

The geometry of soft materials: a primer

Randall D. Kamien*

Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104

(Published 11 September 2002)

This article presents an overview of the differential geometry of curves and surfaces using examples from soft matter as illustrations. The presentation requires a background only in vector calculus and is otherwise self-contained.

CONTENTS

I. Introduction	953
II. Local Theory of Curves	953
A. Conformations of polymers: motivating a geometrical description of curves	953
B. Frenet-Serret apparatus: DNA and other chiral polymers	954
III. Global Theory of Curves	956
A. Fenchel's theorem: energetic bounds on closed curves and knots	956
B. Mermin-Ho relation: basis vectors to the rescue	957
C. Link, twist, and writhe: dynamics of twist-storing polymers	958
IV. Local Theory of Surfaces	961
A. Area element: minimal surfaces	961
B. Mean and Gaussian curvature: energetics of membranes	964
V. Global Theory of Surfaces	965
A. Gauss-Bonnet theorem: foams on curved surfaces	965
B. Euler characteristic and the genus: defects on surfaces	967
VI. Three Dimensions and Beyond	969
Outlook	970
Acknowledgments	970
References	970

I. INTRODUCTION

Though geometry is a common part of our early schooling, a rigorous and thorough education in physics usually tries to expunge it from our thought. Because of their predictive powers, there is good reason to emphasize numbers and formulas. Analytic geometry is the usual emphasis, while classical geometry is relegated to popular expositions. Differential geometry is a bridge between shapes and analytic expressions and is often the appropriate language for modern physics. Nonetheless, when necessary, geometry is often slipped in as a bitter or, at least, tasteless pill—just enough is presented to get on with the analysis. In these lectures I have made an attempt to introduce the basics of differential geometry in the style of a mathematics text: the ideas are grouped by mathematical subject as opposed to physical subject. Nonetheless, I follow each newly developed topic with

an example from soft matter, not only to illustrate the usefulness of the mathematical structure but also to aid the reader with a concrete example.

This is by no means a textbook and many details of the physics are left for the diligent reader to find in the references. For technical details of many of the topics discussed here, the reader is referred to *Elements of Differential Geometry* by Millman and Parker (Millman and Parker, 1977), though any standard reference on classical differential geometry should suffice. A word on notation: I have tried throughout to explicitly display the functional dependence of all the fields and functions in formulas. However, sometimes this would make the notation awkward and the dependencies are dropped. In each case the context should make clear any lack of precision.

Finally, I have tried to reduce as much as possible the use of the powerful formalism of differential geometry. While this formalism is useful for performing complex calculations unambiguously, great expertise is often required to extract the geometrical and physical meaning of these calculations. An excellent complement to these notes are the lecture notes by François David (David, 1989) which present the mathematical elegance and logical compactness of this subject.

II. LOCAL THEORY OF CURVES

A. Conformations of polymers: motivating a geometrical description of curves

Random walks abound in physics. They are the basis for understanding diffusion, heat flow, and polymers. However, polymers are the most interesting of the three: polymers, unlike diffusing particles, leave a “tail” behind them which they must avoid. They are described by self-avoiding random walks. As an introduction to the power of geometrical modeling, in this section we will consider the behavior of stiff polymers at the shortest length scale amenable to a continuum analysis. At these scales, polymers are not random walks at all, but should be thought of as stiff rods. We take this as our starting point. When we describe a polymer as stiff, we are ascribing to it an energy cost for being bent. To model this, we consider a curve, $\mathbf{R}(s)$, parametrized by its arc length s and construct the tangent vector to our curve $\mathbf{T}(s) = d\mathbf{R}(s)/ds$ at a point s on the curve. In the next section we will show that the tangent vector is of unit length. If the

*Electronic address: kamien@physics.upenn.edu

tangent vector is constant along the curve then it is a straight line and does not bend. Thus the energy should depend on derivatives of the unit tangent vector $\mathbf{T}(s)$. Indeed we call the magnitude $d\mathbf{T}(s)/ds$ the curvature of the curve. In the next section we will discuss the geometry of curves in greater detail. In the meantime, we write the energy as

$$E_{\text{curv}} = \frac{1}{2} A \int_0^L \left[\frac{d\mathbf{T}(s)}{ds} \right]^2 ds, \quad (1)$$

where s is a parameter that measures the arc length of the curve, and A is a measure of the stiffness. To study the statistical mechanics of the curve we write the partition function for \mathbf{T} :

$$Z(\mathbf{T}_L) = \int_{\mathbf{T}(0)=\hat{z}}^{\mathbf{T}(L)=\mathbf{T}_L} [d\mathbf{T}] e^{-E_{\text{curv}}[\mathbf{T}]/k_B T}. \quad (2)$$

This is the partition function for the curve which starts with its tangent vector along the z direction and ends with its tangent vector equal to \mathbf{T}_L . Note that if we can calculate $\langle \mathbf{T}(s)\mathbf{T}(s') \rangle$, then we can integrate with respect to s and s' to obtain

$$\langle [\mathbf{R}(L) - \mathbf{R}(0)]^2 \rangle = \int_0^L ds \int_0^L ds' \left\langle \frac{d\mathbf{R}(s)}{ds} \cdot \frac{d\mathbf{R}(s')}{ds'} \right\rangle. \quad (3)$$

Calculating the correlation function of the tangent vectors proves to be straightforward. Though there are many ways to proceed (Doi and Edwards, 1986), we will choose here an analogy with quantum mechanics. In Eq. (2) let $s = it$. Then the partition function becomes

$$Z(\mathbf{T}_L, L) = \int_{\mathbf{T}(0)=\hat{z}}^{\mathbf{T}(L)=\mathbf{T}_L} [d\mathbf{T}] \exp \left\{ \frac{i}{k_B T} \int_0^{-iL} dt \frac{A}{2} \left(\frac{d\mathbf{T}}{dt} \right)^2 \right\}; \quad (4)$$

we recognize this as the path-integral solution to Schrödinger's equation for a single particle where $k_B T$ replaces \hbar , A replaces the particle's mass, and \mathbf{T} is the position of the particle. Since $\mathbf{T}(t)$ lives on the unit sphere, this is just quantum mechanics on a sphere. The Schrödinger equation is

$$-\frac{k_B T}{i} \frac{\partial Z}{\partial L} = -\frac{(k_B T)^2}{2A} \hat{\mathbf{L}}^2 Z, \quad (5)$$

where $\hat{\mathbf{L}}$ is the angular momentum operator. Changing back to our original coordinate s and defining $L_p \equiv A/k_B T$ we have

$$\frac{\partial Z(\mathbf{T}, s)}{\partial s} = \frac{1}{2L_p} \hat{\mathbf{L}}^2 Z(\mathbf{T}, s). \quad (6)$$

We will soon discover that L_p has an important interpretation.

Since the polymer is the same all along its length, we have $\langle \mathbf{T}(s) \cdot \mathbf{T}(s') \rangle = \langle \mathbf{T}(s-s') \cdot \mathbf{T}(0) \rangle$. Using polar coordinates and enforcing the limits of integration in Eq. (4) so that $\mathbf{T}(0) = \hat{z}$, we discover that we are interested in $\langle \cos \theta(S) \rangle$ at $S = s - s'$:

$$\langle \cos \theta(S) \rangle = \frac{\int_{-1}^1 d(\cos \theta) \cos \theta Z(\cos \theta, S)}{\int_{-1}^1 d(\cos \theta) Z(\cos \theta, S)}, \quad (7)$$

where we have been sure to normalize Z to get a probability. The expectation value we seek satisfies a differential equation since $Z(\cos \theta, S)$ satisfies Eq. (6). Moreover, since $\hat{\mathbf{L}}^2$ is a Hermitian operator, we have

$$\frac{d\langle \cos \theta(S) \rangle}{dS} = \frac{1}{2L_p} \langle \hat{\mathbf{L}}^2 \cos \theta(S) \rangle = -\frac{1}{L_p} \langle \cos \theta(S) \rangle, \quad (8)$$

where we have used the fact that $\hat{\mathbf{L}}^2 \cos \theta = -2 \cos \theta$ and $\hat{\mathbf{L}}^2(1) = 0$. It follows that

$$\langle \cos[\theta(s) - \theta(s')] \rangle = e^{-|s-s'|/L_p}. \quad (9)$$

Integrating this expression as in Eq. (3), we have

$$\langle [\mathbf{R}(L) - \mathbf{R}(0)]^2 \rangle = 2L_p(L - L_p + L_p e^{-L/L_p}). \quad (10)$$

From Eqs. (9) and (10) we glean the meaning of L_p . The first equation shows that the tangent vectors along the curve are uncorrelated after a distance L_p . For this reason L_p is called the persistence length (de Gennes, 1970). The second equation shows us that for L much longer than L_p , a stiff rod behaves as a random walk: i.e., the average square distance that is traveled is proportional to the number of steps or length of the walk, $R^2 \propto LL_p$. For L much shorter than L_p , we may expand Eq. (10) to see that $R^2 \propto L^2$.

The physics of stiff rods can be used to study other phenomena. For instance, vortices in fluids, superfluids, and superconductors have a bending stiffness arising from self-interactions. In the past decade the physics of stiff rods has been adapted to study the mechanical properties of DNA (Marko and Siggia, 1994, 1995), and has been augmented to include twisting degrees of freedom (Kamien *et al.*, 1997; Marko, 1997), the topological constraints imposed by self-avoidance (Vologodskii *et al.*, 1992) and by closed loops (Moroz and Kamien, 1997; Moroz and Nelson, 1997, 1998), and the effects of sequence disorder (Bensimon *et al.*, 1998; Nelson, 1998).

B. Frenet-Serret apparatus: DNA and other chiral polymers

In the previous section we have seen that a simple geometrical description of polymers leads to a precise description of their conformational behavior in a variety of regimes. It is now time to discuss more carefully the geometry that went into making the expression for energy in Eq. (1). A curve in three dimensions is a vector-valued function $\mathbf{R}(s) = [x(s), y(s), z(s)]$ that depends on s , a parameter that runs along the curve. Though we may choose to label points along the curve as we wish, it is usually most convenient to let s be the arc length along the curve. We will see how this simplifies our equations shortly. If the curve is L long, then s runs

from 0 to L . The first thing to do is to construct the unit tangent vector to the curve: as we have already asserted, the magnitude of the rate of change of the unit tangent vector is the curvature κ of the curve. We have

$$\mathbf{T}(s) = \frac{\mathbf{R}'(s)}{\|\mathbf{R}'(s)\|}, \quad (11)$$

where $\mathbf{X}'(s)$ denotes the derivative of $X(s)$ with respect to s , and $\|\mathbf{X}\|$ is the length of the vector \mathbf{X} . We have divided by the length of $\mathbf{R}'(s)$ to make $\mathbf{T}(s)$ a unit vector. However, if s measures the arc length along the curve then $\|\mathbf{R}'(s)\| = 1$! To see this, recall that the length of a curve $\mathbf{R}(t)$ from $t=0$ to $t=t_f$ can be found by adding (integrating) the length of pieces of the curve together, each $\sqrt{d\mathbf{R}(t) \cdot d\mathbf{R}(t)}$ long:

$$\begin{aligned} L(t_f) &= \int_0^{t_f} \sqrt{d\mathbf{R}(t) \cdot d\mathbf{R}(t)} \\ &= \int_0^{t_f} dt \sqrt{\frac{d\mathbf{R}(t)}{dt} \cdot \frac{d\mathbf{R}(t)}{dt}}. \end{aligned} \quad (12)$$

If we choose $t=s$ to be the arc length then we have

$$L = \int_0^L ds \|\mathbf{R}'(s)\|, \quad (13)$$

where the upper limit is the same as the length of the curve. Differentiating both sides of Eq. (13), we see that $\|\mathbf{R}'(s)\| = 1$.

Having constructed the unit tangent vector, we may now take its derivatives to obtain the curvature. Since the derivative of a vector is another vector, we write

$$\mathbf{T}'(s) = \kappa(s)\mathbf{N}(s), \quad (14)$$

where $\kappa(s)$ is the curvature at s and $\mathbf{N}(s)$ is a new vector, *the unit normal vector*, which is also unit length, so that $\|\mathbf{T}'(s)\| = |\kappa(s)|$. It is convention to choose $\kappa(s)$ to be always positive—the sign can always be absorbed in the direction of $\mathbf{N}(s)$. Note that $\mathbf{T}(s) \cdot \mathbf{N}(s) = 0$ since the derivative of *any* unit vector is perpendicular to itself; if $\|\mathbf{X}(s)\| = 1$ then

$$\frac{d}{ds}[\mathbf{X}(s) \cdot \mathbf{X}(s)] = \frac{d}{ds}[1],$$

$$2\mathbf{X}(s) \cdot \mathbf{X}'(s) = 0. \quad (15)$$

Of course, as we move along s , the normal vector changes direction as well. Changes in the direction of the normal vector can come from two contributions: the normal can change by rotating towards or away from the tangent vector (of course the pair rotate together to stay perpendicular). It can also change by rotating *around* the tangent vector. The former case corresponds to the curve staying in the same, flat plane, while the second corresponds to rotations of the plane in which the curve lies at s . This plane is known as the *osculating plane*, from the Greek word for kissing. Moreover, since $\mathbf{N}'(s)$ is perpendicular to $\mathbf{N}(s)$, we must introduce a third unit vector to account for changes in the osculating plane. We choose $\mathbf{B}(s) = \mathbf{T}(s) \times \mathbf{N}(s)$ where \times denotes the

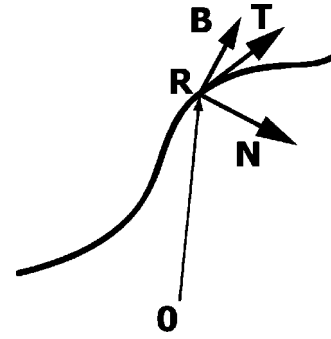


FIG. 1. A curve in space. At \mathbf{R} there is a tangent vector \mathbf{T} , a normal vector \mathbf{N} , and the binormal vector \mathbf{B} .

cross product, as shown in Fig. 1. This vector, *the binormal vector*, is a unit vector perpendicular to both $\mathbf{T}(s)$ and $\mathbf{N}(s)$. We then have

$$\mathbf{N}'(s) = \alpha(s)\mathbf{T}(s) + \tau(s)\mathbf{B}(s), \quad (16)$$

where $\alpha(s)$ is some function of s and $\tau(s)$ is called the *torsion* of the curve. It is a measure of the rate of change of the osculating plane. Why do we not give $\alpha(s)$ a name? Because we note that by differentiating the relation $\mathbf{T}(s) \cdot \mathbf{N}(s) = 0$ we get

$$0 = \mathbf{T}'(s) \cdot \mathbf{N}(s) + \mathbf{T}(s) \cdot \mathbf{N}'(s) = \kappa(s) + \alpha(s), \quad (17)$$

so $\alpha(s) = -\kappa(s)$ and Eq. (16) becomes

$$\mathbf{N}'(s) = -\kappa(s)\mathbf{T}(s) + \tau(s)\mathbf{B}(s). \quad (18)$$

Finally, we may calculate $\mathbf{B}'(s)$ to complete our analysis of the curve. We have

$$\begin{aligned} \mathbf{B}'(s) &= \mathbf{T}'(s) \times \mathbf{N}(s) + \mathbf{T}(s) \times \mathbf{N}'(s) \\ &= \kappa(s)\mathbf{N}(s) \times \mathbf{N}(s) - \kappa(s)\mathbf{T}(s) \times \mathbf{T}(s) \\ &\quad + \tau(s)\mathbf{T}(s) \times \mathbf{B}(s), \\ &= -\tau(s)\mathbf{N}(s) \end{aligned} \quad (19)$$

where the last line follows from the rule $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b})$. Putting Eqs. (14), (18), and (19) together we have the Frenet-Serret equations for a curve in three-dimensions:

$$\frac{d}{ds} \begin{bmatrix} \mathbf{T}(s) \\ \mathbf{N}(s) \\ \mathbf{B}(s) \end{bmatrix} = \begin{bmatrix} 0 & \kappa(s) & 0 \\ -\kappa(s) & 0 & \tau(s) \\ 0 & -\tau(s) & 0 \end{bmatrix} \begin{bmatrix} \mathbf{T}(s) \\ \mathbf{N}(s) \\ \mathbf{B}(s) \end{bmatrix}. \quad (20)$$

This shows that $\kappa(s)$ is the rate of rotation of $\mathbf{T}(s)$ about $\mathbf{B}(s)$ and similarly, $\tau(s)$ is the rate of rotation of $\mathbf{N}(s)$ about $\mathbf{T}(s)$. Written as one matrix equation, the Frenet-Serret formula tells us something very important: given a curvature $\kappa(s)$ and a torsion $\tau(s)$, we can reconstruct our entire curve up to a translation (since we can change the origin) and a rotation (since we can rotate the initial orthonormal triad $\{\mathbf{T}(0), \mathbf{N}(0), \mathbf{B}(0)\}$). Once we have set the location and orientation of that triad, however, the entire curve is determined by only two parameters, not three as one might have thought.

There is a difficulty with the Frenet-Serret frame: when the curvature vanishes, $\mathbf{N}(s)$ is not well defined

and therefore $\mathbf{B}(s)$ and, more importantly, the torsion are not defined either. Thus curves that have straight segments are problematic from the point of view of the Frenet-Serret frame. Moreover, if we consider a helix

$$\mathbf{R}(s) = \left[r \cos\left(\frac{qs}{\sqrt{(qr)^2+1}}\right), r \sin\left(\frac{qs}{\sqrt{(qr)^2+1}}\right), \frac{s}{\sqrt{(qr)^2+1}} \right] \quad (21)$$

then the curvature and torsion are constant:

$$\kappa(s) = \frac{q^2 r}{1+(qr)^2}, \quad \tau(s) = \frac{q}{1+(qr)^2}. \quad (22)$$

As $r \rightarrow 0$ we approach a straight line. However, though the curvature vanishes in this limit, the torsion does not! This is a problematic feature of the Frenet-Serret frame—since the torsion is the magnitude of the derivative of $\mathbf{N}(s)$, it is only a meaningful quantity when $\mathbf{N}(s) = \mathbf{T}'(s)/\kappa(s)$ is unambiguously defined, and this requires that $\kappa(s) \neq 0$. In Sec. III.B, we will offer a different frame that does not suffer from this problem.

Knowing that there are only two parameters needed to describe a space curve, we can now augment Eq. (1) to include other effects. One interesting effect (Harris *et al.*, 1999) is the behavior of *chiral* polymers. While the curvature does not distinguish between left- and right-handed helices, we can see from Eq. (22) that the torsion is sensitive to the sign of q . Thus we might add terms to Eq. (1) to favor a particular chirality for the stiff polymer. We might be tempted to write

$$E = E_{\text{curv}} + E^* = \int_0^L ds \left\{ \frac{A}{2} \kappa^2(s) + \frac{B}{2} [\tau(s) - \tau_0]^2 \right\}. \quad (23)$$

Though this energy appears acceptable, and favors an average torsion $\tau = \tau_0$, it only accentuates the ambiguity of τ when $\kappa = 0$. Moreover, since Eq. (1) is a functional of $\mathbf{R}(s)$, we should construct the new energy only in terms of $\mathbf{R}(s)$ and its derivatives. In addition to $[\mathbf{R}''(s)]^2 = [\mathbf{T}'(s)]^2$, we can also construct the term (Kamien *et al.*, 1997; Marko, 1997)

$$E^* = -\frac{\alpha}{2} \int_0^L ds \mathbf{T}(s) \cdot [\mathbf{T}'(s) \times \mathbf{T}''(s)]. \quad (24)$$

From the Frenet-Serret formula, we then have

$$E = \int_0^L ds \left\{ \frac{A}{2} \kappa^2(s) + \frac{C}{4} \kappa^4(s) - \frac{\alpha}{2} \kappa^2(s) \tau(s) + \frac{\beta}{2} \kappa^2(s) \tau^2(s) \right\}, \quad (25)$$

where A , C , α , and β are all positive. Note that in the special case of constant curvature and torsion, $[\mathbf{T}(s)$

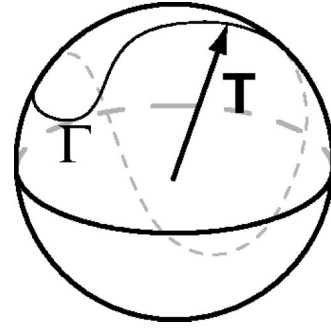


FIG. 2. The tangent spherical map. The tangent vector of the curve \mathbf{T} traces out a curve Γ on the unit sphere. The length of Γ is the integrated curvature along the original curve.

$\times \mathbf{T}''(s)]^2 = \kappa^2 \tau^2$. We have added this term and the extra quartic term to E for reasons that will become clear in the following. This form does not suffer from the torsion ambiguity: when $\kappa(s) = 0$ the torsion drops out of Eq. (25). One can minimize this energy for the helix (21), which has constant curvature and torsion, to find the ground-state conformation of a chiral polymer. For appropriate parameter values, the tendency for torsion can overcome the tendency to be straight and both κ and τ will be nonvanishing.

III. GLOBAL THEORY OF CURVES

A. Fenchel's theorem: energetic bounds on closed curves and knots

When we consider the energetics of closed curves, it is clear that the curvature cannot vanish everywhere (or the curve will not close on itself). We can actually establish a lower bound on the curvature energy by appealing to a theorem on the total curvature of the curve. Not only will this be useful when considering closed polymer loops, the technology of the proof will help us later on when we consider the geometry of surfaces.

Fenchel's theorem states that the total curvature of a closed curve is at least 2π . This is not unreasonable: a circle of radius R has curvature $1/R$ and so the total curvature is $2\pi R/R = 2\pi$. In general, other closed curves only have *more* curvature (in fact, only planar convex curves have an integrated curvature of exactly 2π). The statement of the theorem is

$$\oint_0^L \kappa(s) ds \geq 2\pi \quad (26)$$

for any closed curve, where, as usual, s is the arc length of the curve. Before we prove this fact, note that via the Cauchy-Schwarz inequality we have

$$\oint_0^L (1) \kappa(s) ds \leq \sqrt{\oint_0^L \kappa^2(s) ds} \sqrt{\oint_0^L 1^2 ds}. \quad (27)$$

Squaring both sides and using Eq. (26) we find that the curvature energy (1) satisfies

$$E_{\text{curv}} = \frac{1}{2} k_B T L_p \oint_0^L \kappa^2(s) ds \geq 2 \pi^2 k_B T \frac{L_p}{L}. \quad (28)$$

To prove Fenchel's theorem we introduce the tangent spherical image (Millman and Parker, 1977) of the curve: we take the tail of the unit tangent vector $\mathbf{T}(s)$ at $\mathbf{R}(s)$ and map it to the center of the unit sphere. As shown in Fig. 2, the tip of the tangent vector then traces out a curve Γ on the surface of the unit sphere as s goes from 0 to L . The differential of length of the curve on the tangent sphere is $\|\mathbf{T}(s)'\| ds$, in analogy with Eq. (13). But the magnitude of \mathbf{T}' is just the curvature, so we find that the length of Γ on the unit sphere is

$$l = \int_0^L \kappa(s) ds, \quad (29)$$

the total curvature of our original curve $\mathbf{R}(s)$. Now we note that by definition of $\mathbf{T}(s)$,

$$\mathbf{R}(s) - \mathbf{R}(0) = \int_0^s \mathbf{T}(s') ds', \quad (30)$$

so if the curve is closed, $\mathbf{R}(L) = \mathbf{R}(0)$, and the left-hand side of Eq. (30) is $\mathbf{0}$. This is really three equations, one for each component of \mathbf{T} . It tells us that the curve Γ on the tangent sphere can never be in only one hemisphere. In order for the curve to close, it must turn around. But if the tangent spherical map were in one hemisphere, we could take it to be the upper hemisphere where $T_z(s) > 0$. Yet if $T_z(s)$ is always positive then its integral cannot vanish! Therefore, given any hemisphere on the tangent sphere, the curve must be in it and its complement. But this means that Γ must be *at least* the length of the equator, 2π . We have thus proven Eq. (26). The value of this result is not just the bound on the energetics of a closed polymer. It has introduced us to the tangent map. As we will see in the following it is very useful to take vectors off of curves and surfaces and translate them to the center of the unit sphere. This was our first taste of this procedure.

In closing this section we mention the Fary-Milnor theorem which pertains to the integrated curvature of a non-self-intersecting closed knot. Not surprisingly, this theorem states that a knot has to go around at least twice:

$$\oint_0^L \kappa(s) ds \geq 4\pi. \quad (31)$$

Using the same reasoning that led to Eq. (28), we have

$$k_B T L_p \oint_0^L \kappa^2(s) ds \geq (4\pi)^2 k_B T \frac{L_p}{L} \quad (32)$$

for a knotted closed curve.

B. Mermin-Ho relation: basis vectors to the rescue

Up until this point, we have been able to study the geometry of lines without the introduction of a frame: that is, all our results rely on the original tangent vector

$\mathbf{T}(s)$ and its derivatives—at no point were we required to choose a basis or coordinate system to calculate any quantities. Unfortunately, we cannot continue our discussion of lines and we *certainly* cannot discuss surfaces without introducing some more paraphernalia. We need to introduce a set of spatially varying basis vectors in order to define quantities in addition to the tangent vector, the curvature, and the torsion. We need to make sure, however, that all of our *physical* quantities do not depend on our arbitrary choice. In this section we will derive and explain the Mermin-Ho relation (Mermin and Ho, 1976), originally derived in the context of superfluid $^3\text{He-A}$, a phase characterized by an order parameter with two directions, a “long” direction and a “short” direction perpendicular to it. Though this section is the most technical, it is also straightforward. Fortunately, once we have established this result we will be able to make use of it again and again in the following.

Often we have a vector $\mathbf{n}(\mathbf{x})$ defined everywhere on a surface or in space. It could be the director of a liquid crystal, the normal to a surface or some other field of interest. We now are interested in a vector $\mathbf{N}(\mathbf{x})$ which is always perpendicular to $\mathbf{n}(\mathbf{x})$. This vector might point to the nearest neighbor or in some special direction on the surface. To define this vector, we introduce two new unit vectors, $\mathbf{e}_1(\mathbf{x})$ and $\mathbf{e}_2(\mathbf{x})$, which are defined to be perpendicular to $\mathbf{n}(\mathbf{x})$ and each other so that $\{\mathbf{e}_1(\mathbf{x}), \mathbf{e}_2(\mathbf{x}), \mathbf{n}(\mathbf{x})\}$ is an orthonormal triad and $\mathbf{e}_1(\mathbf{x}) \times \mathbf{e}_2(\mathbf{x}) = \mathbf{n}(\mathbf{x})$. A vector which is defined relative to the spatially varying plane defined by $\mathbf{e}_1(\mathbf{x})$ and $\mathbf{e}_2(\mathbf{x})$ will change not only because its true direction changes (relative to a fixed triad) but also because the basis vectors change. If $\mathbf{N}(\mathbf{x})$ is a unit vector then $\mathbf{N}(\mathbf{x}) = \cos[\theta(\mathbf{x})]\mathbf{e}_1(\mathbf{x}) + \sin[\theta(\mathbf{x})]\mathbf{e}_2(\mathbf{x})$ is always perpendicular to \mathbf{n} . Its derivatives in the i direction are vectors as well. In the plane normal to \mathbf{n} their components are

$$\mathbf{e}_1(\mathbf{x}) \cdot \partial_i \mathbf{N}(\mathbf{x}) = -\sin\theta(\mathbf{x})[\partial_i \theta(\mathbf{x}) - \mathbf{e}_1(\mathbf{x}) \cdot \partial_i \mathbf{e}_2(\mathbf{x})], \quad (33)$$

$$\begin{aligned} \mathbf{e}_2(\mathbf{x}) \cdot \partial_i \mathbf{N}(\mathbf{x}) &= \cos\theta(\mathbf{x})[\partial_i \theta(\mathbf{x}) + \mathbf{e}_2(\mathbf{x}) \cdot \partial_i \mathbf{e}_1(\mathbf{x})] \\ &= \cos\theta(\mathbf{x})[\partial_i \theta(\mathbf{x}) - \mathbf{e}_1(\mathbf{x}) \cdot \partial_i \mathbf{e}_2(\mathbf{x})], \end{aligned} \quad (34)$$

where we have used the fact that $\mathbf{e}_1(\mathbf{x}) \cdot \mathbf{e}_2(\mathbf{x}) = 0$ in the final equality. Notice that the derivatives of $\mathbf{N}(\mathbf{x})$ depend on both gradients of $\theta(\mathbf{x})$ as well as derivatives of the spatially varying basis vectors $\mathbf{e}_i(\mathbf{x})$. Since the basis vectors were chosen arbitrarily, one might be concerned that the derivatives of $\mathbf{N}(\mathbf{x})$ are poorly defined. However, there is a concomitant change in $\theta(\mathbf{x})$ whenever the basis vectors change so that the gradients of $\mathbf{N}(\mathbf{x})$ are well defined. The problem is with gradients of $\theta(\mathbf{x})$.

To disentangle the arbitrary dependence on basis vectors, we start by considering a vector $\mathbf{N}_0(\mathbf{x})$ which is constant in the instantaneous plane perpendicular to $\mathbf{n}(\mathbf{x})$ so that Eqs. (33) and (34) both vanish. For $\mathbf{N}_0(\mathbf{x})$ to be constant in the plane perpendicular to $\mathbf{n}(\mathbf{x})$, $\theta(\mathbf{x})$ must be equal to some $\theta_0(\mathbf{x})$ with $\nabla \theta_0(\mathbf{x}) = \mathbf{e}_1(\mathbf{x}) \cdot \nabla \mathbf{e}_2(\mathbf{x}) \equiv \boldsymbol{\Omega}(\mathbf{x})$, where the last equality defines a new vector field $\boldsymbol{\Omega}(\mathbf{x}) = \mathbf{e}_1(\mathbf{x}) \cdot \nabla \mathbf{e}_2(\mathbf{x})$ called the *spin connection*. Note that we can only solve this equation for $\theta_0(\mathbf{x})$

if $\nabla \times \Omega = 0$. If we now consider an arbitrary vector perpendicular to $\mathbf{n}(\mathbf{x})$, $\mathbf{N}(\mathbf{x})$, we should focus not on gradients of $\theta(\mathbf{x})$, but rather on $[\theta(\mathbf{x}) - \theta_0(\mathbf{x})]$, which is a measure of how much $\theta(\mathbf{x})$ deviates from its “constant” value. But $\nabla \theta_0(\mathbf{x}) = \Omega(\mathbf{x})$, so $\nabla[\theta(\mathbf{x}) - \theta_0(\mathbf{x})] = \nabla \theta(\mathbf{x}) - \Omega(\mathbf{x})$. Thus we see that by subtracting $\Omega(\mathbf{x})$ from $\nabla \theta(\mathbf{x})$, we remove that part of $\theta(\mathbf{x})$ that is induced by a spatially varying basis. Moreover, even if $\nabla \times \Omega$ is non-vanishing, we can generalize this discussion to form the combination $\mathbf{D}\theta(\mathbf{x}) \equiv \nabla \theta(\mathbf{x}) - \Omega(\mathbf{x})$, the *covariant derivative*. It measures the true changes in $\theta(\mathbf{x})$, relative to $\theta_0(\mathbf{x})$. If $[\theta(\mathbf{x}) - \theta_0(\mathbf{x})]$ is a smooth field then $\nabla \times \nabla[\theta(\mathbf{x}) - \theta_0(\mathbf{x})] = 0$, or

$$\nabla \times \nabla \theta(\mathbf{x}) = \nabla \times \Omega(\mathbf{x}). \quad (35)$$

We might be tempted to set the left-hand side equal to zero, since usually the curl of a gradient vanishes. However, as we are about to show, the curl of $\Omega(\mathbf{x})$ is not always zero! This means that there must be some sort of singularities or *defects* in $\theta(\mathbf{x})$ [and $\theta_0(\mathbf{x})$ —the difference $\theta(\mathbf{x}) - \theta_0(\mathbf{x})$ is smooth]. We will talk more about defects in Sec. V.B.

The i th component of the curl is

$$\begin{aligned} [\nabla \times \Omega(\mathbf{x})]_i &= \epsilon_{ijk} \partial_j [e_1^\alpha(\mathbf{x}) \partial_k e_2^\alpha(\mathbf{x})] \\ &= \epsilon_{ijk} [\partial_j e_1^\alpha(\mathbf{x})] [\partial_k e_2^\alpha(\mathbf{x})] \\ &\quad + e_1^\alpha(\mathbf{x}) \epsilon_{ijk} \partial_j \partial_k e_2^\alpha(\mathbf{x}) \\ &= \epsilon_{ijk} [\partial_j e_1^\alpha(\mathbf{x})] [\partial_k e_2^\alpha(\mathbf{x})] + 0, \end{aligned} \quad (36)$$

where the last term vanishes due to the antisymmetry of ϵ_{ijk} and we have used indices to make the calculation unambiguous.¹ We have used the Einstein summation convention of dropping summation signs for repeated indices. Unless otherwise indicated, an index should be summed over if it appears twice in any formula. Both the greek and roman indices run from 1 to 3. Now consider the object $\partial_j e_1^\alpha(\mathbf{x})$. Since $\mathbf{e}_1(\mathbf{x})$ is a unit vector, its derivative is perpendicular to it. Therefore we can write $\partial_j e_1^\alpha(\mathbf{x})$ in terms of the basis vectors $n^\alpha(\mathbf{x})$ and $e_2^\alpha(\mathbf{x})$:

$$\partial_j e_1^\alpha(\mathbf{x}) = A_j(\mathbf{x}) n^\alpha(\mathbf{x}) + B_j(\mathbf{x}) e_2^\alpha(\mathbf{x}). \quad (37)$$

Similarly, we can do the same for $\partial_k e_2^\alpha(\mathbf{x})$:

$$\partial_k e_2^\alpha(\mathbf{x}) = C_k(\mathbf{x}) n^\alpha(\mathbf{x}) + D_k(\mathbf{x}) e_1^\alpha(\mathbf{x}). \quad (38)$$

Putting these expressions into Eq. (36) and using the orthogonality of our triad, we find

$$[\nabla \times \Omega(\mathbf{x})]_i = \epsilon_{ijk} A_j(\mathbf{x}) C_k(\mathbf{x}). \quad (39)$$

By taking the dot product of Eq. (37) with $\mathbf{n}(\mathbf{x})$, we find $A_j(\mathbf{x}) = n^\beta(\mathbf{x}) \partial_j e_1^\beta(\mathbf{x}) = -e_1^\beta(\mathbf{x}) \partial_j n^\beta(\mathbf{x})$. Similarly, $C_k(\mathbf{x}) = -e_2^\gamma(\mathbf{x}) \partial_k n^\gamma(\mathbf{x})$ and so

$$[\nabla \times \Omega(\mathbf{x})]_i = e_1^\beta(\mathbf{x}) e_2^\gamma(\mathbf{x}) \epsilon_{ijk} [\partial_j n^\beta(\mathbf{x})] [\partial_k n^\gamma(\mathbf{x})]. \quad (40)$$

We see then that the curl of $\Omega(\mathbf{x})$ does not always vanish, but it appears that it depends on our arbitrary vectors $\mathbf{e}_1(\mathbf{x})$ and $\mathbf{e}_2(\mathbf{x})$. Note, however, that if we interchange the indices β and γ in Eq. (40) that it is equivalent to interchanging j and k , which would introduce a minus sign. Thus we have

$$\begin{aligned} [\nabla \times \Omega(\mathbf{x})]_i &= \frac{1}{2} [e_1^\beta(\mathbf{x}) e_2^\gamma(\mathbf{x}) - e_1^\gamma(\mathbf{x}) e_2^\beta(\mathbf{x})] \\ &\quad \times \epsilon_{ijk} [\partial_j n^\beta(\mathbf{x})] [\partial_k n^\gamma(\mathbf{x})] \\ &= \frac{1}{2} \epsilon_{\alpha\beta\gamma} n^\alpha(\mathbf{x}) \epsilon_{ijk} \partial_j n^\beta(\mathbf{x}) \partial_k n^\gamma(\mathbf{x}), \end{aligned} \quad (41)$$

where we have used the orthonormality of $\{\mathbf{e}_1(\mathbf{x}), \mathbf{e}_2(\mathbf{x}), \mathbf{n}(\mathbf{x})\}$. This is the celebrated Mermin-Ho relation (Mermin and Ho, 1976). Note that because of the antisymmetry of ϵ_{ijk} , it is unnecessary to keep the brackets around the gradients of $\mathbf{n}(\mathbf{x})$.

This relation between $\mathbf{n}(\mathbf{x})$ and vectors perpendicular to $\mathbf{n}(\mathbf{x})$ is rather remarkable. Note that our choice of $\mathbf{e}_1(\mathbf{x})$ and $\mathbf{e}_2(\mathbf{x})$ is arbitrary so that $\Omega(\mathbf{x})$ is not a constant of the geometry. However, Eq. (41) shows that $\nabla \times \Omega(\mathbf{x})$ only depends on $\mathbf{n}(\mathbf{x})$, and not our choice of basis vectors! This is why the Mermin-Ho relation has been introduced in this section on global properties (Kamien, 1998): if we have a closed curve Γ of length L , then

$$\begin{aligned} &\oint_\Gamma [\nabla \theta(\mathbf{x}) - \Omega(\mathbf{x})] \cdot d\mathbf{R} \\ &= [\theta(L) - \theta(0)] - \iint_M [\nabla \times \Omega(\mathbf{x})] \cdot d\mathbf{S} \\ &= [\theta(L) - \theta(0)] \\ &\quad - \iint_M \frac{1}{2} \epsilon_{\mu\nu\rho} \epsilon_{\alpha\beta\gamma} n^\alpha(\mathbf{x}) \partial_\nu n^\beta(\mathbf{x}) \partial_\rho n^\gamma(\mathbf{x}) dS_\mu, \end{aligned} \quad (42)$$

where we have used Stokes theorem to change an integral around a curve Γ into an integral over a capping surface M . In Eq. (42) $d\mathbf{S}$ is an element of area. Thus if we consider changes in a vector around a closed curve, our choice of $\mathbf{e}_i(\mathbf{x})$ is irrelevant. Though it is necessary that $\mathbf{N}(L) = \mathbf{N}(0)$, $\theta(L) - \theta(0)$ can change by an integral multiple of 2π . The rest of the change in the direction of $\mathbf{N}(\mathbf{x})$ comes from the geometry of $\mathbf{n}(\mathbf{x})$. We will see how this extra term plays an important role in the next section, in the physics of surfaces and the physics of defects in three dimensions.

C. Link, twist, and writhe: dynamics of twist-storing polymers

One of the more interesting stiff polymers is DNA. Though it is well known that it has great biological significance (Watson and Crick, 1953), it is of interest in materials for *at least* two other reasons. First, the persis-

¹The only tensor on which we will rely is the *antisymmetric tensor* ϵ_{ijk} . It is defined by

$$\epsilon_{ijk} = \begin{cases} +1 & \text{if } ijk \text{ is an even permutation of } 123 \\ -1 & \text{if } ijk \text{ is an odd permutation of } 123 \\ 0 & \text{if any two of } i, j, \text{ or } k \text{ are the same.} \end{cases}$$

tence length of DNA is quite long, roughly 50 nm. More importantly, because there are a plethora of enzymes available to cleave DNA, a sample of monodisperse polymers can be prepared much more readily than in any synthetic system. We have already discussed conformations of chiral polymers like DNA in Sec. II. However, DNA has one more interesting element: it can form into a closed loop. This is interesting because DNA is actually a double stranded helix. Therefore the number of times that one strand wraps around the other is fixed when the loop is closed (when the strand is open the helix can unravel). Thus there is a conserved quantity and this leads to a constraint on the possible dynamics of the double strand. More generally, any polymers that cannot unwind along their axis are known as twist-storing polymers.

When we have two closed curves Γ and Γ' , we can assign a linking number Lk to them which counts the number of times one loop passes through the other. There is a simple way to calculate this given the two curves $\mathbf{R}(s)$ and $\mathbf{R}'(s')$ using an analogy with Ampère's law. If we think of the first curve Γ as being a wire carrying a current I , then we know that the line integral of the generated magnetic field \mathbf{B} around the closed curve Γ' is $4\pi nI$, where n is the number of times that the current passes through the closed loop.² Setting $I=1$, we have

$$Lk = n = \frac{1}{4\pi} \oint_{\Gamma'} \mathbf{B}(\mathbf{x}') \cdot d\mathbf{x}'. \quad (43)$$

We may calculate the resulting field $\mathbf{B}(\mathbf{x}')$ from the wire by use of the Biot-Savart Law:

$$\mathbf{B}(\mathbf{x}') = \oint_{\Gamma} \frac{d\mathbf{l} \times \hat{\mathbf{r}}}{r^2} = \oint_{\Gamma} \frac{d\mathbf{x} \times [\mathbf{x}' - \mathbf{x}]}{|\mathbf{x}' - \mathbf{x}|^3}. \quad (44)$$

Putting these together we find that (with the curves having length L and L' , respectively)

$$\begin{aligned} Lk &= \frac{1}{4\pi} \oint_{\Gamma} \oint_{\Gamma'} \frac{[\mathbf{x} - \mathbf{x}'] \cdot (d\mathbf{x} \times d\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3} \\ &= \frac{1}{4\pi} \int_0^L ds \int_0^{L'} ds' \left[\frac{d\mathbf{R}(s)}{ds} \times \frac{d\mathbf{R}'(s')}{ds'} \right] \\ &\quad \cdot \frac{[\mathbf{R}(s) - \mathbf{R}'(s')]}{|\mathbf{R}(s) - \mathbf{R}'(s')|^3} \\ &\equiv G(\Gamma, \Gamma'), \end{aligned} \quad (45)$$

the last equality defines $G(\Gamma, \Gamma')$, the *Gauss invariant*. Thus two closed curves that cannot separate and rejoin must keep this invariant constant.

While this may be elegant, it is not especially useful when studying molecules like DNA. At long length scales, DNA appears as a single filament. It would be useful to recast the linking number in terms of the single polymer picture of DNA. To do this, we consider two

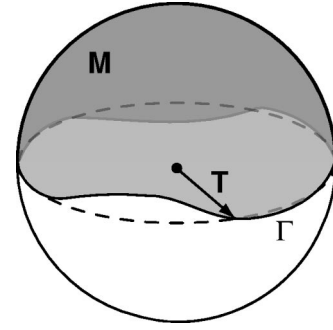


FIG. 3. The tangent spherical map. The area of the patch M on the sphere is, through Stokes theorem, the line integral around Γ of $\mathbf{e}_1 \cdot \nabla \mathbf{e}_2$.

curves that are close together (White, 1969; Fuller, 1971). The first curve is $\mathbf{R}(s)$, while the second curve is

$$\mathbf{R}'(s') = \mathbf{R}(s') + \epsilon \mathbf{u}(s'), \quad (46)$$

where, as is the tradition, ϵ is a small number and $\mathbf{u}(s')$ is a unit vector that is perpendicular to the tangent vector $d\mathbf{R}(s')/ds = \mathbf{T}(s')$. Inserting these expressions into Eq. (45), we have

$$\begin{aligned} Lk &= \frac{1}{4\pi} \oint_0^L ds \int_0^{L'} ds' \mathbf{T}(s) \\ &\quad \times \left[\mathbf{T}(s') + \epsilon \frac{d\mathbf{u}(s')}{ds'} \right] \cdot \frac{[\mathbf{R}(s) - \mathbf{R}(s') - \epsilon \mathbf{u}(s')]}{|\mathbf{R}(s) - \mathbf{R}(s') - \epsilon \mathbf{u}(s')|^3}. \end{aligned} \quad (47)$$

We now want to take ϵ to zero. While the numerator in Eq. (47) does not make this a problem, note that the denominator vanishes whenever $s=s'$. Thus as long as $|s-s'| \geq \delta$ we can take $\epsilon \rightarrow 0$ (Kleinert, 1990). We thus have

$$\begin{aligned} Lk &= \frac{1}{4\pi} \int_0^{L'} ds' \left\{ \int_0^{s'-\delta} ds + \int_{s'+\delta}^L ds \right\} \\ &\quad \times \frac{\mathbf{T}(s) \times \mathbf{T}(s') \cdot [\mathbf{R}(s) - \mathbf{R}(s')]}{|\mathbf{R}(s) - \mathbf{R}(s')|^3} \\ &\quad + \frac{1}{4\pi} \int_0^{L'} ds' \int_{s'-\delta}^{s'+\delta} ds \mathbf{T}(s) \\ &\quad \times \left[\mathbf{T}(s') + \epsilon \frac{d\mathbf{u}(s')}{ds'} \right] \cdot \frac{[\mathbf{R}(s) - \mathbf{R}(s') - \epsilon \mathbf{u}(s')]}{|\mathbf{R}(s) - \mathbf{R}(s') - \epsilon \mathbf{u}(s')|^3}. \end{aligned} \quad (48)$$

The first integral bears a resemblance to the Gauss invariant, while the second integral depends on the vector $\mathbf{u}(s)$. Since δ is small, we can expand $\mathbf{R}(s)$ around s' to find

$$\mathbf{R}(s) = \mathbf{R}(s') + (s-s')\mathbf{T}(s') + \dots \quad (49)$$

and

$$\mathbf{T}(s) = \mathbf{T}(s') + (s-s') \frac{d\mathbf{T}(s')}{ds} + \dots \quad (50)$$

Inserting this into the second integrand and using the fact that $\mathbf{T}(s) \cdot \mathbf{u}(s) = 0$, we find

²For simplicity we work in cgs units with $c=1$.

$$Lk = \frac{1}{4\pi} \oint_0^L ds \oint_0^{L'} ds' \mathbf{T}(s) \times \mathbf{T}(s') \cdot \frac{\mathbf{R}(s) - \mathbf{R}(s')}{|\mathbf{R}(s) - \mathbf{R}(s')|^3} - \frac{1}{4\pi} \oint_0^{L'} ds' \int_{s'-\delta}^{s'+\delta} ds \times \frac{\left[\mathbf{T}(s') + (s-s') \frac{d\mathbf{T}(s')}{ds} \right] \times \left[\mathbf{T}(s') + \epsilon \frac{d\mathbf{u}(s')}{ds} \right] \cdot \boldsymbol{\epsilon} \mathbf{u}(s')}{[(s-s')^2 + \epsilon^2]^{3/2}}, \quad (51)$$

where we have taken $\epsilon \rightarrow 0$ in the first integral and have used $\mathbf{T}(s') \times \mathbf{T}(s') = 0$ in the second. Note that the s dependence of the second integral is manifest and we may do the integration over s , let ϵ go to zero *first* and then let δ go to zero. We have separated the link into two integrals. The first is called the *writhe*, Wr , and it only depends on the curve $\mathbf{R}(s)$. The second is called the *twist*, Tw , and it depends on $\mathbf{u}(s)$. We have

$$Wr = \frac{1}{4\pi} \oint_0^L ds \oint_0^L ds' \mathbf{T}(s) \times \mathbf{T}(s') \cdot \frac{\mathbf{R}(s) - \mathbf{R}(s')}{|\mathbf{R}(s) - \mathbf{R}(s')|^3}, \quad (52)$$

$$Tw = \frac{1}{2\pi} \oint_0^L ds \mathbf{T}(s) \cdot \left[\mathbf{u}(s) \times \frac{d\mathbf{u}(s)}{ds} \right] \quad (53)$$

and thus we arrive at White and Fuller's celebrated result (White, 1969; Fuller, 1971):

$$Lk = Tw + Wr. \quad (54)$$

Though the expression for writhe (52) bares a strong resemblance to that for the Gauss invariant (45), it is not the same. In the Gauss invariant we were considering two different curves that did not touch and there was no need to expunge a singularity from the integration. The expression for writhe, on the other hand, is for the same curve and is only defined in the limit described above. Note further that the writhe is nonlocal, while the twist is local. This is the price we must pay: the link is topological, the writhe depends only on the backbone $\mathbf{R}(s)$ but is nonlocal, and the twist is local but we must know about $\mathbf{u}(s)$, i.e., the other curve.

Note that the expression for the twist (53) suggests a use of the results of Sec. III.B. Since $\mathbf{u}(s)$ is perpendicular to $\mathbf{T}(s)$, we may expand it in a set of basis vectors $\mathbf{e}_1(s)$ and $\mathbf{e}_2(s)$ both perpendicular to $\mathbf{T}(s)$, so that $\mathbf{u}(s) = \cos \theta(s) \mathbf{e}_1(s) + \sin \theta(s) \mathbf{e}_2(s)$. Then

$$\mathbf{T}(s) \cdot \left[\mathbf{u}(s) \times \frac{d\mathbf{u}(s)}{ds} \right] = \partial_s \theta(s) - \mathbf{e}_1(s) \cdot \partial_s \mathbf{e}_2(s), \quad (55)$$

where we have used the fact that $\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{T}$ (plus cyclic permutations). But this is the form we considered in the discussion of the change in a vector. Following the argument at the end of the previous section, we find that

$$\begin{aligned} Tw &= \frac{1}{2\pi} \oint_0^L ds [\partial_s \theta(s) - \mathbf{e}_1(s) \cdot \partial_s \mathbf{e}_2(s)] \\ &= m - \iint_M \frac{1}{4\pi} \epsilon_{\mu\nu\rho} \epsilon_{\alpha\beta\gamma} T^\alpha(\mathbf{x}) \partial_\nu T^\beta(\mathbf{x}) \partial_\rho T^\gamma(\mathbf{x}) dS_\mu \\ &= Lk - \tilde{W}r, \end{aligned} \quad (56)$$

where m is an integer [since $\theta(L) - \theta(0) = 2\pi m$]. We would be tempted to identify the second integral $\tilde{W}r$ as the writhe, since we have an equation that reads just as Eq. (54). This is not always correct and depends on our choice of basis vectors. However, it does establish a weaker result of Fuller's (Fuller, 1978) that the writhe is the last term in Eq. (56) *mod* 1. As we will see in the next section, the integral has a geometric meaning: it is the area swept out by the tangent curve on the tangent spherical map, as shown in Fig. 3.

The difficulty with writhe is that it is nonlocal and therefore cannot be easily included in a local set of dynamical laws—the conformation of the entire curve must be known to calculate writhe. Moreover, writhe is only defined for closed curves and so the identity (54) would not appear to apply to open strands. While the total amount of writhe including the integral part is important for calculating ground states of a particular twisted ribbon (Julicher, 1994; Marko and Siggia, 1994; Fain *et al.*, 1996; Fain and Rudnick, 1999) when considering changes in writhe, the constant integer part is less important. If the time scale for the diffusion of twist along the polymer is long enough, then even an open strand should feel the constraint of conserved link.

To this end, a useful result for the change in writhe of a curve as a function of time t (Aldinger *et al.*, 1995),

$$\partial_t Wr(t) = \frac{1}{2\pi} \oint ds \mathbf{T}(s,t) \cdot [\partial_t \mathbf{T}(s,t) \times \partial_s \mathbf{T}(s,t)], \quad (57)$$

can be used. Note that this result follows from the preceding discussion: if we choose a coordinate system, then (57) (multiplied by dt) is the differential of area swept out by the tangents of two closed curves with tangent vectors $\mathbf{T}(s,0)$ and $\mathbf{T}(s,dt)$. Since the twist is really the local torsional strain of the polymer, we denote the twist density as $\omega(s,t)$ and then we have

$$\begin{aligned} \partial_t Lk &= \oint ds \{ \partial_t \omega(s,t) + \partial_t \mathbf{T}(s,t) \cdot [\partial_s \mathbf{T}(s,t) \\ &\quad \times \mathbf{T}(s,t)] \} + \partial_t n, \end{aligned} \quad (58)$$

where n is the integer difference between Wr and $\tilde{W}r$. Since continuous evolution cannot lead to discontinuous changes in the integer n and since link is conserved we have

$$0 = \oint ds \{ \partial_t \omega + \partial_t \mathbf{T} \cdot [\partial_s \mathbf{T} \times \mathbf{T}] \}. \quad (59)$$

This conservation law need not be satisfied locally: the curve can twist in one place and writhe at some distant

location to satisfy Eq. (59). In addition, the integer can change if the curve develops cusps and evolves in a non-smooth way. Because physics is local, however, we might expect that linking number is locally conserved and changes via a “link current” j . This would lead to a local conservation law

$$\partial_s j = \partial_t \omega + \partial_t \mathbf{T} \cdot [\partial_s \mathbf{T} \times \mathbf{T}], \quad (60)$$

which satisfies Eq. (59). This conservation law enforces total link conservation since $\oint ds \partial_s j = 0$, yet allows for local deviations in the twist and writhe (Kamien, 1998).

This geometrically inspired conservation law has been verified in numerical experiments (Goldstein *et al.*, 1998; Wolgemuth *et al.*, 2000) which show that there are two modes of relaxation in the dissipative dynamics: *twirling*, in which the link relaxes through torsional modes along the polymer (like a speedometer cable), and *whirling*, in which the link relaxes through crank-shaft-like motions of the entire chain.

IV. LOCAL THEORY OF SURFACES

A. Area element: minimal surfaces

While it is easy to generalize the equation of a curve $\mathbf{R}(s)$ to an equation for a surface $\mathbf{X}(s_1, s_2)$, it turns out that the theory of surfaces is much richer than that for curves. In the first place, there is no way to talk about arc length when defining s_1 and s_2 , so we will take them to be arbitrary coordinates of the surface. We can, however, consider the area of a small patch of our surface. To do this, we need to construct vectors tangent to the surface. As shown in Fig. 4, if we have two vectors $\mathbf{a}(s_1, s_2)$ and $\mathbf{b}(s_1, s_2)$ in the surface, then they span a surface area $\Delta S = |\mathbf{a}(s_1, s_2)| |\mathbf{b}(s_1, s_2)| \sin \theta$, where θ is the angle between the two vectors. We recognize this expression for ΔS as the magnitude of the cross product $\mathbf{a}(s_1, s_2) \times \mathbf{b}(s_1, s_2)$. Since both of these vectors are tan-

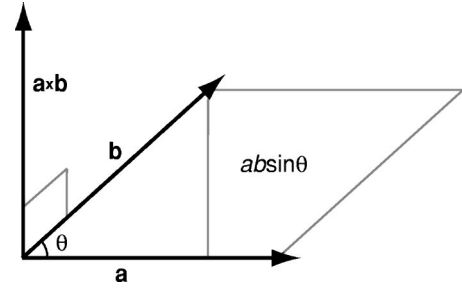


FIG. 4. The magnitude of the cross product of two vectors is the area of the parallelogram swept out by them.

gent to the surface, their cross product is parallel to the unit layer normal $\mathbf{n}(s_1, s_2)$. Thus we have

$$\Delta S = |\mathbf{n}(s_1, s_2) \cdot [\mathbf{a}(s_1, s_2) \times \mathbf{b}(s_1, s_2)]|. \quad (61)$$

But, with this construction, we can also construct the unit layer normal out of $\mathbf{a}(s_1, s_2)$ and $\mathbf{b}(s_1, s_2)$:

$$\mathbf{n}(s_1, s_2) = \pm \frac{\mathbf{a}(s_1, s_2) \times \mathbf{b}(s_1, s_2)}{\|\mathbf{a}(s_1, s_2) \times \mathbf{b}(s_1, s_2)\|}. \quad (62)$$

Putting this expression together with Eq. (61), we have

$$\begin{aligned} \Delta S &= \frac{[\mathbf{a}(s_1, s_2) \times \mathbf{b}(s_1, s_2)] \cdot [\mathbf{a}(s_1, s_2) \times \mathbf{b}(s_1, s_2)]}{\|\mathbf{a}(s_1, s_2) \times \mathbf{b}(s_1, s_2)\|} \\ &= \sqrt{[\mathbf{a}(s_1, s_2) \times \mathbf{b}(s_1, s_2)]^2} \\ &= \sqrt{\mathbf{a}^2(s_1, s_2) \mathbf{b}^2(s_1, s_2) - [\mathbf{a}(s_1, s_2) \cdot \mathbf{b}(s_1, s_2)]^2}. \end{aligned} \quad (63)$$

We now need to construct two, nonparallel vectors $\mathbf{a}(s_1, s_2)$ and $\mathbf{b}(s_1, s_2)$. The surface provides us with two: $\mathbf{a}(s_1, s_2) = \partial_{s_1} \mathbf{X}(s_1, s_2) ds_1$ and $\mathbf{b}(s_1, s_2) = \partial_{s_2} \mathbf{X}(s_1, s_2) ds_2$. This tells us that the differential area element is

$$dS = \sqrt{[\partial_{s_1} \mathbf{X}(s_1, s_2)]^2 [\partial_{s_2} \mathbf{X}(s_1, s_2)]^2 - [\partial_{s_1} \mathbf{X}(s_1, s_2) \cdot \partial_{s_2} \mathbf{X}(s_1, s_2)]^2} ds_1 ds_2. \quad (64)$$

Often we also need the vector surface element $d\mathbf{S} = \mathbf{n}(s_1, s_2) dS$, for instance when calculating electric and magnetic flux. The area of the whole surface M is simply

$$A = \int \int_M \sqrt{(\partial_{s_1} \mathbf{X})^2 (\partial_{s_2} \mathbf{X})^2 - (\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \mathbf{X})^2} ds_1 ds_2. \quad (65)$$

What happens if we choose to parametrize the surface in terms of a new set of coordinates? If we have two new coordinates σ_1 and σ_2 defined by $\sigma_i = \sigma_i(s_1, s_2)$, then by the chain rule

$$\partial_{s_i} \mathbf{X} = \frac{\partial \mathbf{X}}{\partial s_i} = \frac{\partial \mathbf{X}}{\partial \sigma_j} \frac{\partial \sigma_j}{\partial s_i} = \frac{\partial \sigma_j}{\partial s_i} \partial_{\sigma_j} \mathbf{X}, \quad (66)$$

where, as usual, there is an implicit sum over j . Because of this sum, it appears that it would be somewhat tedious to reexpress Eq. (65) in terms of the new coordinates. However, we note that if we define a matrix \mathbf{g}

$$\mathbf{g}(s_1, s_2) = \begin{bmatrix} \partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \mathbf{X} & \partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \mathbf{X} \\ \partial_{s_2} \mathbf{X} \cdot \partial_{s_1} \mathbf{X} & \partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \mathbf{X} \end{bmatrix}, \quad (67)$$

then the expression in the radical of Eq. (65) is simply the determinant of this matrix. In component form, this matrix is $g_{ij}(s_1, s_2) = \partial_{s_i} \mathbf{X}(s_1, s_2) \cdot \partial_{s_j} \mathbf{X}(s_1, s_2)$. This matrix is known as the *metric tensor* which some people find useful (Minser *et al.*, 1973). Moreover, we note that the transformation (66) amounts to matrix multiplication of $\mathbf{g}(s_1, s_2)$:

$$g_{ij}(s_1, s_2) = \frac{\partial \sigma_k}{\partial s_i} \frac{\partial \sigma_m}{\partial s_j} \tilde{g}_{km}(\sigma_1, \sigma_2), \quad (68)$$

where $\tilde{g}_{km}(\sigma_1, \sigma_2)$ is the corresponding matrix in the new coordinate system. If we define a new matrix \mathbf{O} by $O_{ij} = \partial \sigma_j / \partial s_i$, then $\mathbf{g} = \mathbf{O}^T \tilde{\mathbf{g}} \mathbf{O}$. Thus when we change coordinates, we have

$$\begin{aligned} A &= \iint_M \sqrt{\det \mathbf{g}} ds_1 ds_2 \\ &= \iint_M \sqrt{\det \mathbf{O}^T \tilde{\mathbf{g}} \mathbf{O}} ds_1 ds_2 \\ &= \iint_M \sqrt{\det \tilde{\mathbf{g}}} |\det \mathbf{O}| ds_1 ds_2. \end{aligned} \quad (69)$$

But the determinant of \mathbf{O} is just the Jacobian determinant of the transformation from s_i to σ_i , and so $|\det \mathbf{O}| ds_1 ds_2 = d\sigma_1 d\sigma_2$ or

$$A = \iint_M \sqrt{\det \mathbf{g}} ds_1 ds_2 = \iint_M \sqrt{\det \tilde{\mathbf{g}}} d\sigma_1 d\sigma_2, \quad (70)$$

and so the area is invariant under coordinate transformations. This invariance (known as *diffeomorphism covariance* by the *cognoscenti*) is useful as it allows us to choose the most convenient set of coordinates for a given problem. One often used choice is the so-called *Monge gauge* or *height representation*, where s_1 and s_2 are chosen to be the x and y components of the surface vector $\mathbf{X}(s_1, s_2)$ and the z component is a function of x and y :

$$\begin{aligned} x &= X_1(s_1, s_2) = s_1, \\ y &= X_2(s_1, s_2) = s_2, \\ z &= X_3(s_1, s_2) = h(s_1, s_2) = h(x, y). \end{aligned} \quad (71)$$

It is straightforward to check that the expression for the area (65) becomes the more familiar expression:

$$A = \iint_M \sqrt{1 + \left(\frac{\partial h}{\partial x}\right)^2 + \left(\frac{\partial h}{\partial y}\right)^2} dx dy. \quad (72)$$

The height representation requires that for every x and y , there is only one height or, in other words, the surface can have no overlaps. Sometimes it is necessary to break the surface up into different regions or *patches* in order to make a good height representation.

The height representation proves useful when considering the energetics of fluid membranes. By *fluid* we mean that the membrane has no internal structure and the molecules in it can flow freely. Often we are interested in the shape and dynamics of these membranes when they are under a uniform applied tension. Soap films are a common example, though in many cases the interface between two fluids or two distinct phases can also be thought of as a membrane under tension. The free energy of this system is simply $F = \varsigma A$, where ς is the surface tension and A is the area of the interface. Minimizing the free energy thus amounts to minimizing the area of the surface. The expression for the area in

Eq. (72) looks like an action from classical mechanics. If we define the Lagrange density \mathcal{L} ,

$$\mathcal{L} = \sqrt{1 + \left(\frac{\partial h}{\partial x}\right)^2 + \left(\frac{\partial h}{\partial y}\right)^2}, \quad (73)$$

then the Euler-Lagrange equation is (Goldstein, 1980)

$$\frac{\partial}{\partial x} \frac{\partial \mathcal{L}}{\partial (\partial_x h)} + \frac{\partial}{\partial y} \frac{\partial \mathcal{L}}{\partial (\partial_y h)} = \frac{\partial \mathcal{L}}{\partial h}, \quad (74)$$

or, by inserting Eq. (73),

$$\frac{\partial}{\partial x} \left(\frac{\partial_x h}{\sqrt{1 + (\nabla h)^2}} \right) + \frac{\partial}{\partial y} \left(\frac{\partial_y h}{\sqrt{1 + (\nabla h)^2}} \right) = 0. \quad (75)$$

This equation has an important interpretation in terms of the surface normal. If we construct the two tangent vectors to the surface as we did above, we have

$$\begin{aligned} \mathbf{u}_1 &= \partial_{s_1} \mathbf{X}(s_1, s_2) = [1, 0, \partial_x h], \\ \mathbf{u}_2 &= \partial_{s_2} \mathbf{X}(s_1, s_2) = [0, 1, \partial_y h], \end{aligned} \quad (76)$$

and so the surface normal is parallel to $\mathbf{u}_1 \times \mathbf{u}_2 = [-\partial_x h, -\partial_y h, 1]$. Normalizing this vector gives us $\mathbf{n}(s_1, s_2) = [-\partial_x h, -\partial_y h, 1] / \sqrt{1 + (\nabla h)^2}$ and so Eq. (75) becomes $\nabla \cdot \mathbf{n} = 0$. In other words, a surface which extremizes its area has a divergence-free unit normal. We will address this further in the following section. Such a surface is called a *minimal surface*. Typically, Eq. (75) is written equivalently as (Nitsche, 1989) the minimal surface equation:

$$[1 + (\partial_y h)^2] \partial_x^2 h - 2 \partial_x h \partial_y h \partial_x \partial_y h + [1 + (\partial_x h)^2] \partial_y^2 h = 0. \quad (77)$$

Though we were seeking a surface with a minimum area, the Euler-Lagrange equations only guarantee an extremum—we might have a saddle point or, worse, a maximum! Rigorous proofs that the area is minimized in general are difficult. However, we can prove a restricted result easily. If the projection of the boundary ∂M onto the xy plane is convex, then it is clear that any overhangs on the surface add extra area and that the surface need not “stick out” past the boundary of M projected onto the xy plane. Thus we can use the height representation and solve Eq. (77) subject to a boundary condition $h(x, y) = f(x, y)$ on the boundary in the xy plane. As depicted in Fig. 5, we consider an arbitrary surface \tilde{M} with the *same boundary* as the surface of which we are interested. Since the same range of x and y map out both M and \tilde{M} , we can define $\mathbf{n}(x, y)$ on \tilde{M} and in the volume between the surfaces. Define

$$T = \iint_M \mathbf{n}(\mathbf{x}) \cdot d\tilde{\mathbf{S}}, \quad (78)$$

where $\mathbf{n}(\mathbf{x})$ is the unit normal of the original surface M , and $d\tilde{\mathbf{S}}$ is the vector element of surface area, pointing along the normal to the surface $\tilde{\mathbf{n}}(\mathbf{x})$. If we consider the same integral for the original surface M , then since $\mathbf{n}(\mathbf{x})$

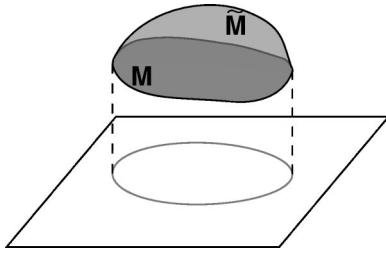


FIG. 5. The original minimal surface \mathbf{M} and the comparison surface $\tilde{\mathbf{M}}$.

is parallel to $d\mathbf{S}$, we have $\mathbf{n}(\mathbf{x}) \cdot d\mathbf{S} = dS$ the element of surface area! Integrating this over the whole surface just gives us the area A . Thus

$$A = \iint_M dS = \iint_M \mathbf{n}(\mathbf{x}) \cdot d\mathbf{S}. \quad (79)$$

Moreover, we have the inequality $\tilde{\mathbf{n}}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) \leq 1$, since the dot product of two unit vectors is the cosine of some angle and is therefore less than or equal to 1. We arrive at

$$T = \iint_{\tilde{M}} \mathbf{n}(\mathbf{x}) \cdot d\tilde{\mathbf{S}} = \iint_{\tilde{M}} \mathbf{n}(\mathbf{x}) \cdot \tilde{\mathbf{n}}(\mathbf{x}) d\tilde{S} \leq \iint_{\tilde{M}} d\tilde{S} = \tilde{A}, \quad (80)$$

where \tilde{A} is the area of the comparison surface. We are almost done. Note that if we construct a closed surface \hat{M} by gluing together M and \tilde{M} along their common boundary, then the integral of $\mathbf{n}(\mathbf{x})$ over the whole surface is

$$\iint_{\hat{M}} \mathbf{n}(\mathbf{x}) \cdot d\hat{\mathbf{S}} = \iint_M \mathbf{n}(\mathbf{x}) \cdot d\mathbf{S} - \iint_{\tilde{M}} \mathbf{n}(\mathbf{x}) \cdot d\tilde{\mathbf{S}} = A - T, \quad (81)$$

where the relative minus sign arises because we want the normal to the closed surface to always point outward. But by Gauss's law, the first integral in Eq. (81) can be converted into an integral over the volume \hat{V} enclosed by \hat{M} , so

$$A - T = \iint_{\hat{M}} \mathbf{n}(\mathbf{x}) \cdot d\hat{\mathbf{S}} = \iiint_{\hat{V}} \nabla \cdot \mathbf{n}(\mathbf{x}) d\hat{V} = 0, \quad (82)$$

since M extremizes the area so $\nabla \cdot \mathbf{n}(\mathbf{x}) = 0$. Using the inequality (80) we have $A = T \leq \tilde{A}$ and thus the comparison surface has an area at least as large as the original

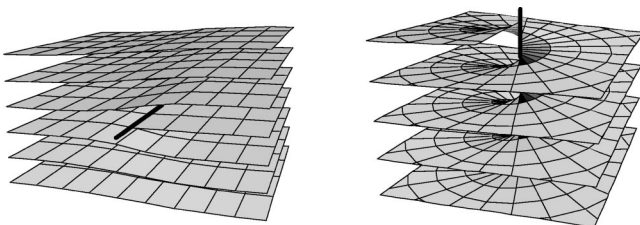


FIG. 6. Edge (left) and screw (right) dislocations in a smectic-A liquid crystal.

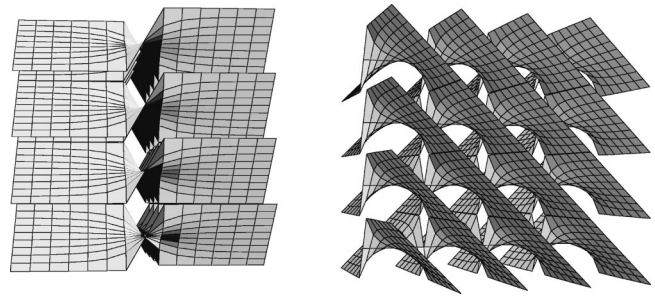


FIG. 7. Scherk's first surface. This minimal surface connects together two layered structures with different orientations.

surface M . Since \tilde{M} was arbitrary we have shown that our surface M minimizes the area.³

Examples of minimal surfaces abound. The smectic-A liquid crystal phase is a one-dimensional crystal, composed of periodic layers of spacing d . Each layer is fluidlike, with no internal ordering. Thus, in some sense, each layer behaves as a minimal surface and the layer spacing is set by a compression modulus (Kamien and Lubensky, 1999). The smectic phase can have topological defects of two types. As shown in Fig. 6, an *edge dislocation* adds or ends a layer in the bulk, while a *screw dislocation* connects adjacent layers together. A screw defect can be represented by the height function

$$h(x, y) = \frac{nd}{2\pi} \tan^{-1} \left(\frac{y}{x} \right), \quad (83)$$

where n is an integer. By choosing n to be an integer, we are guaranteed that when going around the origin once we move up exactly n layers. It is straightforward to check that the screw dislocation is a minimal surface by direct calculation of Eq. (77).

Another nice example also arises in smectic liquid crystals. Scherk's first surface (shown in Fig. 7) (Scherk, 1835) is a minimal surface which smoothly connects two smectic regions which are rotated with respect to each other. It is composed of an infinite set of parallel, screw-like dislocations (Gido *et al.*, 1993; Kamien, 2001) which must be slightly squashed to make the surface satisfy Eq. (77). For those who are interested, the height function is defined for arbitrary rotation angle α :

$$h[x, y; \alpha] \equiv -\sec \left(\frac{1}{2} \alpha \right) \tan^{-1} \left\{ \frac{\tanh \left[\frac{1}{2} x \sin(\alpha) \right]}{\tan \left[y \sin \left(\frac{1}{2} \alpha \right) \right]} \right\}. \quad (84)$$

It is interesting to note that as $\alpha \rightarrow 0$ the height function becomes that of a simple screw dislocation (83) with $nd = 2\pi$.

³In this context, the vector field $\mathbf{n}(\mathbf{x})$ is called a *calibration*. The interested (or disinterested) reader will find this and other results in Morgan (2000).

B. Mean and Gaussian curvature: energetics of membranes

When we discussed curves in three dimensions there were not many options for the definition of curvature: there was only one coordinate on the curve, and if we wanted something that was independent of our parametrization, we need only take derivatives with respect to arc length. As we have seen in the last section, however, surfaces can be reparametrized and have many more degrees of freedom. As a result, there is more than one way to define curvature, in fact there are two.

We have already hinted at one definition in the discussion of minimal surfaces. For curves, we know that a straight line minimizes the distance between two points in space, or in other words, $\kappa(s)=0$ is the equation of a length-minimizing curve. Generalizing this to surfaces, we might define one curvature which is 0 for surfaces that minimize the area spanned by a given boundary. Indeed, this is called the *mean curvature*, H and $H = -\frac{1}{2}\nabla \cdot \mathbf{n}(\mathbf{x})$ where $\mathbf{n}(\mathbf{x})$ is the surface normal. The other type of curvature is the *Gaussian curvature* K , which we will define in the following.

In order to understand this better and to introduce K , we need to step back and consider our surface $\mathbf{X}(s_1, s_2)$. Since the curvature of a line is $\kappa(s) = -\mathbf{T}(s) \cdot \mathbf{N}'(s)$, we would like to take derivatives of the surface normal along tangent directions of the surface. As in Sec. III.B, we now reintroduce the basis vectors perpendicular to $\mathbf{n}(s_1, s_2)$, $\mathbf{e}_1(s_1, s_2)$, and $\mathbf{e}_2(s_1, s_2)$, of unit length and mutually orthogonal. We can take directional derivatives⁴ of $\mathbf{n}(s_1, s_2)$ along $\mathbf{e}_i(s_1, s_2)$, $(\mathbf{e}_i \cdot \nabla)$. Now instead of one number, we have four, which form a matrix:

$$\mathbf{L} = - \begin{bmatrix} \mathbf{e}_1 \cdot [\mathbf{e}_1 \cdot \nabla] \mathbf{n} & \mathbf{e}_2 \cdot [\mathbf{e}_1 \cdot \nabla] \mathbf{n} \\ \mathbf{e}_1 \cdot [\mathbf{e}_2 \cdot \nabla] \mathbf{n} & \mathbf{e}_2 \cdot [\mathbf{e}_2 \cdot \nabla] \mathbf{n} \end{bmatrix}. \quad (85)$$

The two dot products in the matrix are somewhat confusing. To be concrete we write this matrix using its indices:

$$L_{ij} = -e_i^\alpha(s_1, s_2) e_j^\beta(s_1, s_2) \frac{\partial n^\alpha(s_1, s_2)}{\partial x^\beta(s_1, s_2)}. \quad (86)$$

This matrix is also a tensor, known as the *Weingarten map* or *second fundamental form*.⁵ We can diagonalize this matrix via a similarity transform $\mathbf{\Lambda} = \mathbf{S}^{-1} \mathbf{L} \mathbf{S}$. In this diagonal basis the entries of $\mathbf{\Lambda}$ are precisely what we are after: the first entry in the upper left is the derivative of the surface normal along a direction $\hat{\mathbf{e}}_1$. Moreover, since the upper right entry vanishes, this derivative has no

⁴It may seem odd to use the gradient in these definitions since $\mathbf{n}(s_1, s_2)$ is a function of s_1 and s_2 . The derivatives of $\mathbf{n}(s_1, s_2)$ along the surface tangent vectors $\partial_{s_j} \mathbf{X}(s_1, s_2)$ are just $\partial_{s_j} \mathbf{n}(s_1, s_2)$. By writing $\mathbf{e}_i(s_1, s_2)$ as the linear combination $\mathbf{e}_i = A_{ij} \partial_{s_j} \mathbf{X}$, the chain rule implies that $(\mathbf{e}_i \cdot \nabla) = A_{ij} \partial_{s_j}$.

⁵The metric tensor \mathbf{g} is sometimes called the *first fundamental form*.

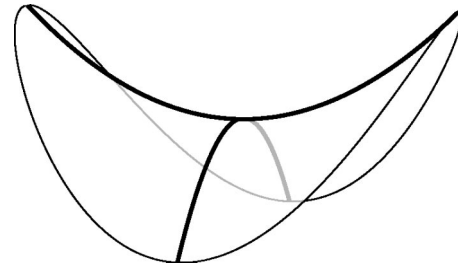


FIG. 8. Curves on a surface (heavy lines) with tangents along the principal directions at the point of intersection.

components along $\hat{\mathbf{e}}_2$. Therefore the upper left corner is the curvature of the curve in the surface at (s_1, s_2) , tangent to $\hat{\mathbf{e}}_1$. Similarly, the lower right entry of $\mathbf{\Lambda}$ gives us another curvature. Note that the two directions $\hat{\mathbf{e}}_1$ and $\hat{\mathbf{e}}_2$ remain orthogonal.⁶ They are known as the *principal directions* on the surface and their associated curvatures κ_1 and κ_2 are the *principal curvatures* (see Fig. 8). Equivalently, we can define the two principal radii of curvature through $R_i = 1/\kappa_i$. We can easily extract these curvatures from the original matrix \mathbf{L} . Note that $\kappa_1 \kappa_2 = \det \mathbf{\Lambda}$ and $\kappa_1 + \kappa_2 = \text{Tr } \mathbf{\Lambda}$. Moreover, since the trace is cyclic and $\det \mathbf{A} \mathbf{B} = \det \mathbf{A} \det \mathbf{B}$, we see that

$$\begin{aligned} \text{Tr } \mathbf{L} &= \kappa_1(s_1, s_2) + \kappa_2(s_1, s_2), \\ \det \mathbf{L} &= \kappa_1(s_1, s_2) \kappa_2(s_1, s_2). \end{aligned} \quad (87)$$

The product of the curvatures is known as the *Gaussian curvature*, $K = \kappa_1 \kappa_2$, while the average of the curvatures is the *mean curvature*, $H = \frac{1}{2} [\kappa_1 + \kappa_2]$. The Gaussian and mean curvature contain all the information to describe the bending of our surface.

The reader may be concerned that our definitions of curvature remain dependent on our choice of basis vectors. This is not the case. Indeed, using the expression (86) we have

$$\begin{aligned} H &= -\frac{1}{2} \sum_{i=1}^2 e_i^\alpha e_i^\beta \frac{\partial n^\alpha}{\partial x^\beta}, \\ K &= (e_1^\alpha e_1^\beta e_2^\gamma e_2^\delta - e_1^\alpha e_2^\beta e_2^\gamma e_1^\delta) \frac{\partial n^\alpha}{\partial x^\beta} \frac{\partial n^\gamma}{\partial x^\delta}. \end{aligned} \quad (88)$$

These expressions still look dependent on \mathbf{e}_i . However, since \mathbf{n} is a unit vector, $n^\alpha n^\beta \partial_\beta n^\alpha = 0$ and so we have

$$\begin{aligned} H &= -\frac{1}{2} [e_1^\alpha e_1^\beta + e_2^\alpha e_2^\beta + n^\alpha n^\beta] \frac{\partial n^\alpha}{\partial x^\beta} \\ &= -\frac{1}{2} \delta^{\alpha\beta} \frac{\partial n^\alpha}{\partial x^\beta} = -\frac{1}{2} \nabla \cdot \mathbf{n}, \end{aligned} \quad (89)$$

where we have used the fact that $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{n}\}$ form an or-

⁶The orthogonality of the two vectors $\hat{\mathbf{e}}_1$ and $\hat{\mathbf{e}}_2$ follows from the fact the $\mathbf{S}^{-1} = \mathbf{S}^T$. In other words, \mathbf{L} can be diagonalized via an orthogonal transformation since it is symmetric: $L_{12} - L_{21} = -(e_1^\alpha e_2^\beta - e_2^\alpha e_1^\beta) \partial_\beta n^\alpha = -\mathbf{n} \cdot [\nabla \times \mathbf{n}] = 0$ when \mathbf{n} is normal to a surface, i.e., $\mathbf{n} = \nabla \phi / |\nabla \phi|$.

thonormal triad. Thus, as we hinted at the beginning of this section, the mean curvature is proportional to the divergence of the surface normal. Note that the sign of H depends on the choice of the sign of \mathbf{n} .

We have to do a little more work on the expression for the Gaussian curvature. The form in Eq. (88) reminds us of the expressions in the derivation of the Mermin-Ho relation. Using the orthonormality of the basis vectors,

$$\begin{aligned} K &= e_1^\alpha e_2^\gamma (e_1^\beta e_2^\delta - e_2^\beta e_1^\delta) \partial_\beta n^\alpha \partial_\delta n^\gamma \\ &= e_1^\alpha e_2^\gamma \epsilon_{\beta\delta\rho} n^\rho \partial_\beta n^\alpha \partial_\delta n^\gamma. \end{aligned} \quad (90)$$

Finally, as in the discussion of Eq. (40), we have

$$K = \frac{1}{2} \epsilon_{\alpha\beta\gamma} n^\gamma \epsilon_{ijk} n^k \partial_i n^\alpha \partial_j n^\beta = \mathbf{n} \cdot [\nabla \times \mathbf{\Omega}]. \quad (91)$$

Moreover, the Gaussian curvature has a simple interpretation. Consider the *normal spherical map* defined by analogy with the tangent spherical map. For each point on the surface we identify a point on the unit sphere which corresponds to the surface normal at that point. This map is also known as the *Gauss map*. Since the surface element on our surface is just $dS_\mu = n_\mu dS$, we see from the discussion at the beginning of Sec. IV.A, that the Gaussian curvature is just the area swept out by the surface normal on the Gauss map as we move along the surface. In other words, if we have a small region M of our surface, then

$$\iint_M K dS = \iint_M \frac{1}{2} \epsilon_{\alpha\beta\gamma} n^\gamma \epsilon_{ijk} \partial_i n^\alpha \partial_j n^\beta dS_k. \quad (92)$$

Thus the Gaussian curvature is the ratio between the infinitesimal area swept out on the Gauss map and the infinitesimal area of the original surface to which it corresponds.

We have seen that the two principal curvatures depend only on the surface normal and not on our choice of coordinates or basis vectors. This is useful when we consider the energetics of fluid membranes. They are called “fluid” because they have no internal structure. Therefore it would be unphysical to build an energy out of anything but the two invariants H and K . Note that if we have an open surface there is no distinction between inside and outside, so the layer normal $\mathbf{n}(s_1, s_2)$ is defined only up to a sign. Though the Gaussian curvature is independent of this sign, the mean curvature is not. To get around this, we insist that the free energy be even in powers of H (Canham, 1970; Helfrich, 1973):

$$F_{\text{CH}} = \iint dS \{2\kappa H^2 + \bar{\kappa} K\}, \quad (93)$$

where κ and $\bar{\kappa}$ are (confusingly) the standard symbols for the bending moduli. This is known as the *Canham-Helfrich* free energy for fluid membranes. Note that they both have units of energy, since the dimensions of K and H^2 cancel the dimensions of the surface. Also note that the integration is done with respect to the actual surface area, so that $dS = \sqrt{\det \mathbf{g}} ds_1 ds_2$. We will see in the fol-

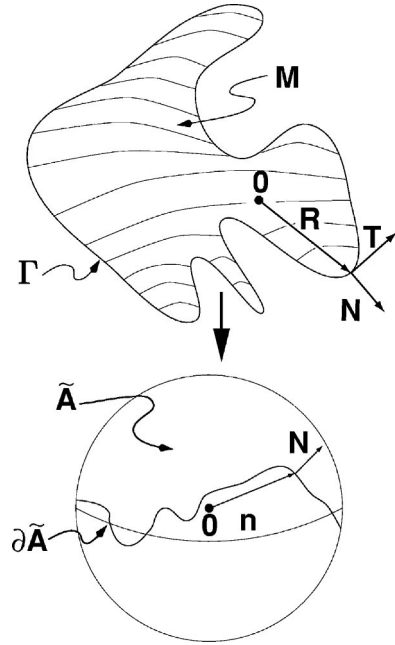


FIG. 9. Geometry for the Gauss-Bonnet theorem. Note that now we are using the Gauss map which maps the surface normal $\mathbf{n}(\mathbf{x})$ unto the unit sphere.

lowing sections that the integral of the Gaussian curvature is a constant and so the term proportional to $\bar{\kappa}$ is usually neglected. We will also discover that not only is the Gaussian curvature independent of our choice of basis vectors, but that it also can be measured with no knowledge of the surface normal $\mathbf{n}(s_1, s_2)$. Finally, if the membrane is *tethered* then it has internal elastic degrees of freedom. These degrees of freedom couple to the geometry and produce a variety of singular structures (DiDonna *et al.*, 2002).

V. GLOBAL THEORY OF SURFACES

A. Gauss-Bonnet theorem: foams on curved surfaces

We have seen that there are two sorts of curvature that we can consider on a surface, the mean and Gaussian curvatures. We have studied both of these by considering deviations of the normal vector to the surface. However, a large part of differential geometry focuses on *intrinsic* properties, those quantities which can be measured without reference to the space in which the manifold is embedded. It turns out that the Gaussian curvature is intrinsic: by measuring the lengths and diameters of small circles entirely in the surface, one can determine K .

We have all the technology necessary to demonstrate this remarkable property. To show this, we consider a patch on our surface M , with boundary ∂M . As in Sec. III.C, we can take the unit normal at each point and map it to the unit sphere. As shown in Fig. 9, when we traverse the patch M on the surface, we traverse a patch \tilde{M} on the unit sphere. As we discussed, the area of \tilde{M} is the integral of the Gaussian curvature of the patch M :

$$\begin{aligned}\bar{A} &= \iint_{\bar{M}} d\bar{S} = \iint_M K dS \\ &= \iint_M \frac{1}{2} \epsilon_{\mu\nu\rho} \epsilon_{\alpha\beta\gamma} n^\alpha(\mathbf{x}) \partial_\nu n^\beta(\mathbf{x}) \partial_\rho n^\gamma(\mathbf{x}) dS_\mu, \quad (94)\end{aligned}$$

where we have used Eq. (91). Now consider the boundary curve ∂M . This curve has a tangent vector $\mathbf{T}(s)$ in space. We know from our discussion of curves that $\mathbf{T}'(s) = \kappa(s)\mathbf{N}(s)$. However, while $\mathbf{T}(s)$ is also tangent to the sphere, $\mathbf{N}(s)$ is, in general, *not*. The normal to the curve, $\mathbf{N}(s)$, can have a component along the normal to the surface $\mathbf{n}(s)$ and perpendicular to it. If we lived in the surface and could only make measurements in the surface, like some kind of bug, then we could only measure the component of the curve's normal in the surface, the *surface normal* $\boldsymbol{\gamma}(s)$:

$$\boldsymbol{\gamma}(s) = \frac{\mathbf{T}'(s) - [\mathbf{T}'(s) \cdot \mathbf{n}(s)]\mathbf{n}(s)}{\kappa(s) \sqrt{1 - [\mathbf{N}(s) \cdot \mathbf{n}(s)]^2}}. \quad (95)$$

This is a unit vector which lies in the plane tangent to the surface and which is perpendicular to $\mathbf{T}(s)$, that is, the tangent to the boundary ∂M of the patch M . We would like to define a curvature which measures the *extra* curvature in the curve, not arising from the curvature of the surface in which it is embedded. By analogy to the Frenet-Serret formulas, we define the *geodesic curvature*⁷ κ_g to be

$$\kappa_g(s) = \mathbf{T}'(s) \cdot \boldsymbol{\gamma}(s) = \kappa(s) \sqrt{1 - [\mathbf{N}(s) \cdot \mathbf{n}(s)]^2}. \quad (96)$$

Since $\{\mathbf{T}(s), \boldsymbol{\gamma}(s), \mathbf{n}(s)\}$ form an orthonormal triad, we have

$$\kappa_g(s) = \mathbf{T}'(s) \cdot [\mathbf{n}(s) \times \mathbf{T}(s)] = \mathbf{n}(s) \cdot [\mathbf{T}(s) \times \mathbf{T}'(s)], \quad (97)$$

a form we have seen before in the context of the Mermin-Ho relation! Since the unit tangent vector lies in the plane perpendicular to $\mathbf{n}(\mathbf{x})$, we may write it in terms of our arbitrary basis $\mathbf{e}_1(\mathbf{x})$ and $\mathbf{e}_2(\mathbf{x})$, introduced in Sec. III.B, $\mathbf{T}(s) = \cos \theta \mathbf{e}_1 + \sin \theta \mathbf{e}_2$. Using Eq. (97) to calculate κ_g , we find that

$$\kappa_g(s) = \partial_s \theta(s) - e_1^\alpha(s) \partial_s e_2^\alpha(s), \quad (98)$$

where we have used the fact that for a unit vector $\mathbf{u}(s)$, $\partial_s[\mathbf{u}(s)]^2 = 2\mathbf{u}(s) \cdot \partial_s \mathbf{u}(s) = 0$. In analogy with Eq. (42), we can integrate the geodesic curvature around the boundary of M to find

$$\oint_{\partial M} \kappa_g(s) ds = \oint_{\partial M} [\nabla \theta(\mathbf{x}) - \boldsymbol{\Omega}(\mathbf{x})] \cdot d\mathbf{R}. \quad (99)$$

As with our discussion of the relation between link, twist, and writhe, we may transform the line integral of $\boldsymbol{\Omega}(\mathbf{x})$ into a surface integral over M so that

⁷We may also define the *normal curvature*, $\kappa_n(s)$ which is the curvature of our embedded curve that is imposed by the surface. We define $\kappa_n(s) = \mathbf{T}'(s) \cdot \mathbf{n}(s)$ and thus the total curvature satisfies $\kappa(s) = \sqrt{\kappa_n^2(s) + \kappa_g^2(s)}$.

$$\oint_{\partial M} \partial_s \theta(s) ds = \oint_{\partial M} \kappa_g(s) ds + \iint_M K dS, \quad (100)$$

where we have used Eq. (41) to rewrite the final integrand as the Gaussian curvature (94). Finally, if the boundary curve does not intersect itself, then the tangent vector rotates around $\mathbf{n}(\mathbf{x})$ exactly once, so $\theta(s)$ changes by 2π around the curve. We have thus established the *Gauss-Bonnet theorem*:

$$\iint_M K ds + \oint_{\partial M} \kappa_g ds = 2\pi. \quad (101)$$

By integrating the geodesic curvature around a closed loop, we can calculate the integrated Gaussian curvature that we surround. As the loop shrinks ever smaller around a point \mathbf{x}_0 , we can calculate the Gaussian curvature at that point through division by the surface area enclosed. This is remarkable since the *geodesic curvature can be measured without the use of the layer normal!* In other words, since the geodesic curvature is intrinsic, so is the Gaussian curvature! We can take this result a step further by considering regions that have discrete angles in their boundaries (i.e., polygons). When there is a sharp bend in the boundary curve ∂M , we cannot calculate the geodesic curvature $\kappa_g(s)$. We can, however, integrate the geodesic curvature along the smooth parts of the boundary. If there are j sharp bends with angles $\Delta \theta_j$, respectively, then when we integrate around the boundary as in Eq. (99) we will have a deficit of $\sum_j \Delta \theta_j$, or, in other words, the smooth part of $\theta(s)$ does not have to change by 2π for the curve to come around. We thus have

$$\iint_M K dS + \oint_{\partial M} \kappa_g ds + \sum_j \Delta \theta_j = 2\pi, \quad (102)$$

where we understand that the integral around the boundary should be broken into smooth segments of the boundary. The jump angles account for the discontinuities. Note that we can have lines for which $\kappa_g(s) = 0$. These are the “straight” lines on the surface and are called *geodesics*. If we were to build a polygon with n -sides, all of which are geodesics, then Eq. (102) reads

$$2\pi - \sum_j \Delta \theta_j = \iint_M K dS. \quad (103)$$

Since the sum of the external angles $\Delta \theta_j$ of a polygon in flat space is 2π , we see that the Gaussian curvature is a measure of the excess (or deficit) angle in a polygon. If the Gaussian curvature is positive the sum of the external angles is smaller than 2π so the sum of the *internal* angles $\sum_j (\pi - \Delta \theta_j) = (n-2)\pi + \iint_M K dS$ is larger than we might expect. For example, imagine the triangle on the sphere connecting the North Pole (90°N), Pontianak, Indonesia (roughly 0°N , $109^\circ 20'\text{W}$) and Loolo, The Congo (roughly 0°S , $19^\circ 20'\text{W}$) along great circles. At each vertex of this triangle the arcs of great circles meet at 90° and so the sum of the interior angles is $3\pi/2$. Since $n=3$ for a triangle, we discover that $\iint_M K dS = \pi/2$ for this triangle that covers $\frac{1}{8}$ of the globe. However, we

know that for a sphere of radius R , the Gaussian curvature is just $1/R^2$ and so $\iint K dS = \frac{1}{8}(4\pi R^2/R^2) = \pi/2$, as we expect!

There are many beautiful uses and examples of the interplay between Gaussian curvature and geodesic curvature in physical systems. For instance, Avron and Levine (Avron and Levine, 1992) have considered dry foams on curved, two-dimensional surfaces. In this context, “dry” refers to the fact that there is no fluid between adjacent bubbles so that the walls between them can be treated as lines and the vertices may be treated as points. There are two key ingredients to the physics of dry foams: surface tension and pressure. In two dimensions the surface tension amounts to a line tension σ along the interfaces between neighboring bubbles. If the soap film is uniform then the line tension is constant as well. Whenever the lines meet, mechanical equilibrium must be maintained. The pressure in the bubbles exerts a force on their boundaries. If two cells are separated by a boundary line then the pressure difference ΔP must be balanced by the boundary. As with the derivation of the wave equation, the force exerted by the boundary curve Γ is $\sigma \mathbf{T}'(s)$. However, the pressure exerts forces only within the surface and so we are only interested in the component of the force in the plane of the surface. According to the above discussion, that force is $\sigma \kappa_g(s) \boldsymbol{\gamma}(s)$, and so the magnitude of the force is $\sigma \kappa_g(s) = \Delta P$. In the case of planar foams, $\kappa_g(s) = \kappa(s)$ and this is known as the Young-Laplace law. To model the diffusion of gas from one cell to the other, we assume a simple dynamics where

$$\frac{dN}{dt} = -C \sum_j \Delta P_j l_j, \quad (104)$$

where N is the number of gas molecules in the cell of interest, ΔP_j is the pressure difference between it and its j th neighbor, l_j is the length of the boundary separating the cell from its j th neighbor, and $C > 0$ is a diffusion constant. This dynamics captures the simple idea that if a bubble is higher pressure than its neighbors so that $\Delta P_j > 0$, then it loses gas, while if it has lower pressure it gains gas. Using the generalization of the Young-Laplace law and the Gauss-Bonnet theorem we have

$$\begin{aligned} \frac{dN}{dt} &= -C \sigma \oint_{\partial M} \kappa_g(s) ds \\ &= C \sigma \left\{ \iint_M K dS + \sum_j (\pi - \alpha_j) - 2\pi \right\}, \end{aligned} \quad (105)$$

where α_j are the internal angles.

For flat surfaces with $K=0$ it is known that a hexagonal honeycomb network of boundaries minimizes the length of the cell walls (Morgan, 2000; Hales, 2001). If we imagine starting with a stable configuration on a flat membrane and distorting the membrane, we should take each internal angle to be $2\pi/3$. Doing so Avron and Levine (Avron and Levine, 1992) found

$$\frac{dN}{dt} = C \sigma \left\{ \frac{\pi}{3}(n-6) + \iint_M K dS \right\}. \quad (106)$$

If we seek a stationary configuration so that N is time independent, we see that for flat membranes $K=0$ and only bubbles with $n=6$ sides are stationary: hexagons. If the surface has positive curvature, $K > 0$, and there is an instability. To make $(dN/dt)=0$, we see that $n < 6$. However, even if an area and an n were found to make N time independent, we can see that there is an instability: if the area grows then (dN/dt) becomes positive so the bubble grows *more*. Likewise, if the bubble shrinks then the integral over K gets smaller and so gas flows out of the bubble and it shrinks some more. Thus on a positively curved surface the only stable situation is one for which one bubble overtakes the whole system. On the other hand, when the surface has negative Gaussian curvature, it follows from a similar argument that when the area of the bubble increases $(dN/dt) < 0$ and when the area shrinks $(dN/dt) > 0$. In the critical case of a flat membrane the stationary solution is a hexagon and it is neither stable nor unstable towards growth (Avron and Levine, 1992).

B. Euler characteristic and the genus: defects on surfaces

We can take Eq. (102) one step further by considering a closed surface. In this case we can integrate the Gaussian curvature over the whole manifold. If we triangulate the entire surface to form a net, then we can use the Gauss-Bonnet theorem to establish a relation between the topology of the network and the total Gaussian curvature. At each vertex there is a total angle of 2π which is divided into the internal angles of the triangles meeting at that point. Each face contributes π to the total angle at the vertices in addition to the excess angle from that face, $\iint K dS$. Adding all the triangles together we have

$$2\pi V = \pi F + \iint_M K dS, \quad (107)$$

where V is the number of vertices and F is the number of faces. Each face contributes three edges E , but each edge is shared by two triangles, so $3F=2E$. We then have

$$V - E + F = \frac{1}{2\pi} \iint_M K dS. \quad (108)$$

This is a remarkable result. As we show in the bottom two graphs of Fig. 10, if we remove an edge from our network ($E \rightarrow E-1$), two faces join into one ($F \rightarrow F-1$), and we lose two vertices ($V \rightarrow V-2$), so $V-E+F$ does not change! Likewise if we add an edge $V-E+F$ is unchanged. Thus our result does not require the use of triangles. This invariant is known as the *Euler characteristic* $\chi = V - E + F$. Note that if we have any network that can be deformed into a sphere without cut-

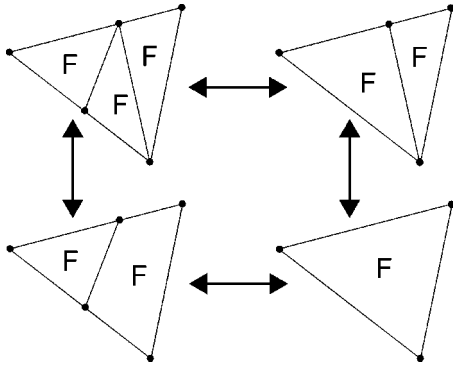


FIG. 10. Removing or adding an edge from a graph maintains the value of the Euler characteristic $V - E + F$.

ting the edges or changing the vertices, then χ can be calculated using Eq. (108) for a perfect sphere, where the surface area is $4\pi R^2$ and the Gaussian curvature is $1/R^2$. Thus we find that

$$V - E + F = \chi = 2 \tag{109}$$

for a network with the topology of a sphere. Suppose we take two faces on the sphere, deform them to be the same triangle, and place them together, face-to-face. In the process we lose three edges ($E \rightarrow E - 3$), two faces ($F \rightarrow F - 2$), and three vertices ($V \rightarrow V - 3$) so $\chi = 0$. What have we done? We have made a doughnut (or a torus) with one handle and have reduced the Euler characteristic by 2. Clearly, any time we add a handle χ is reduced by 2. If we define the *genus* g to be the number of handles of the surface, then

$$\frac{1}{2\pi} \iint_M K dS = V - E + F = \chi = 2(1 - g). \tag{110}$$

Though the Gaussian curvature was a geometric property, when integrated over the entire surface it becomes a topological invariant, independent of the local geometry.

The Euler characteristic can be used to understand the topology of defects on closed surfaces. Suppose that we have a unit vector field $\mathbf{v}(\mathbf{x})$ living in the local tangent plane to a closed surface of genus g . Since $\mathbf{v}(\mathbf{x})$ lies in the tangent plane, we have $\mathbf{v}(\mathbf{x}) = \cos \theta(\mathbf{x}) \mathbf{e}_1(\mathbf{x}) + \sin \theta(\mathbf{x}) \mathbf{e}_2(\mathbf{x})$, where we have reintroduced our vectors $\mathbf{e}_1(\mathbf{x})$ and $\mathbf{e}_2(\mathbf{x})$ which are everywhere perpendicular to the unit normal $\mathbf{n}(\mathbf{x})$ of the surface. We will try to cover the surface with a vector field that is single valued. In other words, if we integrate derivatives of $\mathbf{v}(\mathbf{x})$ around a closed curve, we should get back the same vector. But as we discussed in Sec. III.B, to ensure this property we should focus on the covariant derivative $\mathbf{D}\theta(\mathbf{x})$. We will have a single-valued vector field when the curl of this derivative vanishes. In this case Eqs. (41) and (91) give

$$\iint_M [\nabla \times \nabla \theta(\mathbf{x})] \cdot d\mathbf{S} = \iint_M K dS = 2\pi\chi. \tag{111}$$

Usually $\nabla \times \nabla \theta = 0$, but this result tells us that it is not true on an arbitrary genus surface M .

Without presenting a detailed discussion on topological defects (Chaikin and Lubensky, 1995), we can see that $\nabla \times \nabla \theta(\mathbf{x})$ does not vanish in the presence of a defect configuration and that on a two-dimensional surface:

$$\nabla \times \nabla \theta(\mathbf{x}) = \sum_i m_i \delta^2(\mathbf{x} - \mathbf{x}_i), \tag{112}$$

where m_i is the charge of the defect and \mathbf{x}_i is the corresponding position. To see this, we consider

$$\theta(\mathbf{x}) = m \tan^{-1}\left(\frac{y}{x}\right). \tag{113}$$

It is straightforward to calculate $\nabla \theta(\mathbf{x})$ and then to integrate the gradient around a closed curve γ that contains the origin. We have

$$2\pi m = \oint_\gamma \nabla \theta(\mathbf{x}) \cdot d\mathbf{x} = \iint_M \nabla \times \nabla \theta(\mathbf{x}) \cdot d\mathbf{S}, \tag{114}$$

where γ is the boundary of M . Since we may arbitrarily shrink the path γ around the origin and the integral remains constant, it must be that $\nabla \times \nabla \theta = m \delta^2(\mathbf{x})$. Generalizing this to an arbitrary collection of defects leads us to Eq. (112). Using Eq. (112) in Eq. (111) gives us

$$\sum_i m_i = 2(1 - g) = \chi. \tag{115}$$

This result is known as the *Poincaré-Brouwer theorem*. Since a vector field must be single valued on the surface, m_i must be an integer so that $\cos \theta$ and $\sin \theta$ are well defined. This result tells us, for instance, that a vector field on the surface of a sphere must have two $+1$ defects or one $+2$ defect. On a torus, however, no defects are necessary.

There is another, more geometric way to establish Eq. (115). We first argue that two vector fields on the same surface must have the same total topological charge. Consider two vector fields $\mathbf{u}(\mathbf{x})$ and $\mathbf{v}(\mathbf{x})$. We can triangulate the surface so that each triangle contains at most one defect in $\mathbf{u}(\mathbf{x})$ and one defect of $\mathbf{v}(\mathbf{x})$ and that every defect is in some triangle. In each triangle we may now integrate $\nabla \theta_u(\mathbf{x})$ and $\nabla \theta_v(\mathbf{x})$ around the triangle boundary. The total topological charge χ_u of $\mathbf{u}(\mathbf{x})$ is

$$\chi_u = \sum_{\text{triangles}} \sum_{\text{edges}} \int \nabla \theta_u(\mathbf{x}) \cdot d\mathbf{x} \tag{116}$$

and similarly for $\mathbf{v}(\mathbf{x})$. Note that to establish the angles $\theta_u(\mathbf{x})$ and $\theta_v(\mathbf{x})$, we must choose basis vectors $\mathbf{e}_i(\mathbf{x})$. Since we know that $\nabla \theta_u(\mathbf{x})$ depends on our choice of these vectors, the integrals of $\nabla \theta_u(\mathbf{x})$ around each triangle depend on the underlying geometry through the Mermin-Ho relation (41). More importantly, each vector field on M may require a different choice of basis vectors $\mathbf{e}_i(\mathbf{x})$, since a defect in any particular triangle forces a constraint on $\mathbf{e}_i(\mathbf{x})$. Thus it is not, in general, possible

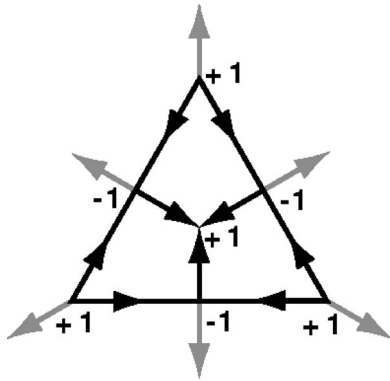


FIG. 11. A vector field on a triangular patch. Note that the $+1$ defects at the vertices and at the center force -1 defects on the edges.

to calculate χ_u and χ_v using the same set of $\mathbf{e}_i(\mathbf{x})$. However, the difference between χ_u and χ_v is just

$$\chi_u - \chi_v = \sum_{\text{triangles}} \sum_{\text{edges}} \int \nabla[\theta_u(\mathbf{x}) - \theta_v(\mathbf{x})] \cdot d\mathbf{x} = 0. \quad (117)$$

The difference vanishes since the angle between $\mathbf{u}(\mathbf{x})$ and $\mathbf{v}(\mathbf{x})$ is independent of our choice of basis vectors. Therefore when we sum over all triangles we get each edge twice but in opposite directions. Since we do not need the basis vectors to define the integrals, it is clear that the integrals in opposite directions cancel and we get 0. Thus any two vector fields have the same topological charge.

We will now count the defects of a vector field that we construct on triangulated surface, via the following rules: (i) put a defect with $+1$ charge at each vertex, (ii) put a defect with $+1$ charge at each center, and (iii) put a defect with -1 charge on each edge. We can see that this will produce a consistent vector field everywhere else, as shown in Fig. 11. Adding together the charges we have

$$\chi_u = V - E + F = \chi, \quad (118)$$

and so the topological charge is the Euler characteristic for any vector field on the surface.

A striking example of the effect of these defects can be seen when there is a coupling between the bending of the surface and the configuration of the vector field (MacKintosh and Lubensky, 1991; Park *et al.*, 1992). For instance, if we have a vector field living on a closed surface then it must have two $+1$ defects. But because the defects are discrete, when we consider Eq. (111) we find that the Gaussian curvature K must also be confined to the core of the defect since we may also integrate Eq. (41) over any submanifold of M . But then all the Gaussian curvature sits at the defects. If instead of viewing Eq. (41) as a constraint, we can write it as an energy

$$F = \frac{1}{2} \iint dS A [\nabla \theta(\mathbf{x}) - \mathbf{\Omega}(\mathbf{x})]^2 + F_{\text{CH}}, \quad (119)$$

where F_{CH} is the free energy for the surface (93), and A

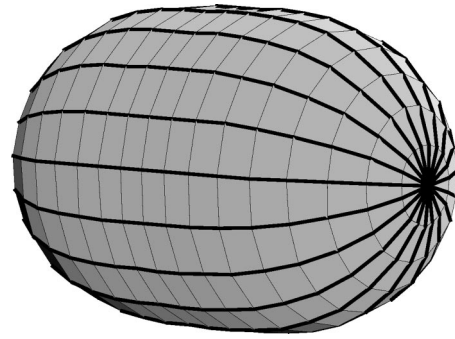


FIG. 12. A vesicle with some sort of vector order parameter. The Poincaré-Brouwer theorem assures that there are two defects in the vector field, drawn as the dark lines on the vesicle. The vesicle distorts so that the Gaussian curvature is larger in the vicinity of the defects.

is the spin stiffness. On a flat surface we can always choose $\mathbf{\Omega} = \mathbf{0}$ and so the first term in Eq. (119) just describes the usual Goldstone mode of broken rotational invariance. On a curved surface the inclusion of $\mathbf{\Omega}$ is necessary in order to make the free energy uniquely defined. The energy (119) now represents a competition between surface bending and the requirement of two $+1$ topological defects which serve as a source for Gaussian curvature. Both of these energies vanish for the plane, but if we restrict our topology to a sphere, for instance, then neither term can be made to vanish everywhere. Balancing these two effects, a vesicle would form an oblate shape, as shown in Fig. 12.

VI. THREE DIMENSIONS AND BEYOND

We have covered the basic elements of the geometry of curves and surfaces. Things become significantly more abstract in three dimensions because there is no analog of the normal vector, binormal vector, or surface normal: the three-dimensional system is not embedded in a higher-dimensional space. Though it might be hard to visualize how something can have curvature without an embedding space, we already know that it is possible. Recall that the Gaussian curvature could be measured without any knowledge of the surface normal. Quantities with that property are called *intrinsic*. The general theory of relativity relies on intrinsic quantities to describe the curvature of four-dimensional space-time (Misner *et al.*, 1973)—there is no reference to a larger space in which our universe lives. Though we will not go into any mathematical detail, it is worthwhile to describe two examples from soft materials.

The first is the blue phases of chiral liquid crystals (Meiboom *et al.*, 1981; Wright and Mermin, 1989). These are phases in which there is a three-dimensional, periodic modulation of the nematic director $\mathbf{n}(\mathbf{x})$ with a length scale comparable to that of visible light. These usually exist only over a narrow temperature range and are stabilized by an often neglected term in the nematic free energy. The Frank free energy for a nematic liquid crystal is

$$\begin{aligned}
 F = & \frac{1}{2} \int d^3x \{ K_1 (\nabla \cdot \mathbf{n})^2 + K_2 [\mathbf{n} \cdot (\nabla \times \mathbf{n})]^2 \\
 & + K_3 [\mathbf{n} \times (\nabla \times \mathbf{n})]^2 \\
 & + 2K_{24} \nabla \cdot [(\mathbf{n} \cdot \nabla) \mathbf{n} - (\nabla \cdot \mathbf{n}) \mathbf{n}] \}. \quad (120)
 \end{aligned}$$

The standard elastic constants K_1 , K_2 , and K_3 are a measure of the energy cost for splay, twist, and bend modes, respectively (Chaikin and Lubensky, 1995). The last term with elastic constant K_{24} is known as the *saddle-splay* term. Though it is a total derivative, when there are defects present it can contribute to the energy. The blue phase is riddled with precisely those defects that contribute to saddle splay. The remarkable thing about this term is that it is precisely the Gaussian curvature as in Eq. (91):

$$\mathbf{n}(\mathbf{x}) \cdot \nabla \times \boldsymbol{\Omega}(\mathbf{x}) = -\frac{1}{2} \nabla \cdot [(\mathbf{n} \cdot \nabla) \mathbf{n} - (\nabla \cdot \mathbf{n}) \mathbf{n}]. \quad (121)$$

What does this mean? If there are surfaces to which the nematic director is normal, then the saddle splay is the Gaussian curvature of those surfaces. However, the saddle splay is more general. Even if the director is not a field of layer normals, the saddle splay is a measure of curvature. Indeed, the blue phases can be understood as unfrustrated systems in curved three-dimensional space (Sethna *et al.*, 1983). Projecting this texture into flat space leads to the topological defects in these phases. A compelling and alternative way of viewing the blue phases is to view them as decorations of space filling minimal surfaces (Pansu and Dubois-Violette, 1989). It is the connection between topological defects and curvature that makes it possible to identify and locate defects in lattice simulations of liquid crystals (Priezjev and Pelcovits, 2001).

Another three-dimensional system that can be understood in terms of curvature is the melting and freezing of hard-sphere fluids. In two dimensions, as hard spheres condense from the fluid phase to the crystalline phase, they form close-packed triangles. Eventually these can assemble into a triangular lattice. The situation is not so happy in three dimensions. There, four spheres can close pack into a tetrahedron, but tetrahedra cannot assemble to fill space. As a result, all lattices have a packing density lower than the best local packing. This difficulty can be viewed as a geometric frustration along the same lines as the blue phase. In positively curved space, however, this frustration can also be eliminated (Nelson, 1983; Nelson and Widom, 1983), and tetrahedral close packing can fill space.

OUTLOOK

Though the mathematics described here has a certain elegance and beauty, I hope that I have conveyed the utility of differential geometry in a variety of physical problems. While a geometric description of a system often leads to an intuitive perspective, there are many other arenas in which a geometric formulation of the

problem is not only useful but essential. Put together with statistical mechanics, differential geometry has been and will continue to be a powerful tool in the study of soft materials.

ACKNOWLEDGMENTS

I thank I. Bluestein, B. A. DiDonna, G. M. Grason, M. I. Kamien, W. Y. Kung, T. C. Lubensky, T. R. Powers, P. Ziherl, and especially G. Jungman and P. Nelson for careful and critical reading of this review. This work was supported by NSF Grants DMR01-29804 and INT99-10017, and a gift from L. J. Bernstein.

REFERENCES

- Aldinger, J., I. Klapper, and M. Tabor, 1995, *J. Knot Theory Ramif.* **4**, 343.
- Avron, J. E., and D. Levine, 1992, *Phys. Rev. Lett.* **69**, 208.
- Bensimon, D., D. Dohmi, and M. Mezard, 1998, *Europhys. Lett.* **42**, 97.
- Canham, P., 1970, *J. Theor. Biol.* **26**, 61.
- Chaikin, P. M., and T. C. Lubensky, 1995, *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge, England).
- David, F., 1989, in *Statistical Mechanics of Membranes and Surfaces: Jerusalem Winter School for Theoretical Physics*, edited by D. R. Nelson, T. Piran, and S. Weinberg (World Scientific, Singapore).
- de Gennes, P.-G., 1970, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, NY).
- DiDonna, B. A., T. A. Witten, S. C. Venkataramani, and E. M. Kramer, 2002, *Phys. Rev. E* **65**, 016603.
- Doi, M., and S. F. Edwards, 1986, *The Theory of Polymer Dynamics* (Oxford University Press, New York).
- Fain, B., and J. Rudnick, 1999, *Phys. Rev. E* **60**, 7239.
- Fain, B., J. Rudnick, and S. Östlund, 1996, *Phys. Rev. E* **55**, 7364.
- Fuller, F. B., 1971, *Proc. Natl. Acad. Sci. U.S.A.* **68**, 815.
- Fuller, F. B., 1978, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3557.
- Gido, S. P., J. Gunther, E. L. Thomas, and D. Hoffman, 1993, *Macromolecules* **26**, 4506.
- Goldstein, H., 1980, *Classical Mechanics*, 2nd ed. (Addison-Wesley, Reading, MA).
- Goldstein, R. E., T. R. Powers, and C. H. Wiggins, 1998, *Phys. Rev. Lett.* **80**, 5232.
- Hales, T. C., 2001, *Discrete Comput. Geom.* **25**, 1.
- Harris, A. B., R. D. Kamien, and T. C. Lubensky, 1999, *Rev. Mod. Phys.* **71**, 1745.
- Helfrich, W., 1973, *Z. Naturforsch. C* **28C**, 693.
- Julicher, F., 1994, *Phys. Rev. E* **49**, 2429.
- Kamien, R. D., 1998, *Eur. Phys. J. B* **1**, 1.
- Kamien, R. D., 2001, *Appl. Math. Lett.* **14**, 797.
- Kamien, R. D., and T. C. Lubensky, 1999, *Phys. Rev. Lett.* **82**, 2892.
- Kamien, R. D., T. C. Lubensky, P. Nelson, and C. S. O'Hern, 1997, *Europhys. Lett.* **38**, 237.
- Kleinert, H., 1990, *Path Integrals in Quantum Mechanics Statistics and Polymer Physics* (World Scientific, Singapore), Chap. 16.
- MacKintosh, F. C., and T. C. Lubensky, 1991, *Phys. Rev. Lett.* **67**, 1169.

- Marko, J. F., 1997, *Europhys. Lett.* **38**, 183.
- Marko, J. F., and E. D. Siggia, 1994, *Macromolecules* **24**, 981.
- Marko, J. F., and E. D. Siggia, 1995, *Phys. Rev. E* **52**, 2912.
- Meiboom, S., J. P. Sethna, P. W. Anderson, and W. F. Brinkman, 1981, *Phys. Rev. Lett.* **46**, 1216.
- Mermin, N. D., and T. L. Ho, 1976, *Phys. Rev. Lett.* **36**, 594.
- Millman, R. S., and G. D. Parker, 1977, *Elements of Differential Geometry* (Prentice-Hall, Englewood Cliffs, NJ).
- Misner, C. W., K. S. Thorne, and J. A. Wheeler, 1973, *Gravitation* (Freeman, San Francisco).
- Morgan, F., 2000, *Geometric Measure Theory*, 3rd ed. (Academic Press, San Diego).
- Moroz, J. D., and R. D. Kamien, 1997, *Nucl. Phys. B* **506**, 695.
- Moroz, J. D., and P. Nelson, 1997, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 14418.
- Moroz, J. D., and P. Nelson, 1998, *Macromolecules* **31**, 6333.
- Nelson, D. R., 1983, *Phys. Rev. B* **28**, 5515.
- Nelson, D. R., and M. Widom, 1983, *Nucl. Phys. B* **240**, 113.
- Nelson, P., 1998, *Phys. Rev. Lett.* **80**, 5810.
- Nitsche, J. C. C., 1989, *Lectures on Minimal Surfaces* (translated by J. M. Feinberg) (Cambridge University Press, Cambridge, England).
- Pansu, B., and E. Dubois-Violette, 1989, *Europhys. Lett.* **10**, 43.
- Park, J., T. C. Lubensky, and F. C. MacKintosh, 1992, *Europhys. Lett.* **20**, 279.
- Priezjev, N. V., and R. A. Pelcovits, 2001, *Phys. Rev. E* **64**, 031710.
- Scherk, H. F., 1835, *J. Reine Angew. Math.* **13**, 185 (also see Nitsche, 1989).
- Sethna, J. P., D. C. Wright, and N. D. Mermin, 1983, *Phys. Rev. Lett.* **51**, 467.
- Vologodskii, A. V., S. D. Levene, K. V. Klenin, M. Frank-Kamenetskii, and N. R. Cozzarelli, 1992, *J. Mol. Biol.* **227**, 1224.
- Watson, J. D., and F. H. C. Crick, 1953, *Nature (London)* **171**, 737.
- White, J. H., 1969, *Am. J. Math.* **91**, 693.
- Wolgemuth, C. W., T. R. Powers, and R. E. Goldstein, 2000, *Phys. Rev. Lett.* **84**, 1623.
- Wright, D. C., and N. D. Mermin, 1989, *Rev. Mod. Phys.* **61**, 385.