that in order to capture the constraints imposed by a tube geometry associated with a discrete chain, the conventional notion of pairwise interactions between particles has to be augmented by appropriately chosen three-body interactions to capture the steric constraints imposed by the tube.

Such a tube-like description of a chain leads to many of the standard results of polymer physics when the tube thickness is small compared to other length scales in the problem [28]. However, when the tube size becomes comparable to the range of the effective attractive interactions resulting from the hydrophobicity, the novel phase populated by biopolymers results [28]. This self-tuning of length scales occurs naturally for proteins because the steric effects [29, 30] associated with the backbone and side chain atoms of the amino acids, on the one hand, lead to a tube like description and control the tube thickness and the same side chains, on the other hand, have an effective attractive interaction which is at an Angstrom scale due to the screening effects of the water.

It has been demonstrated [28] that this novel phase has several characteristics of the phase populated by protein native structures including the ability to expel water from the hydrophobic core, a vast simplification in the energy landscape with relatively few putative native state structures, a prediction that helices, zig-zag hairpins and sheets of the correct geometry are the building blocks of protein structures, a simple explanation of the cooperative nature of the folding transition of globular proteins and an explanation of why protein structures are flexible and versatile. Thus the idea of a tube and the use of appropriate many-body potentials introduced here not only lead to a better description of a polymer but allows one to bridge the gap between polymer science and protein science and provides, for the first time, a framework based on geometry for an understanding of the common character of all globular proteins [28]. We look forward to similar applications of our new ideas on sheets of non-zero thickness.

References

[1] Yamakawa, H., Modern theory of polymer solutions, Harper and Row, New York (1971).

[2] de Gennes, P.G., Scaling concepts in polymer physics, Cornell University Press, Ithaca (1979).

[3] Doi, M. and Edwards, S.F., The theory of polymer dynamics, Clarendon Press, New York (1993).

[4] des Cloiseaux, G. and Jannink, J.F., Polymers in solution: their modeling and structure, Clarendon Press, Oxford (1990).

[5] Nelson, D.R., Piran, T. and Weinberg, S., Editors, Statistical mechanics of membranes and surfaces, Volume 5, Jerusalem Winter School for Theoretical Physics, World Scientific, Singapore (1989).

[6] Wiese, K.J., Polymerized membranes: a review, in Phase transitions and critical phenomena, Volume 19, Domb, C. and Lebowitz, J.L., Editors, Academic Press, New York (2000).

[7] Wilson, K.G. and Fisher, M.E., Critical exponents in 3.99 dimensions, Phys. Rev. Lett. 28 (1972) 240–243.

[8] Zinn-Justin, J., Quantum field theory and critical phenomena, Clarendon Press, New York (1993).

[9] Kadanoff, L.P., Critical behavior, universality and scaling in critical phenomena, Proc. of the Int. School of Phys. "E. Fermi" (Varenna) Course LI, Green, M.S., Editor, Academic Press, New York (1971).

[10] Stanley, H.E., Scaling, universality and renormalization: three pillars of modern critical phenomena, Rev. Mod. Phys. 71 (1999) S358–S366.

[11] Cantarella, J., Kusner, R. and Sullivan, J.M., On the minimum rope-length of knots and links, Inventiones Math, to appear.

[12] Gonzalez, O. and de la Llave, R., Existence of ideal knots, Journal of Knot Theory and its Ramifications, to appear.

[13] Gonzalez, O., Maddocks, J.H., Schuricht, F. and von der Mosel, H., Global curvature and self-contact of nonlinearly elastic curves and rods, Calculus of Variations 14 (2002) 29–68 (DOI 10.1007/s005260100089, April 2001).

[14] Schuricht, F. and von der Mosel, H., Global curvature for rectifiable loops, Mathematische Zeitschrift, to appear.

[15] Katritch, V., Bednar, J., Michoud, D., Scharein, R.G., Dubochet, J. and Stasiak, A., Geometry and physics of knots, Nature 384 (1996) 142–145.

[16] Gonzalez, O. and Maddocks, J.H., Global curvature, thickness and the ideal shapes of knots, Proc. Natl. Acad. Sci. USA 96 (1999) 4769–4773.

[17] Maritan, A., Micheletti, C., Trovato, A. and Banavar, J.R., Optimal shapes of compact strings, Nature 406 (2000) 287–290.

[18] Gonzalez, O., Maddocks, J.H. and Smutny, J., Curves, circles, and spheres, to appear in Contemporary Mathematics, American Mathematical Society, Providence, R.I. (2002).

[19] Stasiak, A., Katritch, V. and Kauffman, L.H., Editors, Ideal Knots, World Scientific Publishing, Singapore (1998).

[20] Struik, D.J., Lectures on Classical Differential Geometry, Second Edition, Dover, New York (1988).

[21] Wiese, K.J. and David, F., Self-avoiding tethered membranes at the tricritical point, Nucl. Phys. B 450 (1995) 495–557.

[22] Baumgärtner, A., Does a polymerized membrane crumple?, J. Phys. I France 1 (1991) 1549–1556.

[23] Baumgärtner, A. and Renz, W., Crumpled self-avoiding tethered surfaces, Europhys. Lett. 17 (1992) 381–386.

[24] Kroll, D.M. and Gompper, G., Floppy tethered networks, J. Phys. I France 3 (1993) 1131.

[25] Kusner, R.B. and Sullivan, J.M., Möbius invariant knot energies, in Ideal Knots, Chapter 17, Stasiak, A., Katritch, V. and Kauffman, L.H., Editors, World Scientific Publishing, Singapore (1998).

[26] Anfinsen, C., Principles that govern the folding of protein chains, Science 181 (1973) 223–230.

[27] Chothia, C., One thousand families for the molecular biologist, Nature 357 (1992) 543–544.

[28] Banavar, J.R., Maritan, A., Micheletti, C. and Trovato, A., Geometry and physics of proteins, Proteins 47 (2002) 315–322 (DOI 10.1002/prot.10091). Banavar, J.R., Flammini, A., Marenduzzo, D., Maritan, A. and Trovato, A., Geometry of compact tubes and protein structures, ComPlexUs (in press).

[29] Ramachandran, G.N. and Sasisekharan, V., Conformations of polypeptides and proteins, Adv. Protein Chem. 23 (1968) 283–438.

[30] Srinivasan, R. and Rose, G.D., LINUS - A hierarchical procedure to predict the fold of a protein, Proteins 22 (1995) 81–99.

Acknowledgments. We are indebted to Philip Anderson and Leo Kadanoff for valuable comments on the manuscript. The following generous support is gratefully acknowledged: JRB the Penn State MRSEC under NSF grant DMR-0080019 and NASA, OG the US National Science Foundation under grant DMS-0102476, JHM the Swiss National Science Foundation, and AM the MURST Cofin'99.

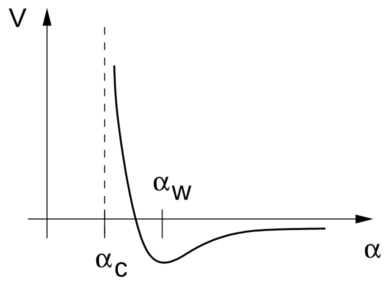

Figure 1: Example potential function with hard-core and potential well parameters α_c and α_w.

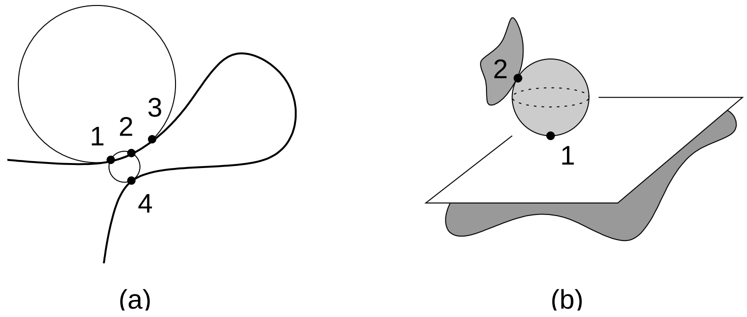

Figure 2: Interpretation of multi-point distances for a curve and a surface. (**a**), Three-point distance for curve. Given any three distinct points, r is the radius of the unique circle that contains the points. When the points are from the same neighborhood on a curve, such as points 1, 2 and 3, r is close to the local radius of curvature. When points, such as 1, 2 and 4, are taken from two different neighborhoods of the curve that are close to intersection, r approximates (half of) the distance of closest approach of the curve to itself. (**b**), Tangent-point distance for a surface. Given two distinct points 1 and 2 on a surface, `tp` is the radius of the unique sphere that contains both points and is tangent to the surface at point 1. When the points are neighbors on the surface, `tp` approximates the absolute value of the local normal radius of curvature in the direction defined by the two points (not illustrated). When points are taken from different neighborhoods that are close to intersection, `tp` approximates (half of) the distance of closest approach.

Figure 3: Graphical interpretation of forces derived from potentials depending on two- and three-point distance functions. (**a**), two particles interacting via a potential depending on two-point or pairwise distance. Holding particle 2 fixed and moving particle 1 on the circumference of the circle leaves the energy unchanged. This implies that the force on each particle is along the joining line, and that the resultant of the forces is zero due to the translational invariance of the energy. (**b**), three particles interacting via a potential depending on three-point distance. Here any combination of the particles may be moved along the circumference of the circle without changing their interaction energy. This implies that for any distinct triplet the force on each particle is along a radial line from the center of the circle to the particle, and that the resultant of the forces is again zero due to the translational invariance of the energy.

Caption for Figure 1

Figure 1: Example potential function with hard-core and potential well parameters α_c and α_w.

Caption for Figure 2

Figure 2: Interpretation of multi-point distances for a curve and a surface. (**a**), Three-point distance for curve. Given any three distinct points, r is the radius of the unique circle that contains the points. When the points are from the same neighborhood on a curve, such as points 1, 2 and 3, r is close to the local radius of curvature. When points, such as 1, 2 and 4, are taken from two different neighborhoods of the curve that are close to intersection, r approximates (half of) the distance of closest approach of the curve to itself. (**b**), Tangent-point distance for a surface. Given two distinct points 1 and 2 on a surface, tp is the radius of the unique sphere that contains both points and is tangent to the surface at point 1. When the points are neighbors on the surface, tp approximates the absolute value of the local normal radius of curvature in the direction defined by the two points (not illustrated). When points are taken from different neighborhoods that are close to intersection, tp approximates (half of) the distance of closest approach.

Caption for Figure 3

Figure 3: Graphical interpretation of forces derived from potentials depending on two- and three-point distance functions. (**a**), two particles interacting via a potential depending on two-point or pairwise distance. Holding particle 2 fixed and moving particle 1 on the circumference of the circle leaves the energy unchanged. This implies that the force on each particle is along the joining line, and that the resultant of the forces is zero due to the translational invariance of the energy. (**b**), three particles interacting via a potential depending on three-point distance. Here any combination of the particles may be moved along the circumference of the circle without changing their interaction energy. This implies that for any distinct triplet the force on each particle is along a radial line from the center of the circle to the particle, and that the resultant of the forces is again zero due to the translational invariance of the energy.