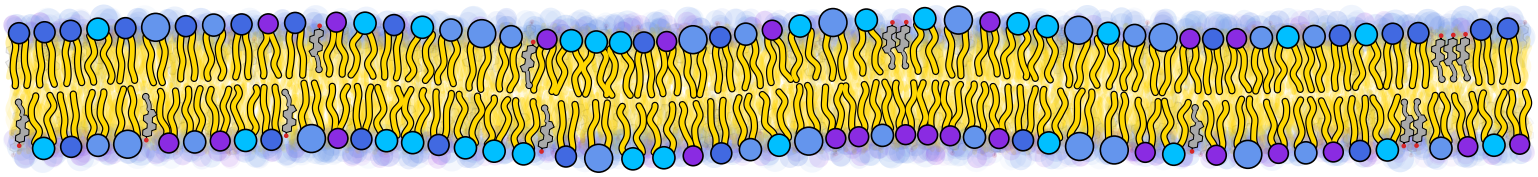


Lipid Membranes

Assembly, Geometry, Elasticity, and Asymmetry

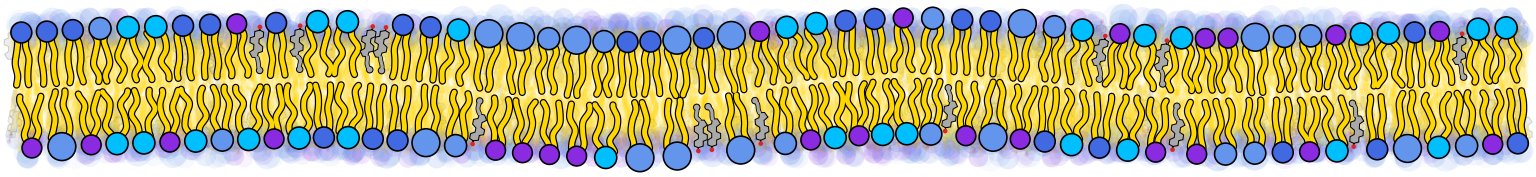
— Version 1.1 —

BY

Markus Deserno

Lecture notes for the 2024 Boulder Summer School on
"Self-Organizing Matter: From Inanimate to the Animate"

Boulder, Colorado, July 1–26, 2024



Contents

1 Setting the stage

1.1 Bonds

How many types of chemical bonds are there?

Those of us who have ever been exposed to a course in chemistry will recall the eagerness with which this kind of classification is sometimes pursued. There are, of course, covalent bonds, ionic bonds, and metallic bonds. But that's not all: covalent bonds can be polar and hence have an "ionic component." They can also be "single," "double," or even "triple" (but we are also admonished not to take these numbers *too* seriously). Down the road we learn of other things, such as hydrogen bonds, which are usually first explained as an oriented electrostatic dipole interaction (so we start to think of them as essentially ionic in nature), until we later learn that there's also a bit of electron sharing happening and hence there's just a *wee* bit of covalent flavor. Then there are van der Waals interactions, which are usually not called bonds at all, but they still somehow can bind things together, and they rely on fluctuating charges—so maybe they are ionic in nature?—but with some quantum origin due to the nature of the fluctuations.

In short, the more you know, the more this gets a real mess.

Far be it from me to belittle the attempts of chemistry to systematize what really is an outstandingly complicated quantum-multibody problem. Nevertheless, overly eager classifications into neat distinct man-made classes are almost never useful in a world where most anything is a matter of degree. Often a more meaningful way to proceed is to focus on a particular property, one that's continuously varying, and then assess how it compares to other quantities of the same dimension that are, somehow, relevant to the problem at hand. We should then contemplate "regimes" and "cross-over regions," but not pristine boundaries.

1.2 Energy scales

A very obvious continuous quantity for "bonding" is to look at the bonding strength: a quantity of dimension "energy" that tells us how much energy is needed to separate the two things that are bonded. What other quantity of dimension "energy" could we compare this to? Well, bonds that are based on quantum mechanical electron sharing all somehow scale with an energy that derives from the quantum mechanical treatment of electrons electrostatically bound to a nucleus. For instance, the Bohr energy—the ground state energy of the hydrogen atom—is given by $E_{\mathrm{Bohr}} = -\frac{1}{2}\alpha^2 m_{\mathrm{e}}c^2$, where $\alpha = e^2/4\pi\varepsilon_0\hbar c \approx 1/137$ is the *fine structure constant* and $m_{\mathrm{e}}c^2 \approx 511\,\mathrm{keV}$ is the *rest energy of an electron*. We find $E_{\mathrm{Bohr}} \approx -13.6\,\mathrm{eV}$. Once we look at molecules, of course we can hardly do any exact calculations, but some approximate variational calculations have been tried early on (James and Coolidge, 1933), finding that the answer is some fraction of this, maybe like a third, about 4.5 eV. And thus "a few electron-volts" is a good reference point for a binding strengths that involves bonding based on quantum-mechanical "sharing" of electrons. Whenever some other energy of comparable magnitude arises, chemical bond-making or -breaking is happening. For instance, the energy of a UV photon of $\lambda = 300\,\mathrm{nm}$ wavelength is $E_{300\,\mathrm{nm}} = hc/\lambda \approx 4.1\,\mathrm{eV}$. This can break important chemical bonds in your body and this is why you should put on sunscreen—especially in places as high up as Boulder where more UV radiation from the sun reaches you.

But there are other scales of interest. Here's one: the thermal energy $k_{\mathrm{B}}T$, which at room temperature has the approximate value

$$k_{\mathrm{B}}T_{\mathrm{room}} \approx 4.1\times10^{-21}\,\mathrm{J} \approx 4.1\,\mathrm{pN\,nm} \approx \frac{1}{40}\mathrm{eV}$$
$$\approx 0.6\,\mathrm{kcal/mol} \approx 2.5\,\mathrm{kJ/mol} \approx 200\,\mathrm{cm}^{-1}\ .$$

Can energy of that magnitude break chemical bonds? Hardly. This is why increasing the temperature typically does not make chemicals fall apart. But we very well know that there are things that come apart at lower energies. Why? Maybe because they are held together by weaker forces, weaker "bonds," whose characteristic strength is measured in $k_{\mathrm{B}}T_{\mathrm{room}}$ and not in eV. What are these types of materials?

1.3 Moduli

Here's a nice pointer to this, which involves yet another cheap-but-great physics scaling argument: the rigidity of materials is measured by elastic moduli, such as the shear modulus G, and this modulus has dimension *energy per volume*. In the absence of any other knowledge, the easiest guess we can make about the modulus of a material is therefore to take its characteristic energy scale and divide it by the cube of a characteristic length. For instance, for materials bound together by chemistry, that scale is the electron volt. For metals, it's the Fermi energy, which is again in the electron volt range. For ionic crystals, we again end up having electron volts. Moreover, all these materials have typical distances between atoms or molecules in the couple-of-Angstrom range. So we can estimate that their modulus is of the magnitude

$$G \sim \frac{\text{few eV}}{(\text{few Å})^3} \sim \frac{\text{eV}}{\text{few}^2(\text{Å})^3} \approx \frac{1.6 \times 10^{-19}\,\text{Nm}}{10 \times 10^{-30}\,\text{m}^3} \sim \text{tens of GPa} .$$

What stuff has moduli in that range? Stuff that you would ordinarily call solids. Many metals, hard plastic, glass, bone, concrete. Solids are congealed quantum mechanics.

What do we get for thermal scale interactions? Things that are held together by the strength of thermal energy, or which are "fluffed up" by the strength of thermal fluctuations, also happen to have slightly larger length scales: larger molecular constituents (polymers, colloids, ...) or at least intrinsic scales (say, a polymeric entanglement length). These scales are often in the few nanometer range, but can be significantly larger. What are typical moduli for stuff of that type?

$$G \sim \frac{\text{few}\, k_\text{B}T_\text{room}}{(\text{few nm})^3} \sim \frac{4.1 \times 10^{-21}\,\text{Nm}}{\text{few}^2(10^{-9}\,\text{m})^3} \approx \text{few hundreds of kPa} . \quad (1.1)$$

Matter of that type is quite squishy—soft rubbers, gels—and might easily start to flow if external forces are applied (foams, emulsions, bicontinuous lamellar phases), unless beyond some strain the deformation is arrested because stronger chemical bonds kick in (like in rubber, a vulcanized polymer network). Soft matter is congealed thermal energy.

Whether we need eV-scale energy to separate two particles or $k_\text{B}T$-scale energy makes a big practical difference for people who live at room temperature.[1] Without doing any chemistry (and, fingers crossed,

Heads up 1

"Few" $\sim \sqrt{\mathcal{O}(10)}$ is a sneaky number that lets you aim for a slightly higher precision in order-of-magnitude estimations. It follows useful rules such as

$$1 + \text{few} = \text{few}$$
$$\text{few} + \text{few} = \text{few}$$
$$\text{few} \times \text{few} = 10$$

[1] At around 12 000 K these two scales are the same, so beings at that scale don't have to worry about the distinction. But then, the existence of that scale separation is *hugely* important for life as we know it.

without really having to ponder too much quantum mechanics!) we can make or break things by exploiting the energy available in thermal fluctuations. "Making things" then often means "having things come together spontaneously," and this is what's typically called "*self assembly*." To be sure, people seem to have a predilection to restrict this word to things that do not just assemble into formless three-dimensional blobs or chunks. The energy between solvated ions in water is very much on the $k_{\mathrm{B}}T$ scale, thanks to water's huge dielectric constant, but the formation of ionic crystals out of a salty solution is typically not regarded as self assembly.

But wait, there's more: Given how small the thermal scale is compared to forces we can readily apply to materials without breaking a sweat, it is very easy to drive soft matter out of equilibrium. Hence, a lot of non-equilibrium thermodynamics most readily happens in the arena of soft matter physics.

1.4 Roadmap for this set of lectures

My goal is to pick one specific example from soft-matter self assembly and show you how to climb the hierarchy from the assembly process itself all the way to the emergent physics of the resulting aggregates—lipid membranes in this case—and how to reason about them. In the first lecture I will therefore discuss assembly of amphiphiles (such as surfactants or lipids), and how depending on their resulting geometry we do get aggregates whose "size statistics" is very different. In the second lecture I show how in the case of 2d-assemby we can develop a marvelous coarse-grained theory of the resulting aggregates and describe them not in terms of molecules and interaction forces but in terms of geometry and elasticity. In the third lecture I then walk you through some fun applications in the field of membrane-biophysics, especially some thoughts on the recent and quite interesting topic of membrane asymmetry.

1.5 Things to think about

1. I claimed that chemical reactions happen at the eV scale and are thus not bothered by temperature. Does that strike you as plausible? If so, why do you think chemistry labs have Bunsen burners? Let's think a bit harder.
   Consider a chemical reaction that requires crossing an energy bar-

rier E_{barrier}. To be specific, let's say $E_{\text{barrier}} = 1\,\text{eV} = 40\,k_\text{B}T_\text{room}$. Thermal energy is too small to cross it (in a sense, we're $39\,k_\text{B}T_\text{room}$ "short"), but there's a small chance that a pretty *enormous* thermal fluctuation kicks you over it, and we expect that chance to be proportional to the Boltzmann factor $\text{e}^{-E_\text{barrier}/k_\text{B}T}$. That's teeny-tiny, but if we have a mole of material, we get lots of opportunities to try. Now: what happens if we change the temperature from 300 K to 600 K? Instead of $39\,k_\text{B}T_\text{room}$ we're now "only" $38\,k_\text{B}T_\text{room}$ short—it doesn't seem we made much headway. But that's of course the entirely wrong way to think about it. Doing it properly, how much more likely is it that the same chemical reaction is now triggered by a thermal fluctuation?

(You also just learned something deep about cooking.)

2. The interaction energy between two ions in solution is given by $U(r) = q_1 q_2/4\pi\varepsilon_0\varepsilon_\text{r} r$, where q_1 and q_2 are the ionic charges, r is their separation, and ε_r is the dielectric constant.

   a) The so-called *Bjerrum length* ℓ_B is defined as the distance at which two monovalent ions in solution have an interaction energy equal to $k_\text{B}T$. Find a formula for it and calculate its value for water at room temperature!

   b) The dielectric constant of water actually depends a bit on temperature. A good approximation between 0 °C and 100 °C, due to Malmberg and Maryott (1956), is

   $$\varepsilon_\text{r}(T) \approx 10^{2.48151 - 0.001972(T/\text{K})} \approx \text{e}^{5.71388 - 0.004541(T/\text{K})}\,. \quad (1.2)$$

   Show that this implies that the strength of electrostatic interactions in water *increases* with temperature. Does that strike you as weird?

   c) That the dielectric constant ε_r depends on temperature is a stark reminder that the electrostatic interaction in water is really a *free* energy. As such, it has an *energetic* and an *entropic* "component:" $\psi(r) = \psi_E(r) - T\psi_S(r)$, if you will. Leaning on some standard thermodynamic lore, such as $F = E - TS$, or $S = -\partial F/\partial T$, calculate $\psi_E(r)/\psi(r)$ and show that for the specific temperature dependence from Eqn. (1.2) the energetic part of the interaction *has the "wrong" sign* (meaning, two like charges in water would attract, if interaction *energy* were the only thing that matters. (The fact that they still repel means that the entropic part, $\psi_S(r)$, must overcompensate this blunder. This puts a new spin on

the discovery that electrostatic interactions in water become stronger as the temperature increases.)

3. Later in these notes, you will (maybe...) derive in a homework problem (Chapter 3, Problem 6c, specifically: Eqn. (3.83)) an approximate formula that relates the bending rigidity of *one leaflet* of a lipid membrane, κ_{m}, with its thickness d and the elastic modulus Y of the material from which it is made:

$$\kappa_{\mathrm{m}} = \frac{Y d^3}{12} \ . \tag{1.3}$$

(In case you're wondering: the whole lipid *bilayer* has a rigidity that's just twice that: $\kappa = 2\kappa_{\mathrm{m}}$.)

a) If I tell you that $\kappa_{\mathrm{m}} \approx 15\, k_{\mathrm{B}}T$ and $d \approx 2\,\mathrm{nm}$, what value for the elastic modulus would you predict? Google! What type of matter does this suggest?

b) A membrane's bending rigidity κ is *not* of order $k_{\mathrm{B}}T$. And it's not of order of hundreds of $k_{\mathrm{B}}T$ either. Ponder why the few-tens-of-$k_{\mathrm{B}}T$ value might be biophysically attractive! *(Hint: throwing another energy scale into the mix: the energy released by the hydrolysis of a single ATP molecule under physiological conditions is around* $20\, k_B T$*, sometimes also a bit more.)*

c) What if modern nanotechnology buffs decided to build such membranes out of steel? We'd like to keep the bending rigidity at the value I told you, for the reasons you hopefully unearthed in the last part, but we replace it with a modern-era material that has a modulus of 200 GPa. What thickness would it have to have? Does this strike you as realizable? What does it say, maybe in retrospect, about the fact that all of us are squishy beings?

4. I said that soft matter systems are particularly easy to drive out of thermal equilibrium. Can you imagine arguments that would show how they would also be great examples where the concept of active matter arises?

2 Self assembly of amphiphiles

As promised, the topic of this chapter is to understand the thermal equilibrium physics of self assembly—specifically that of amphiphilic molecules such as surfactants or lipids. As it turns out, the size-statistics of the resulting aggregates *massively* depends on their geometry, and their geometry in turn is heavily influenced by the shape of an individual surfactant molecule. Now, what precisely we mean by "the shape of an individual surfactant molecule" is a very dicey question about which we could (but will not) talk a lot more. Suffice to say: it's subtle, easily misused, but *not entirely silly*. We take that as a sufficient reason to proceed. With caution. What makes this so interesting is that extremely minor molecular changes can, via the lever arm of the aggregation mechanism, make huge differences in the morphology and properties of the resulting aggregates. This is an incredibly lucrative game to play if you are a chemical engineer. (And a vital one, if you are nature.) So let us in a first step explore the question of aggregate shape, from which we will then transition in a second step to statistical properties of the aggregates.

Full disclosure: this chapter is a minimally revised excerpt from lecture notes I had prepared a few years back for a summer school in Udine, Italy (Terzi and Deserno, 2018), and it mostly channels the highlights from the groundbreaking paper by Israelachvili et al. (1976). Good discussions can also be found in textbooks on soft condensed matter physics, such as (Jones, 2002; Witten and Pincus, 2004).

2.1 From surfactant shape to aggregate geometry

Surfactant molecules are amphiphiles: they comprise different chemical moieties that are soluble in different solvents. Since they are linked together chemically, this requires nature to grapple with an interesting problem: how best to lower the free energy, given that no matter what the solvent conditions are, some chemical moieties will likely be "unhappy." Nature's solution to this is *self assembly*—a process by which larger scale structures form cooperatively, such that unfavorable sol-

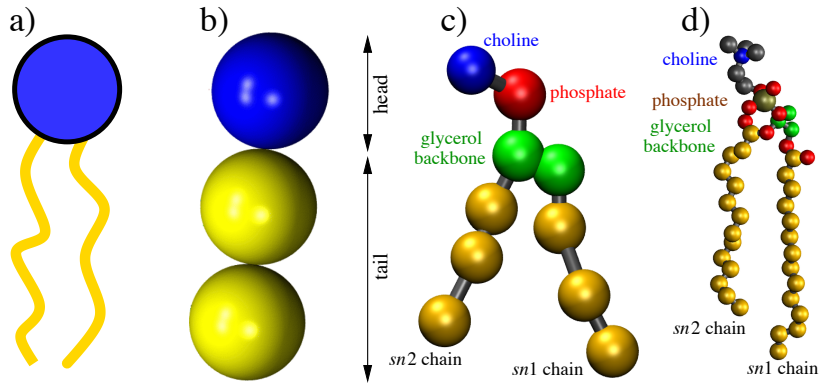


Figure 2.1 | Illustration of the morphology of a lipid molecule. Panel (a) shows a typical physicist's cartoon—a hydrophilic head group with two schematic tails; panel (b) takes this sketch serious and translates it into a highly coarse grained model (Cooke et al., 2005); panel (c) illustrates a lipid on the MARTINI level (Marrink et al., 2007), where the number of beads is increased, but still each bead accounts for approximately 3-4 heavy atoms; and panel (d) displays a united-atom lipid model of a very particular lipid, DMPC (dimyristoylphosphatidylcholine) (Berger et al., 1997), in which every atom (except non-polar hydrogens) are explicitly accounted for. Adapted from (Wang and Deserno, 2016).

vent contact is largely avoided. Self assembly is an amazing and hugely important example of an *emergent phenomenon*, in that it creates new physical entities (namely, the aggregates) which can be much bigger than the individual molecules they are made of.

2.2 Morphology

Amphiphiles are molecules which are typically divided into a "head" and a "tail." The head is hydrophilic (water soluble), for instance because it has polar groups (*e. g.*, hydroxyl or carbonyl groups), or because it is charged (*e. g.*, amino, carboxyl, or phosphate groups). The tail, on the other hand, is hydrophobic (water insoluble), and for lipids generally consists of two aliphatic chains. They typically contain between 12 and 22 carbon atoms, usually connected by single bonds, but sometimes with one or more double bonds (in the latter case one speaks of "unsaturated lipids"). Figure 2.1 gives a simple illustration of this, by showing pictures of lipids using some commonly employed computational models for studying them. Notice that only one of these models strives for a detailed chemical resolution. The others simplify the chemical architecture more or less drastically, but they all keep one key aspect: lipids are amphiphiles.

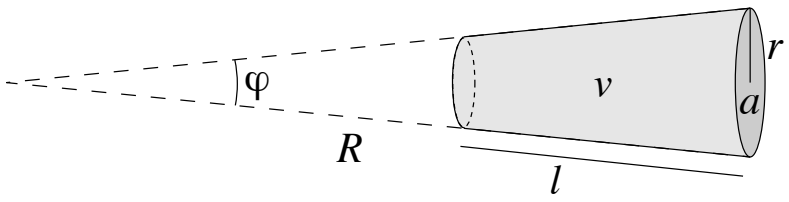


Figure 2.2 | Simplified shape-description of a surfactant as a blunted cone.

The key effect on which self assembly relies is a cooperative aggregation of surfactants that tries to bury the water-insoluble tails in the interior of the aggregate, shielding them from the aqueous solvent by a layer of hydrophilic head groups. Interestingly, there are numerous different morphologies in which that could happen, and this depends on the *shape* of the surfactant. For instance, if the lipid has a relatively large head group and a thin tail—if it looks like an ice cream cone—then we can imagine these surfactants packing together to form little spheres. But if the shape of a lipid is less obviously pointed, then lower curved structures seem more likely—such as cylindrical aggregates or even planar sheets. As we will now see, Israelachvili et al. (1976) have developed a beautifully simple way to make this intuition quantitative.

Let us represent a lipid schematically as a building block that is approximately cylindrical, but with a somewhat tapered tail region, as illustrated in Fig. 2.2, so that it looks like a blunted cone. The area of its head-group surface is $a = \pi r^2$, its volume is v, and its length is l.

2.2.1 Spheres

Imagine we need N of these object to piece them together into a sphere of radius R_{sph}. It is then obvious that we must have

$$Nv = V_{\text{sph}} = \frac{4}{3}\pi R_{\text{sph}}^3 \ , \tag{2.1a}$$

$$Na = A_{\text{sph}} = 4\pi R_{\text{sph}}^2 \ . \tag{2.1b}$$

Dividing these two equations, N cancels, and we get an equation for the radius of that sphere:

$$\frac{v}{a} = \frac{1}{3}R_{\text{sph}} \ . \tag{2.2}$$

At the center of the sphere we cannot have any empty space. Hence the radius R_{sph} which we found cannot be larger than the length l of the amphiphile—imagine for instance that there is a largest length to which

the tails can stretch, and that limits the sphere's radius: $R_{\text{sph}} \leq l$. This results in the condition

$$\text{spheres:} \qquad \frac{v}{al} =: P \leq \frac{1}{3} \ , \tag{2.3}$$

where we defined the important dimensionless *packing parameter* P. We hence find that if this condition on P is satisfied, these lipid building blocks will indeed like to aggregate into spherical objects, which go under the name *spherical micelles*.[1]

2.2.2 Cylinders

We can repeat this argument, but now instead consider packing the building blocks into a cylinder of radius R_{cyl} and length L_{cyl}; Assuming that L_{cyl} is large enough to ignore end effects, we then get

$$Nv = V_{\text{cyl}} = \pi R_{\text{cyl}}^2 L_{\text{cyl}} \ , \tag{2.4a}$$

$$Na = A_{\text{cyl}} = 2\pi R_{\text{cyl}} L_{\text{cyl}} \ . \tag{2.4b}$$

Again dividing these two equations cancels not just N but also L_{cyl} (as it should) and yields

$$\frac{v}{a} = \frac{1}{2} R_{\text{cyl}} \ . \tag{2.5}$$

Once more, requiring that the resulting value for the cylinder's radius is not larger than the lipid length l leads to the condition

$$\text{cylinders:} \qquad \overbrace{\frac{1}{3} \leq \underbrace{P \leq \frac{1}{2}}_{\text{from this section}}}^{\text{from last section}} \ , \tag{2.6}$$

where the lower cutoff comes from the previous case: if P is even smaller than $\frac{1}{3}$, we already know that we get spheres!

[1] The more skeptical among you might worry that if all those surfactant heads happily sit on the surface of a sphere and send their hydrophobic tails straight towards that sphere's center, it would get *awfully* crowded at the center! Absolutely true. It *would*. A perfectly radial arrangement of tails is inconsistent with a uniform liquid-like density. But then, not all tails really go to the center, for exactly that reason. Some have to bend back, possibly all the way to the surface of the sphere. If one wants to be quantitative about such packing issues, one needs to go beyond such naïve pictures as the one we're using here, and people have done so. See for instance Dill and Flory (1981).

2.2.3 Planes

Three time's the charm. So let's now pack the amphiphiles into a planar bilayer structure of area A_{bil} and thickness b_{bil}, leading to

$$Nv = V_{\text{bil}} = b_{\text{bil}} A_{\text{bil}} \ , \tag{2.7a}$$
$$Na = A_{\text{bil}} = 2A_{\text{bil}} \ . \tag{2.7b}$$

Dividing these two equations cancels both N and A_{bil} and gives

$$\frac{v}{a} = \frac{1}{2} b_{\text{bil}} \ . \tag{2.8}$$

Again, the thickness of each individual leaflet (*i. e.*, half the bilayer's thickness) cannot exceed the length l to which the lipid can stretch, $\frac{1}{2} b_{\text{bil}} \leq l$, and so we find

$$\text{bilayers:} \qquad \overbrace{\frac{1}{2}}^{\text{from last section}} \leq \underbrace{P \leq 1}_{\text{from this section}} \ . \tag{2.9}$$

The argument, as presented, is remarkably simple; Israelachvili et al. (1976) look at the situation in a fair bit more detail, but the key findings nevertheless hold up. In fact, this line of reasoning works well even for building blocks that are very simple and not very pliable–such as the lipid model from Fig. 2.1b. Cooke and Deserno (2006) showed that by simply changing the head-group size of the three-bead lipid, one can drive the entire morphological transition from spheres over cylinders to bilayers; if one pushes the packing parameter even larger, the lamellar phase becomes unstable. This is illustrated in Fig. 2.3.

Of course, the transitions *themselves* do not yet tell whether the simple packing-parameter theory works; but this theory makes a prediction that can be tested. Taking the area per lipid from a flat bilayer as the value for a, and using one of the transitions (say, spheres to cylinders) to pinpoint v/l, one can write the packing parameter as a function of the head group size of the lipid. This then gives a prediction for the head group size where the other transition (cylinders to bilayer) happens Cooke and Deserno (2006) show that this prediction indeed works.

Heads up 2

Oh dear. As a general piece of advice: if authors tell you that something "works" (or even "works well"), you should ask them *in what sense*.

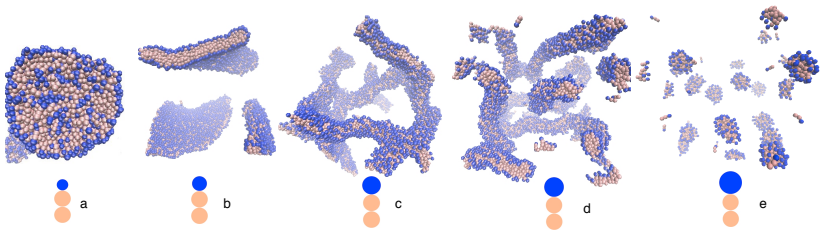


Figure 2.3 | The different morphologies of amphiphilic aggregates are controlled by amphiphile shape, even for models as simple as that from Fig. 2.1b. Reprinted from Cooke and Deserno (2006).

2.2.4 From packing parameter to spontaneous curvature

The geometrical picture we have in mind by now is that a smaller packing parameter P corresponds to a more cone-like shape, while for a larger P the lipid becomes more cylindrical. This intuition can be verified (and made more precise) by a simple calculation: if Ω is the solid angle of the blunted cone, then its volume can be written as

$$v = \frac{1}{3}\Omega\left[R^3 - (R-l)^3\right] = \Omega\left[R^2 l - Rl^2 + \frac{1}{3}l^3\right] \quad . \tag{2.10}$$

Since its head surface is $a = \Omega R^2$, we find $P = 1 - \frac{l}{R} + \frac{1}{3}\left(\frac{l}{R}\right)^2$, a quadratic equation that can be solved for R, from which we get the solid angle. Since, furthermore, $\Omega = 2\pi\left(1 - \cos\frac{\varphi}{2}\right) \approx \frac{1}{4}\pi\varphi^2$, where the last approximation is good for small φ, we arrive at the opening angle

$$\frac{\varphi}{r/l} \approx 3\left[1 - \sqrt{1 - \frac{4}{3}(1-P)}\right] \quad . \tag{2.11}$$

This relation is illustrated in Fig. 2.4. The characteristic ratio r/l defines an angle, and the actual opening angle φ is some multiple of that—twice as big for cones at the boundary between spheres and cylinders, and about 1.3 times as big at the boundary between cylinders and planes. Of course, the angle vanishes at $P = 1$. Notice that we can alternatively also calculate the lipid spontaneous curvature, defined as $J_{0\mathrm{m}} = 2/R$. For P close to 1 we find for this parameter

$$J_{0\mathrm{m}} l \approx 2(1-P) + \frac{2}{3}(1-P)^2 + \cdots \tag{2.12}$$

This provides a (fairly heuristic!) link between the parameter $J_{0\mathrm{m}}$ from continuum Helfrich theory, which we will meet in chapter 3 (see Eqn. (3.42)), and a parameter from the self assembly problem, P.

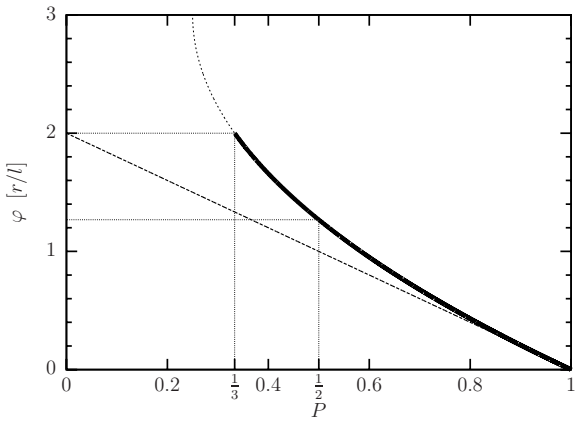


Figure 2.4 | Relation between the opening angle of the blunted cone from Fig. 2.2 (measured in units of r/l) and the packing parameter P. Around $P = 1$ we have $\varphi \approx \frac{2r}{l}(1 - P)$.

2.3 Statistical thermodynamics

Knowing the shape of the aggregate is only the beginning. We surely also want to know, under what conditions such aggregates form, and if they come in different sizes (say, what's the length of a cylindrical micelle?), we want to know what that is, too.

The problem is interesting, because entropy plays a key role. Were it only a matter of energy, any kind of amphiphile would aggregate to any other amphiphile, no matter how weak any attractive interaction is. But when we consider entropy, we realize that aggregation strongly reduces the translational entropy of amphiphiles. To understand this energy-entropy balance better we again follow Israelachvili et al. (1976). Let us therefore define

$$\varepsilon_n : \text{energy per molecule in } n\text{-aggregate} \tag{2.13a}$$

$$\phi_n : \text{concentration of } n\text{-aggregates} \tag{2.13b}$$

$$X_n : \text{concentration of monomers in } n\text{-aggregates}, \ = n\phi_n \ , \tag{2.13c}$$

where an "n-aggregate" or "n-mer" is a self-assembled aggregate of molecules consisting exactly of n molecules (or monomers or 1-aggregates). You may think of X_n in the following way: consider only the n-aggregates in solution (mentally remove all the others) and now ask, what is the overall concentration of all amphiphiles left in the system?

Heads up 3

The definitions here are really important. Make sure you understand them, especially the difference between ϕ_n and X_n.

The total energy of one n-aggregate is of course $E_n = n\varepsilon_n$. Observe that this does *not* imply that $E_n \propto n$, since ε_n also depends on n. The energy density due to n-aggregates is therefore

$$e_n = \phi_n E_n = \phi_n n \varepsilon_n = X_n \varepsilon_n \ . \tag{2.14}$$

For the (translational) entropy density of n-aggregates we will simply assume an ideal gas law, so that we get

$$s_n = -k_\mathrm{B}\, \phi_n \big(\log \phi_n - 1 \big) \ . \tag{2.15}$$

The total free energy density is then the sum of the energetic and entropic terms over all aggregate sizes:

$$f = \sum_{n=1}^{N} \Big\{ e_n - T s_n \Big\} \tag{2.16a}$$

$$= \sum_{n=1}^{N} \left\{ X_n \varepsilon_n + k_\mathrm{B} T \, \frac{X_n}{n} \left(\log \frac{X_n}{n} - 1 \right) \right\} \ , \tag{2.16b}$$

where N is the total number of molecules, and hence also the biggest aggregate we can get.

We are interested in the distribution function of aggregate sizes, X_n, *subject to the constraint* that the total amount of material in the system is fixed, meaning

$$\sum_{n=1}^{N} X_n =: X = \text{fixed} \ , \tag{2.17}$$

where X is the total monomer concentration in the system. We can calculate this by minimizing Eqn. (2.16b) subject to the constraint, which we enforce by means of a Lagrange multiplier μ:

$$0 \stackrel{!}{=} \frac{\partial}{\partial X_n} \left\{ f[X_n] + \mu \left[X - \sum_{m=1}^{N} X_m \right] \right\} \ . \tag{2.18}$$

This readily gives

$$\phi_n = \mathrm{e}^{-\beta n (\varepsilon_n - \mu)} \ , \tag{2.19}$$

where as usual $\beta = 1/k_\mathrm{B}T$. From this we in particular also get the monomer concentration ϕ_1, and so we can eliminate the Lagrange multiplier μ from the expression:

$$\phi_n = \left[\phi_1 \, \mathrm{e}^{\beta(\varepsilon_1 - \varepsilon_n)} \right]^n \ . \tag{2.20}$$

This is a very important general result. How it plays out in reality depends entirely on ε_n, which in turn depends crucially on the geometry of the aggregate—spherical cylindrical, or planar. Regardless: we see that if $\varepsilon_n < \varepsilon_1$, meaning that it is favorable for a monomer to be in an n-aggregate compared to being isolated in solution, the exponential factor becomes large and the concentration of n-aggregates goes up. But let us now specifically look at the individual geometries.

2.3.1 Spherical micelles

What is the energy of a monomer in a micelle consisting of n monomers? This is potentially a difficult question, but we will circumvent it by looking at the physics: packing monomers of some particular curvature into a spherical aggregate will likely result in some particular size—say, m—at which they fit best, and deviations away from that size will be suboptimal. Let us hence assume that, to lowest order, the energy is simply quadratic in the deviation from that particular optimal state:

$$\varepsilon_n = \varepsilon_m + \frac{1}{2}\varepsilon^*(n-m)^2 \ . \tag{2.21}$$

Inserting this into Eqn. (2.19) leads to

$$\phi_n = \exp\left\{-\beta n\left(\varepsilon_m + \frac{1}{2}\varepsilon^*(n-m)^2 - \mu\right)\right\} \ , \tag{2.22}$$

where μ needs to be determined from the normalization condition (2.17). Notice that this distribution is *cubic* in the exponent. However, we can simplify it by expanding the exponent around its maximum, up to quadratic order, and hence find an approximate Gaussian distribution that describes ϕ_n reasonably well. To do so, we need to calculate

$$0 \stackrel{!}{=} \frac{\partial}{\partial n}\left[-\beta n\left(\varepsilon_m + \frac{1}{2}\varepsilon^*(n-m)^2 - \mu\right)\right] \ , \tag{2.23}$$

which leads to the solution n^* at which the function peaks:

$$n^* = \frac{m}{3}\left[2 + \sqrt{1 - \frac{6(\varepsilon_m - \mu)}{\epsilon^* m^2}}\right] \approx m - \frac{\epsilon_m - \mu}{\epsilon^* m} \ , \tag{2.24}$$

where the approximation results from expanding the square root to first order, since the term $6(\varepsilon_m - \mu)/\epsilon^* m^2$ is small. We then find the quadratic expansion

$$\begin{aligned} n\left(\varepsilon_m + \frac{1}{2}\varepsilon^*(n-m)^2 - \mu\right) &\approx \text{const.} + \frac{1}{2}\epsilon^* m\sqrt{1 - \frac{6(\varepsilon_m - \mu)}{\epsilon^* m^2}}(n-n^*)^2 \\ &\approx \text{const.} + \frac{\epsilon^* m}{2}(n-n^*)^2 \ , \end{aligned} \tag{2.25}$$

where we again expanded the square root. This shows that the micelle distribution can be approximated as a Gaussian,

$$\phi_n \approx \text{const.} \times \exp\left\{-\frac{(n-n^*)^2}{2\sigma^2}\right\} \ , \tag{2.26}$$

with the mean value n^* given in Eqn. (2.24) and the variance given by

$$\sigma^2 = \frac{k_\mathrm{B}T}{\epsilon^* m} \ . \tag{2.27}$$

Observe that the distribution widens at larger temperature and is narrower for bigger micelles.

The effects on the structure on a single micelle are curious but minor in the spherical case; what is truly remarkable and very important is the overall aggregation thermodynamics which this model implies. In order to not get bogged down in tedious math (chiefly from dealing with the normalization condition (2.17)), let us instead look at a *two-state system*, in which we only have *monomers coexisting with m-aggregates*, and the normalization condition becomes $X = \phi_1 + m\phi_m$. Furthermore, we have

$$\phi_m \overset{(2.20)}{=} (\phi_1 \mathrm{e}^{\beta(\varepsilon_1 - \varepsilon_m)})^m = (\phi_1 \mathrm{e}^{\alpha})^m \ , \tag{2.28}$$

where we defined $\alpha = \beta(\varepsilon_1 - \varepsilon_m) > 0$ (we know the sign because we know that it is energetically favorable to form an m-aggregate). The normalization condition then becomes

$$X = \phi_1 + m\mathrm{e}^{\alpha m}\, \phi_1^m \ . \tag{2.29}$$

This must be solved for ϕ_1. Uh oh. It's an m^{th} order polynomial equation—how could we *possibly* do this in generality? True, this looks exceedingly troublesome, but it in fact becomes simple to get an approximate solution if we remember that m is likely large: recall from Sec. 2.2 and Fig. 2.2 that the number of surfactants in a spherical micelle can be written as $N = 4\pi l^2/a = 4\pi l^2/\pi r^2 = (2l/r)^2$, and with a reasonable estimate of $a \approx 0.5\,\mathrm{nm}^2$ (and hence $r \approx 0.4\,\mathrm{nm}$) and $\ell \approx 2\,\mathrm{nm}$, we find $N \approx 100$. We then see that the second term in Eqn. (2.29) stays extremely small for sufficiently small ϕ_1 and then very rapidly picks up and completely dominates the value of X—see the left hand graph in Fig. 2.5. The crossover happens where the two terms on the right hand side are approximately equal, leading to

$$\phi_1 = m\mathrm{e}^{\alpha m}\phi_1^m \quad \Longrightarrow \quad \phi_1 = \left(\frac{1}{m}\right)^{\frac{1}{m-1}} \mathrm{e}^{-\frac{\alpha m}{m-1}} \approx \mathrm{e}^{-\alpha} \ , \tag{2.30}$$

where the approximation is very good because $m \gg 1$ (recall in particular that $(1/m)^{1/m} \approx 1 - (\ln m)\, m^{-1} + \mathcal{O}(m^{-2})$).

We just discovered something hugely important: a critical concentration exists, $\phi_{\mathrm{cmc}} = \mathrm{e}^{-\alpha}$, at which an important change happens in the solution: up to that concentration, the normalization condition is

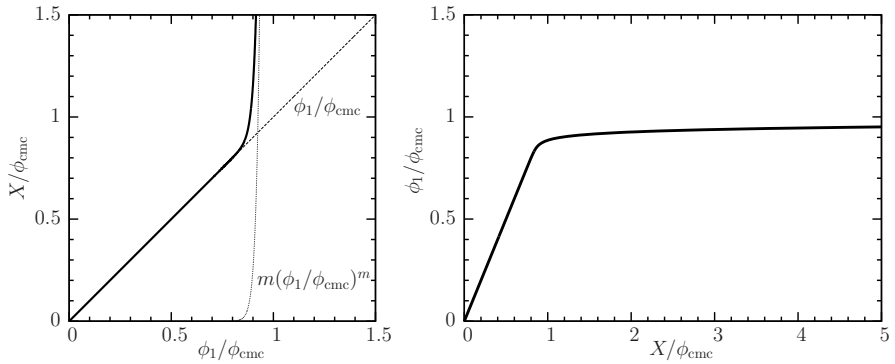


Figure 2.5 | The left plot shows the total surfactant concentration in all aggregates combined, X, as a function of the concentration of *single* monomers, ϕ_1. Since X emerges as a sum of ϕ_1 and a second term $m(\phi_1/\phi_{\rm cmc})^m$ with a large m (in the graph we chose $m = 50$), there is a sharp crossover near $\phi_1 = \phi_{\rm cmc} = {\rm e}^{-\alpha}$. The right picture simply flips the axes and shows the monomer concentration ϕ_1 as a function of the total surfactant concentration. Initially, the monomer concentration grows linearly with the amount of added amphiphiles—up to the concentration $\phi_{\rm cmc}$, at which point it essentially stays constant.

dominated by ϕ_1, and this means that the solution exists almost exclusively of monomers. But at $\phi_{\rm cmc}$ the second term takes over, and *from now on adding extra material will almost exclusively go into aggregates*. This is very visible if we plot the inverse of the normalization condition—see the right hand side of Fig. 2.5: the concentration of monomers initially grows linearly with the amount of added material, but it levels off quite abruptly at $\phi_{\rm cmc}$, meaning that from now on any additional material will form micelles, which so far did not exist. The concentration $\phi_{\rm cmc}$ is called the *critical micelle concentration*, usually abbreviated as "cmc", and it is a fundamentally important quantity for any aggregation problem. We will soon see that the concept remains relevant beyond the case of spherical micelles we have discussed just now. Notice that $\alpha = \beta(\varepsilon_1 - \varepsilon_m)$ is not just positive, but can be a fair amount bigger than 1, since the energy which an amphiphile gains in an aggregate compared to being in isolation can be many $k_{\rm B}T$. This implies, in turn, that the cmc can be very low: not much material needs to be added before micelles form. For instance, the cmc for the standard surfactant sodiumdodecylsulfate (SDS) is about 8 mM in water at 25°C, at which point the aggregation number of the micelles is $m \approx 60$ (Turro and Yekta, 1978).

To be clear: the micellization transition is *not* a phase transition in the classical sense: there is no discontinuity or non-analyticity in any of the thermodynamic functions; the transition is always rounded, since

m is large but finite. Regardless, it is a *very* pronounced change in the system's behavior, and as such it dominates aggregation physics.

2.3.2 Cylindrical micelles

The difference between the spherical and the cylindrical case enters via the energy per monomer in an aggregate, ε_n. For spheres we made the reasonable assumption in Eqn. (2.21) that there is a typical size for a micelle, and that the energy will deviate quadratically as we move away from that value. This cannot be true for cylindrical micelles, though, since they have an unspecified length: we can easily make cylindrical micelles longer by simply adding more amphiphiles to the linear part. The aggregation energy of these amphiphiles will be always the same, for they cannot know how long the cylindrical aggregate is of which they are a part. However, amphiphiles at the two end caps of the micelle must have a different energy, and it must be *larger* than the energy of amphiphiles in the wormlike middle, for if that were not so, spherical micelles would form in the first place. It is hence reasonable to write the total energy of a cylindrical micelle of n monomers as $E_n = n\varepsilon_\infty + 2E_{\text{cap}}$, and hence the energy per monomer is

$$\varepsilon_n = \varepsilon_\infty + \frac{2E_{\text{cap}}}{n} =: \varepsilon_\infty + \frac{\alpha\, k_\text{B}T}{n} \ . \tag{2.31}$$

Notice that the dimensionless number α must be large: it is the excess energy (in units of $k_\text{B}T$) of all end-cap monomers. Since these caps consist of two semi-spheres, they together make up essentially one full *spherical* micelle, whose aggregation number is $\mathcal{O}(100)$, and it seems fair to estimate that the excess energy for each monomer stuck in the wrong local geometry is at least a sizable fraction of $k_\text{B}T$.

Inserting this ansatz for ε_n into Eqn. (2.20), we get

$$\phi_n = \left[\phi_1\, \mathrm{e}^{\beta(\varepsilon_1 - \varepsilon_\infty - \alpha\, k_\text{B}T/n)}\right]^n = \left[\phi_1\, \mathrm{e}^{\beta(\varepsilon_1-\varepsilon_\infty)}\right]^n \mathrm{e}^{-\alpha} = \left[\phi_1\, \mathrm{e}^{\alpha}\right]^n \mathrm{e}^{-\alpha} \ . \tag{2.32}$$

The last step follows since this equation must also be true for $n = 1$.

It is now highly useful to define the scaled concentrations $\tilde{\phi}_n = \phi_n \mathrm{e}^{\alpha}$, because in these variables Eqn. (2.32) becomes

$$\tilde{\phi}_n = \tilde{\phi}_1^n \ . \tag{2.33}$$

The distribution of the $\tilde{\phi}_n$ is *exponential*, which is remarkably wide (we will make this more precise below) and thus *very* different from the spherical case, where the distribution was sharply peaked around an optimal size. Notice that in order for it to be normalizable, we

must have $\tilde{\phi}_1 < 1$, implying that the monomer concentration can never exceed $\mathrm{e}^{-\alpha}$—a concentration we will soon recognize as the cmc for the cylindrical case.

If we define the scaled total concentration of monomers as $\tilde{X} = X\mathrm{e}^{\alpha}$, the normalization condition (2.17) becomes

$$\tilde{X} = \sum_{n=1}^{N} n\,\tilde{\phi}_n = \sum_{n=1}^{N} n\,\tilde{\phi}_1^n \stackrel{*}{=} \frac{\tilde{\phi}_1}{(1-\tilde{\phi}_1)^2} \;, \tag{2.34}$$

where in the last step, at $*$, we used an identity whose proof you will walk through in problem 1. We also took the limit $N \to \infty$.

This relation between $\tilde{X}$ and $\tilde{\phi}$ is a quadratic in $\tilde{\phi}$, which we can therefore easily solve for $\tilde{\phi}(\tilde{X})$:

$$\tilde{\phi}_{1\pm} = \frac{1 + 2\tilde{X} \pm \sqrt{1+4\tilde{X}}}{2\tilde{X}} \;. \tag{2.35}$$

Since we know $\tilde{\phi}_1 < 1$, the minus sign is the correct choice. Expanding the solution for small and large $\tilde{X}$, we find

$$\tilde{\phi}_1 = \begin{cases} \tilde{X} + \mathcal{O}(1) & \text{for } \tilde{X} \ll 1 \\ 1 - 1/\sqrt{\tilde{X}} + \mathcal{O}(\tilde{X}^{-1}) & \text{for } \tilde{X} \gg 1 \end{cases} \;. \tag{2.36}$$

As promised, we can again define a cmc, $\phi_{\mathrm{cmc}} = \mathrm{e}^{-\alpha}$, such that below the cmc the monomer concentration in our solution is proportional to the amount of added material, while for concentrations larger than the cmc any added material goes into micelles, leaving the monomer concentration below ϕ_{cmc}, and approaching it with a very slow $1/\sqrt{X}$ asymptotics. This is illustrated in Fig. 2.6.

We already know that the distribution of micelle sizes is exponential, but we might also want to know what the mean and the variance are. These are easily calculated by working out (weight-averaged) moments of n. For the first one, we find

$$\langle n \rangle = \frac{\sum_{n=1}^{\infty} n\tilde{X}_n}{\sum_{n=1}^{\infty} \tilde{X}_n} = \frac{\sum_{n=1}^{\infty} n^2\tilde{\phi}_n}{\sum_{n=1}^{\infty} n\tilde{\phi}_n} \stackrel{*}{=} \frac{1+\tilde{\phi}_1}{1-\tilde{\phi}_1} \stackrel{\#}{=} \sqrt{1+4\tilde{X}} \;, \tag{2.37}$$

where at $*$ we again used the fact that the sums can be done exactly, as you should verify in problem 1, simplified à la $N \to \infty$, and at $\#$ we inserted the solution (2.35). The average micelle length therefore grows like the *square root* of the concentration: $\langle n \rangle \approx 2\sqrt{X/\phi_{\mathrm{cmc}}}$.

The second moment of n is given by

$$\langle n^2 \rangle = \frac{\sum_{n=1}^{\infty} n^2\tilde{X}_n}{\sum_{n=1}^{\infty} \tilde{X}_n} = \frac{\sum_{n=1}^{\infty} n^3\tilde{\phi}_n}{\sum_{n=1}^{\infty} n\tilde{\phi}_n} \stackrel{*}{=} \frac{1+\tilde{\phi}_1(4+\tilde{\phi}_1)}{(1-\tilde{\phi}_1)^2} \;, \tag{2.38}$$

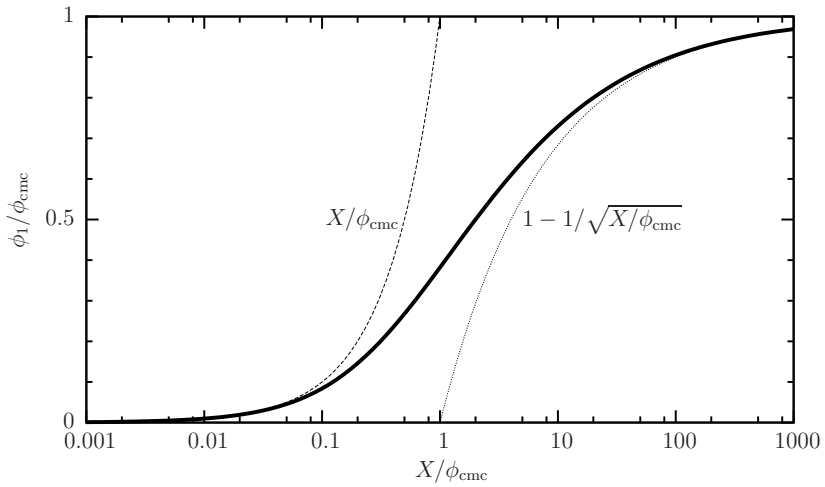


Figure 2.6 | Monomer concentration for the case of a cylindrical micelle aggregation scenario. The dashed and dotted curves indicate the small- and large-concentration limits from Eqn. (2.36). The full solution shows a cross-over at the cmc.

where at $*$ we used for a third time some knowledge about these types of sums, to be proved in problem 1, and again took the limit $N \to \infty$. The variance of n hence turns out to be

$$\sigma_n^2 = \langle n^2 \rangle - \langle n \rangle^2 = \frac{2\tilde{\phi}_1}{(1-\tilde{\phi}_1)^2} \stackrel{\#}{=} 2\tilde{X} \ , \tag{2.39}$$

where at $\#$ we again used the solution (2.35). This answer is important, because it shows that the width of the distribution essentially scales with its mean, and hence

$$\frac{\sigma_n}{\langle n \rangle} = \sqrt{\frac{2\tilde{X}}{1+4\tilde{X}}} = \frac{1}{\sqrt{2}} - \mathcal{O}(\tilde{X}^{-1}) \ . \tag{2.40}$$

Distributions of cylindrical micelles are hence "wide" no matter how large the micelles are; there is no "law of large micelles," or a $1/\sqrt{n}$ like asymptotics towards a sharp mean. As remarkable as this is, it is of course not unexpected, for that is what exponential distributions do.

2.3.3 Planar bilayers

Again, the first question to address is: what is ε_n for an aggregate that assembles in a planar fashion? To make headway, though, we need to make further assumptions about its geometry. We will assume that it stays flat, and that it will be circular. The latter follows because the

amphiphiles at the bilayer disc's edge will have a higher free energy per molecule than the one in the flat region (for reasons analogous to the elevated free energy of monomers at the ends of cylindrical micelles). This excess free energy per unit length acts as a *line tension* (in this case usually called *edge tension*), and minimizing it at constant overall area of the aggregate means that the shape has to be a circle.

If the circular aggregate has area $A = \pi R^2$, its circumference is $C = 2\pi R = 2\sqrt{\pi A}$. The excess free energy of the edge is $E_{\text{edge}} = 2\pi R\gamma = 2\sqrt{\pi A}\gamma$, with γ being the edge tension—a material parameter. Since the number of lipids in the aggregate is approximately $n = 2A/a_\ell$, with a_ℓ being the area per lipid, we get $A = \frac{1}{2}na_\ell$, and hence $E_{\text{edge}} = \sqrt{2\pi n a_\ell}\gamma$. The replacement for Eqn. (2.31) is hence

$$\varepsilon_n = \varepsilon_\infty + \frac{E_{\text{edge}}}{n} = \varepsilon_\infty + \frac{\sqrt{2\pi n a_\ell}\gamma}{n} = \varepsilon_\infty + \frac{\alpha\, k_\text{B}T}{\sqrt{n}}\,, \tag{2.41}$$

where $\alpha = \sqrt{2\pi a_\ell}\beta\gamma$ is a dimensionless number that's again a fair bit larger than 1. To estimate it, let's take the DOPC values of $a_\ell \simeq 0.7\,\text{nm}^2$ (Kučerka et al., 2006) and $\gamma \simeq 20\,\text{pN}$ (Portet and Dimova, 2010), from which we get $\alpha \approx 10$. Notice that the only difference between the cylindrical and the planar case is that in the latter the excess term is proportional to $1/\sqrt{n}$ instead of $1/n$. We will see that this changes the physics in a big way.

Inserting this expression for the energy per monomer into the general form of the aggregate distribution, Eqn. (2.20), we get

$$\begin{aligned} X_n = n\,\phi_n &= n\left[\phi_1\,\mathrm{e}^{\beta(\varepsilon_1-\varepsilon_\infty-\alpha\,k_\text{B}T/\sqrt{n})}\right]^n \\ &= n\left[\phi_1\,\mathrm{e}^{\beta(\varepsilon_1-\varepsilon_\infty)}\right]^n \mathrm{e}^{-\alpha\sqrt{n}} \\ &= n\left[\phi_1\,\mathrm{e}^{\alpha}\right]^n \mathrm{e}^{-\alpha\sqrt{n}}\,, \end{aligned} \tag{2.42}$$

where the last step again follows because this equation must also be true for $n = 1$. The normalization condition (2.17) then becomes

$$X = \sum_{n=1}^{N} X_n = \sum_{n=1}^{N} n\left[\phi_1\,\mathrm{e}^{\alpha}\right]^n \mathrm{e}^{-\alpha\sqrt{n}}\,. \tag{2.43}$$

The term $\mathrm{e}^{-\alpha\sqrt{n}}$ decreases with n, while for the term $[\phi_1\,\mathrm{e}^{\alpha}]^n$ the asymptotic behavior depends on whether $\phi_1\mathrm{e}^{\alpha}$ is bigger or smaller than 1. Assume it is bigger than 1. Then this term grows with n, and it asymptotically grows *faster* than $\mathrm{e}^{-\alpha\sqrt{n}}$ decreases. This might get us worried, for if we again replace $N \to \infty$ (because N will be macroscopically big),

the sum in Eqn. (2.43) would diverge. So let us assume that, instead, the expression $\phi_1 \mathrm{e}^{\alpha}$ is smaller than 1. In that case, we can calculate

$$X = \sum_{n=1}^{\infty} n \left[\phi_1 \, \mathrm{e}^{\alpha}\right]^n \mathrm{e}^{-\alpha\sqrt{n}} \leq \sum_{n=1}^{\infty} n \, \mathrm{e}^{-\alpha\sqrt{n}} \approx \int_0^{\infty} \mathrm{d}n \; n \, \mathrm{e}^{-\alpha\sqrt{n}} = \frac{12}{\alpha^4} \; . \tag{2.44}$$

This is a pretty disastrous finding, though: apparently, the total amount of material we can add to the system is bounded from above. What if we wanted to add more material—who's stopping us? (Not excluded volume—that wasn't part of the model!)

The solution to this conundrum is subtle: the assumption that N can be replaced by infinity is wrong—despite the fact that N could really be an Avogadro number of molecules. But large is not the same as infinite, and the normalization condition only enforces $\phi_1 \mathrm{e}^{\alpha} \leq 1$ if we really sum all the way up to infinity. If the sum is *finite*, there is no reason to demand that $\phi_1 \mathrm{e}^{\alpha} \leq 1$, because finite sums cannot diverge! More specifically, even if this term would ultimately outcompete $\mathrm{e}^{-\alpha\sqrt{n}}$, if $\phi_1 \mathrm{e}^{\alpha}$ is only ever so slightly bigger than 1, this will only happen near the upper bound of the sum—showing us that the value of this sum will likely depend very critically on just how much $\phi_1 \mathrm{e}^{\alpha}$ exceeds 1.

Unfortunately, it is quite tricky to see how this plays out analytically, because the normalization sum (2.43) turns out to be an *extremely* delicate interplay between very small and very large terms. To brace ourselves for what is actually happening here, we shall first look at a numerical example. Let us assume that $\alpha = 10$, that we have $N = 10,000$ molecules in the system (really an incredibly small number by experimental standards, but this might be a typical number to be used in a simulation), and let us demand that we want to ultimately gain a total concentration of $X = 10^{-2}$ (notice that this is larger than the erroneous upper bound of $12/\alpha^4 = 1.2 \times 10^{-3}$). If we abbreviate $\tilde{\phi}_1 = \phi_1 \mathrm{e}^{\alpha}$, then we have to numerically solve the following equation for $\tilde{\phi}_1$:

$$10^{-2} = \sum_{n=1}^{10,000} n \, \tilde{\phi}_1^n \, \mathrm{e}^{-10\sqrt{n}} \; . \tag{2.45}$$

Fig. 2.7 plots the right hand side of this equation as a function of the parameter $\tilde{\phi}_1$ in the interesting range.[2] Up to $\tilde{\phi}_1 \approx 1.1025$, the right hand side grows linearly with $\tilde{\phi}_1$, but at around this point a big change happens, and the sum picks up extremely rapidly—becoming a power

[2] This sum is a tricky mess of precision. I used MATHEMATICA®, defining it as `S[x_]:=Sum[n*x^n*Exp[-10*Sqrt[n]],{n, 1, 10000}]`. To illustrate how finicky this is: if you replace `10` by `10.0`, you get completely wrong answers.

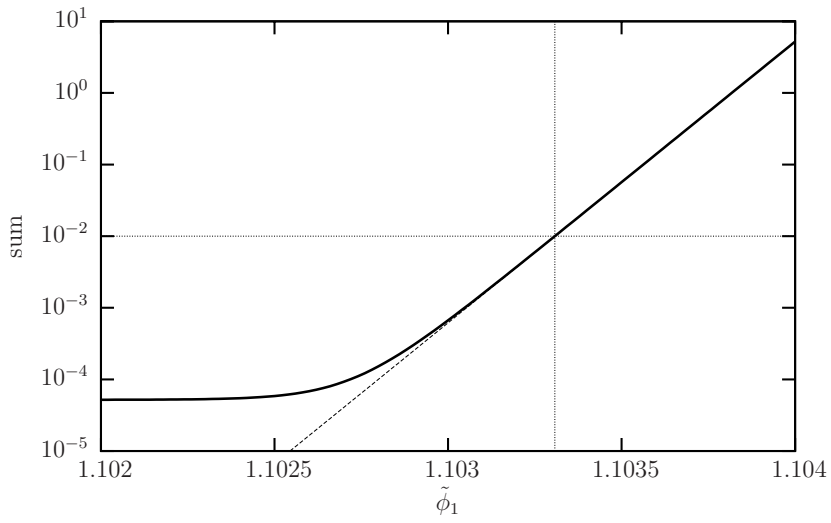


Figure 2.7 | The solid curve is the right hand side of Eqn. (2.45) as a function of $\tilde{\phi}_1$, for $\alpha = 10$ and $N = 10,000$; the dashed curve is the large-n-approximation from Eqn. (2.47). Observe the tiny $\tilde{\phi}_1$-range on the horizontal axis!

law with an exponent of about $10,000$. (This also shows why it is very hard to treat this problem numerically with even bigger values of N.) The value 10^{-2} is reached at $\tilde{\phi}_1 \approx 1.10330764$ and hence $X_1 \approx 5.009 \times 10^{-5}$.

Inserting this value for $\tilde{\phi}_1$ into the distribution function for X_n from Eqn. (2.42), we can plot it over the entire range of permissible n values: from $n = 1$ to $n = 10,000$; this is done in Fig. 2.8. Initially, the distribution function drops precipitously: one finds $X_2 \approx 1.756 \times 10^{-6} \approx X_1/30$ and $X_3 \approx 1.211 \times 10^{-7} \approx X_1/400$. But at $n = 2566$ the function attains a minimum, after which it again begins to rapidly grow. At its largest n-value it becomes $X_{10,000} \approx 4.696 \times 10^{-4} \approx 10X_1$, showing it is about 10 times more likely to find a lipid in that aggregate than to find it in isolation! Another way of looking at this is the following: 99% of all monomers are found in aggregates with a size of at least $9,890$. And yet another illustration is the following: Look at the cumulative normalized distribution of X_n, namely, $f(m) = X^{-1} \sum_{n=1}^{m} X_n$. It rapidly rises from 0 to about 0.0052 when m rises from 1 to 10. However, after that it stays virtually constant, until about $9,800$, when it begins to rise again. In other words, with the exception of about half a percent of small oligomers, *virtually the whole system forms one giant aggregate.*

I wish to strongly emphasize what this means: *surfactants that prefer to assemble two-dimensionally are prone to form macroscopic aggregates.* Macroscopic! Their size is no longer limited by their shape

Heads up 4

This is a really important big-picture lesson. It's the reason why Chapter 3 even makes sense!

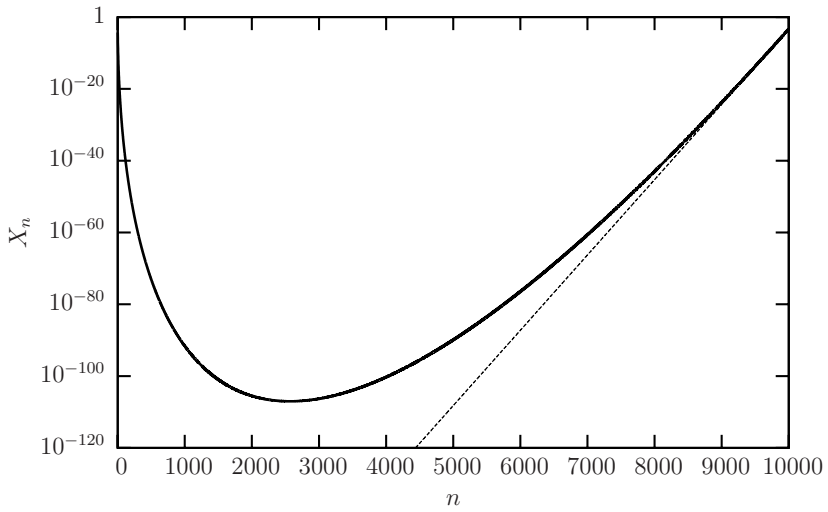


Figure 2.8 | The solid curve is the distribution function $X_n = n\,\phi_n$ from Eqn. (2.42), using the numerical parameters $\alpha = 10$, $N = 10,000$, and $X = 0.01$, which implies the numerical solution $\tilde{\phi}_1 \approx 1.10330764$ and hence $X_1 \approx 5.009 \times 10^{-5}$. The dashed curve is the approximate distribution from Eqn. (2.46), using the value for $\tilde{\phi}_1$ determined via the first-order approximation in Eqn. (2.50), $\tilde{\phi}_1^{(1)} \approx 1.10330882$. Using $\tilde{\phi}_1^{(1)}$ in the full distribution (instead of the exact $\tilde{\phi}_1$) leads to a curve that is indistinguishable from the exact one on this plot, with a normalization that is about 1% off.

(unlike spherical micelles) or the vagaries of a strongly fluctuating size distribution (unlike cylindrical micelles). Instead, they form large entities that could easily consists of thousands, millions, or hundreds of millions of individual molecules. They form *emergent* and *persistent* entities that can be described as new objects that deserve an effective description on their own, one that may not need to refer to their underlying microscopic reality as two-dimensional oceans of amphiphiles. It is this amazing "quirk" of two-dimensional self-assembly that gives lipid membranes their identity as "things to be reckoned with."

What follows from here on is more or less just a clean-up of the math. It goes beyond what you can find in (Israelachvili et al., 1976), but it is not strictly necessary. It is cute, though, and I encourage everyone interested in some fun math to go through it. You do not day-in-day-out encounter situations in which an Avogadro number of objects cannot rightfully be considered effectively infinite, and it is quite nifty how this plays out in the actual math. So, for those curious to be entertained with some really fun distribution functions—here we go!

With these observations we are now in a better position to develop a decent approximate solution for the normalization condition (2.43). No-

tice that we need to analytically describe the region in that sum which strongly increases (the "uptake" in Fig. 2.7), and that this comes from the aggregates—meaning, the large-n part of the distribution function. Hence it is probably a good idea to expand the summands in Eqn. (2.43) around the upper end, $n = N$, and preferably in such a fashion that we can perform the sum. But given the exponential variation of X_n, it is wise to do that expansion in the exponent:

$$\begin{aligned} X_n = n\,\tilde{\phi}_1^n \mathrm{e}^{-\alpha\sqrt{n}} &= \tilde{\phi}_1^n \exp\left\{-\alpha\sqrt{n} + \ln n\right\} \\ &= \tilde{\phi}_1^n \exp\left\{ -\alpha\left[\sqrt{N} + \frac{1}{2\sqrt{N}}(n-N)\right]\right. \\ &\qquad \left. + \ln N + \frac{1}{N}(n-N) + \mathcal{O}\Big((n-N)^2\Big)\right\} \\ &\approx N\mathrm{e}^{-\alpha\sqrt{N}/2}\left(\tilde{\phi}_1\,\mathrm{e}^{-\alpha/2\sqrt{N}}\right)^n \ . \end{aligned} \tag{2.46}$$

This expansion permits us to do the sum, since it turns into a simple geometric series:

$$\sum_{n=1}^{N} n\,\tilde{\phi}_1^n \mathrm{e}^{-\alpha\sqrt{n}} \approx N\mathrm{e}^{-\alpha\sqrt{N}/2}\frac{y^{N+1}-1}{y-1} \qquad \text{with} \quad y = \tilde{\phi}_1\,\mathrm{e}^{-\alpha/2\sqrt{N}} \ . \tag{2.47}$$

Since y is slightly larger than 1, but N is huge, y^{N+1} will be very large compared to 1. (In our above numerical example we would find $y \approx 1.0494987$ and hence $y^{N+1} \approx 1.0494987^{10,001} \approx 6.92 \times 10^{209}$.) We can hence neglect the "-1" in the numerator, but of course not in the denominator.

The normalization condition now becomes

$$\Xi := \frac{X\,\mathrm{e}^{\alpha\sqrt{N}/2}}{N} = \frac{y^{N+1}}{y-1} \ , \tag{2.48}$$

but this is again impossible to solve analytically. However, we can get increasingly good approximations by iteration. First, recall that the right hand side really emerged as a geometric series, and so it is given by $y^N + y^{N-1} + y^{N-2} + \cdots$. Let us take the dominant term, y^N, and solve the equation. We then get

$$y = \Xi^{\frac{1}{N}} \ . \tag{2.49}$$

Even though only approximate, this already looks remarkably good, since it gives $\tilde{\phi}_1 = 1.103645$ for our numerical example, about 0.03% off. And yet, inserting this value into the normalization condition gives

a value about 20 times too big. We need to do better. In fact, we can improve the solution by iterating the defining equation, à la $y^{(i+1)} = [\Xi(y^{(i)} - 1)]^{1/(N+1)}$, where $y^{(0)} = \Xi^{1/N}$ is our initial simple result. At first order we get

$$\begin{aligned}\tilde{\phi}_1^{(1)} &= \mathrm{e}^{\alpha/2\sqrt{N}}\, y^{(1)} \\ &= \mathrm{e}^{\alpha/2\sqrt{N}}\, \left[\Xi\left(\Xi^{1/N} - 1\right)\right]^{1/(N+1)} \\ &= \mathrm{e}^{\alpha/2\sqrt{N}} \left[\frac{X}{N}\, \mathrm{e}^{\alpha\sqrt{N}/2} \left(\left(\frac{X}{N}\right)^{1/N} \mathrm{e}^{\alpha/2\sqrt{N}} - 1\right)\right]^{1/(N+1)} \quad . \end{aligned} \tag{2.50}$$

With the numerical example from above ($X = 0.01$, $\alpha = 10$, and $N = 10,000$), this gives $\tilde{\phi}_1^{(1)} = 1.10330882$, which differs from the exact numerical solution only by 1 part in 10^6, and now the normalization condition is only 1% off. Unfortunately, further iterations do not gain us much anymore, because we are still solving an approximate equation, not the exact one.

There is more to be learned. Even the simplest solution becomes exact in the thermodynamic limit $N \to \infty$. Performing it, we get

$$\phi_1 = \mathrm{e}^{-\alpha} \lim_{N\to\infty} \left\{ \mathrm{e}^{\alpha/2\sqrt{N}} \left(\frac{X\, \mathrm{e}^{\alpha\sqrt{N}/2}}{N} \right)^{1/N} \right\} = \mathrm{e}^{-\alpha} \ , \tag{2.51}$$

showing that—again—we have a critical "micelle" concentration. Since bilayer patches are usually not viewed as "micelles," this is more commonly called the *critical aggregate concentration* and abbreviated as "cac": $\phi_{\mathrm{cac}} = \mathrm{e}^{-\alpha}$.

The scenario looks superficially similar to what we have seen in the spherical case: the normalization condition becomes a polynomial with a constant term, a linear term, and one term with a large power (compare Eqn. (2.29) and Eqn. (2.48)), and the "largeness" of that power makes the transition. However, in the spherical micelle case that power was given by the micelle size, and hence it was *mesoscopic*—of order 10^2. In the bilayer case that power is *macroscopic*—the total number of molecules in the system, conceivably of the order of Avogadro's number, but more importantly: *extensive*. It will *by definition* diverge in the thermodynamic limit. It hence follows that *the aggregation transition for bilayers is a true phase transition*—at least in the model we have studied here.

Alas, our model is defective. Or at least *incomplete*. The $1/\sqrt{n}$ correction to ε_n (see Eqn. (2.41)), on which the whole scenario hinges, comes from the $\sqrt{n}$ divergence of the edge energy for increasingly large

flat circular aggregates. But bilayer patches do not have to stay flat. Once they exceed a critical size, it is preferable for them to close up, make an edgeless spherical vesicle, and pay bending energy instead, because bending energy does not scale with size. This was first discussed by Helfrich (1974a). Hence, vesiculation caps the edge energy, moving the correction term back to a $1/n$ form, for which we expect a wide exponential distribution function like in the case of cylindrical micelles. Unfortunately, in reality things are now a lot more complicated, because we can no longer ignore kinetics. In any case, we still encounter an aggregation transition once the amphiphile concentration in solution exceeds a critical aggregate concentration.

2.4 Things to think about

1. While deriving the size distributions for cylindrical micelles, we encountered some nontrivial sums, which I claimed have closed forms. Let's derive them. Our goal is to evaluate expressions of the type

$$S_n(x) = \sum_{n=1}^{\infty} k^n x^k \qquad \text{with } |x| < 1 \text{ and } n \in \mathbb{N}_0 \,. \tag{2.52}$$

   The way to solve this is, curiously, by exploiting a tiny bit of calculus and some slick notation.

   a) Remind yourself that you (hopefully) already know $S_0(x)$.

   b) Define the differential operator $\hat{D} := x\frac{\mathrm{d}}{\mathrm{d}x}$. Now observe that x^k is an *eigenfunction* of that operator with *eigenvalue* k. Use this to rewrite the sum $S_1(x)$ using $S_0(x)$ and the differential operator $\hat{D}$.

   c) Realize that this trick can be generalized into a formal expression for $S_n(x)$ that involves S_0 and powers of $\hat{D}$.

   d) Explicitly evaluate $S_n(x)$ for $n \in \{1, 2, 3\}$, which we need in Eqns. (2.34), (2.37), and (2.38), respectively.

2. Chemists might describe linear aggregation of monomers into chains as follows. We have monomers, which we denote as "X_1" and n-mers, which we denote by "X_n". And we assume that there are chemical equilibria that describe the addition of a monomer to an n-mer to create an $(n+1)$-mer, which look like this:

$$\mathrm{X}_n + \mathrm{X}_1 \overset{K}{\rightleftharpoons} \mathrm{X}_{n+1} \,. \tag{2.53}$$

The key to a simple analytical answer is assuming that the chemical equilibrium constant K is *the same* for all these reactions. Using the law of mass action, show that this leads *exactly* to the same type of condition (2.33) that gave us all the results we found for the size distribution functions of cylindrical micelles.

3. The mystery around the paradoxical result in connection with Eqn. (2.44), and its resolution, bear some striking resemblance to a similar conundrum (and its solution) in quantum mechanics, that typically arises when one first encounters Bose-Einstein condensation. Those of us who have learned about that material: ponder these two cases and narrate the similarity and differences!

3 Geometry and Elasticity of lipid membranes

The previous chapter has prepared the groundwork for a very important insight: surfactants that prefer to assemble into two-dimensional aggregates form macroscopic persistent entities. Such surfactants are typically called "lipids", and the aggregates are called "lipid bilayers" or "lipid membranes." An *enormous* amount of science has been done about this subject, and it seems hopeless to even scratch the surface. And yet, that's why we're here, and so I'll try to give it my best shot.

3.1 A super quick overview of some membrane properties

The critical micelle (well: *aggregate*) concentration (cmc/cac) of lipids is *very* low: the lipid DPPC has a cmc of $4.6(5) \times 10^{-10}\,\mathrm{M}$ in water (Smith and Tanford, 1972); in fact values in the $10^{-10}\,\mathrm{M}$ range seem typical for most common two-tailed lipids. Let's make sure we understand what that means: water has a molarity of about 55 M, so in an aqueous solution in which a solute is present at a concentration of $10^{-10}\,\mathrm{M}$, water molecules outnumber the solute by about half a trillion to one. Or stated differently, the mean linear separation between two solute molecules is about 2.5 µm (cubically thinking...), which is about 8000 water molecules apart (again, linear distance, cubic lattice thinking). There are virtually no free lipids. They are basically all part of membranes.

Since individual lipids in a membrane are not chemically bonded (recall: we've got binding at $k_\mathrm{B}T$ scale, not eV scale!), they can laterally pass one another. Of course, the same is true for water molecules, but we know that this doesn't guarantee fluidity—solid ice really does exist. In other words: it is conceivable that the lipids in a membrane, even though not chemically bonded, are still locally stuck because the membrane is "frozen" into some sort of "solid" phase. However, it turns out melting and freezing in two dimensions is a rather tricky thing, and rigorously solid phases don't actually exist. All the same, there *is*

A few common lipids:

DPPC:
1,2-dipalmitoyl-*sn*-glycero-3-phosphatidylcholine

DOPC:
1,2-dioleoyl-*sn*-glycero-3-phosphatidylcholine

DMPC:
1,2-dimyristoyl-*sn*-glycero-3-phosphatidylcholine

POPC:
1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphatidylcholine

POPE:
1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphatidylethanolamine

POPS:
1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphatidylserine

a phase transition in membranes where lipids go from the commonly studied fluid phase into a far more ordered "gel phase" in which lipid tails wiggle less, lipids pack better, hence the area per lipid is smaller, and the diffusion constant of lipids drops by at least two orders of magnitude. Do such phases matter in nature? Rarely. In your body, there's only one location where they are common: the stratum corneum (the outermost layer of your epidermis). To be fair, there are people who are *deeply* interested in it: cosmetics companies want to make your skin look nice, or want to deliver all kinds of beneficial substances into deeper skin layers, and so they expensively worry about what's in the way. But if you are instead interested in lipid bilayers because you care about the membranes that surround cells or compartmentalize a variety of organelles inside cells, then you can usually ignore gel phases.

So, we have reached the point where we wish to think about lipid bilayers in their fluid phase. Given the size of lipids, these are something like 4 nm thick, but can laterally extend over many micrometer. That's an aspect ratio of something like 10 000:1, which is more extreme than the aspect ratio of a sheet of paper. Because of that, it seems like a pretty excellent idea to describe such membranes as two-dimensional surfaces embedded in three-dimensional space. How far down in scale this description still makes sense is of course open at this point, but as usual, we will discover that it works far better than it seemingly has any right to, even at the tens of nanometer scale. Also, we might worry that paper is indeed solid, but membranes are fluid, and whether this makes things more complicated. Turns out it actually makes things *simpler* because—in technical terms—the absence of a fixed reference shape also means there is no fixed reference *metric* we need to keep track of.

3.2 The geometry of fluid membranes

Our first task towards a continuum elastic description of lipid membranes is to find a way of representing two-dimensional surfaces embedded into three-dimensional space. There are many ways for how to do this; in fact, an entire field of mathematics—differential geometry—was born from the interest in this subject (and then *vastly* outgrew its original motivation). For the purpose of these lecture notes, I do not have the time to go into much detail, so I will restrict myself to some vignettes, chosen for being common or useful (ideally: both). Along the way I will make quite a few claims that I will not prove, but you can find these things easily in the literature (Kreyszig, 1991; do Carmo, 1976; Willmore, 2012; Spivak, 1970, 1975a,b; do Carmo, 1992; Lovelock

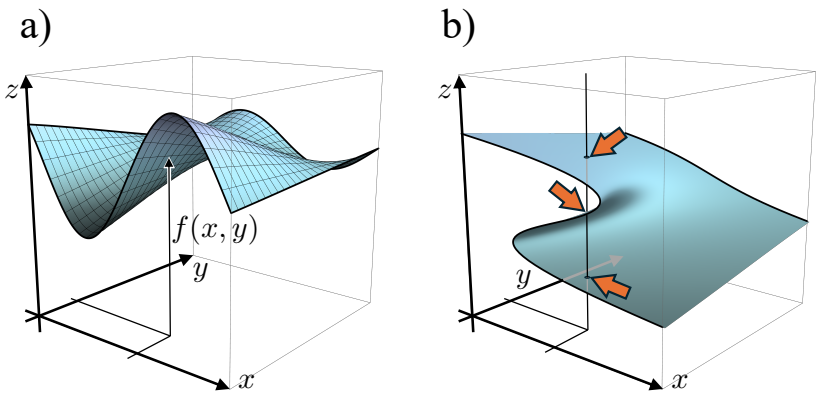


Figure 3.1 | Describing a surface via "Monge gauge": panel (a) shows a surface that can be represented as a function $f(x, y)$ above some reference plane. This is not possible for the surface in panel (b) since it has overhangs. In this specific case, the surface in (b) could also be described in Monge gauge after a 90° degree rotation about the y-axis, but it is not always possible to find such a simple fix.

and Rund, 1989; Frankel, 2004; Schutz, 1980; Darling, 1994; Flanders, 1989). I would also like to shamelessly draw your attention to a review article I wrote some years back in which I very specifically deal with a slightly more sophisticated differential geometric description of membranes, with the specific aim to discuss membrane stresses more efficiently (Deserno, 2015). It goes into much more detail, but still does not offer proofs of the type you expect in math books (because I am not a mathematician). It does try hard, though, to create physical intuition.

3.2.1 Monge gauge

The by far most common parametrization for membranes is called "Monge gauge." This sounds fancy (to be fair, anything with a name sounds fancy), but it really is the most straightforward twodimensionalification of the concept of plotting a function $f(x)$: just add another axis and plot the function $f(x, y)$ over some portion of a horizontal base plane into the third vertical direction. That's it. Panel (a) in Fig. 3.1 illustrates the idea.

A moment's thought reveals that this, of course, cannot describe every possible surface—for the same reason that a function $f(x)$ cannot describe every curve drawn in the two-dimensional plane. Most obviously, we will not be able to handle "overhangs"—surfaces with multiple "sheets" at some given (x, y) position—see panel (b) in Fig. 3.1. More sophisticated parametrizations can deal with this, and we will look at

one of them later, but even such more fancy approaches generally need to describe a surface as a collection of multiple patches. Since each such patch is a mapping (or "map"), the collection of all such maps is aptly called an "atlas", and of course one now has to define quite carefully how we can change between maps, or how we can transform coordinates from one map into another one, if these two maps happen to cover the same region on our surface.[1] The point is: even after cranking up mathematical sophistication we may still have to describe a complicated surface as a collection of multiple patches. Now observe that each individual patch, if small enough, might also be simple enough for Monge gauge to work! Why? Because near any point p a sufficiently smooth surface does not strongly deviate from its local tangent plane T_p, and if we rotate the surface so that T_p becomes horizontal, then in a vicinity of p the surface is perfectly describable by a height function $f(x, y)$. Long story short, Monge gauge is not as clumsy as it might initially sound.

Let's now measure some important things on the surface. The most basic one is area. If we have a little infinitesimal rectangular area element $\mathrm{d}x\,\mathrm{d}y$ on the base plane, this maps to a distorted parallelogram on the surface, and it is no longer flat. What is its area $\mathrm{d}A$? By carefully generalizing a simple thought that works for curves (see problem 1 at the end of this chapter), we find that

$$\mathrm{d}A = \sqrt{1 + \big(\boldsymbol{\nabla} f(x, y)\big)^2}\,\mathrm{d}x\,\mathrm{d}y \;, \tag{3.1}$$

where $\boldsymbol{\nabla} = (\partial/\partial x, \partial/\partial y)^\top$ is the gradient operator in the flat cartesian base plane.

Next, the other thing we really care about in a surface is its local *shape*. Given the key word "local", the obvious thing to try here is to just Taylor-expand the parametrization $f(x, y)$ near some point of interest and see what we get. To be able to write things down a bit more nicely, let's not talk about "x" and "y" but about x_1 and x_2, which together make up a vector $\boldsymbol{r} = (x_1, x_2)^\top$. In components, a local

[1] Deep down, the *real* challenge is this: maps are arbitrary, but our surface is not. We need to make sure that we drag as little as possible of that arbitrariness into our description. This recognition of the arbitrariness of coordinates is foundational to both the mathematical and the physical description, and both mathematicians and physicist will emphasize its importance. But they will use rather different words for doing so.

Taylor expansion of f near $\boldsymbol{r}^{(0)} = (x_1^{(0)}, x_2^{(0)})^\top$ is then:

$$\begin{aligned} f(x_1, x_2) = f(x_1^{(0)}, x_2^{(0)}) &+ \sum_{i=1}^{2} \left.\frac{\partial f}{\partial x_i}\right|_{\boldsymbol{r}^{(0)}} (x_i - x_i^{(0)}) \\ &+ \frac{1}{2} \sum_{i,j=1}^{2} \left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{\boldsymbol{r}^{(0)}} (x_i - x_i^{(0)})(x_j - x_j^{(0)}) \\ &+ \text{higher order terms} \,. \end{aligned} \tag{3.2}$$

Written ever so slightly more compactly in vector notation, we get

$$f(\boldsymbol{r}) = f(\boldsymbol{r}^{(0)}) + \boldsymbol{\nabla} f\Big|_{\boldsymbol{r}^{(0)}} \cdot (\boldsymbol{r} - \boldsymbol{r}^{(0)}) + \frac{1}{2} (\boldsymbol{r} - \boldsymbol{r}^{(0)})^\top \cdot \mathbf{H}\Big|_{\boldsymbol{r}^{(0)}} \cdot (\boldsymbol{r} - \boldsymbol{r}^{(0)}) + \cdots \,, \tag{3.3}$$

where $\mathbf{H}$ is the Hessian matrix

$$\mathbf{H} = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \dfrac{\partial^2 f}{\partial x_2^2} \end{pmatrix} \equiv \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix} \,, \tag{3.4}$$

which in Eqn. (3.3) we of course need to evaluate at the actual expansion point $\boldsymbol{r}^{(0)}$.

It seems eminently plausible that the Hessian matrix, consisting of second derivatives, has something to do with "curvature." This is true, but there are still a few subtleties hiding here. First, let's briefly recall the 1d situation, where a function $f(x)$ would be expanded as

$$f(x) = f(x^{(0)}) + f'(x^{(0)})\,(x - x^{(0)}) + \frac{1}{2} f''(x^{(0)})\,(x - x^{(0)})^2 + \cdots \,. \tag{3.5}$$

Is the second derivative at some point equal to the curvature at that point? Sadly—no. A quick way to convince ourselves that this cannot be so is to recall that for a parabola, $f(x) = x^2$, the second derivative is a constant, 2, and yet a parabola clearly does not have constant curvature—*it's not a circle*! Instead, the curvature is largest at its apex and then gets smaller and smaller the further we go along any of two branches. Again, one way to get some intuition about this is to work out what the curvature actually is for the 1d case, for some smooth function $f(x)$, and problem 2 at the end of the chapter walks you through an elementary calculation for doing so. This exercise shows that, indeed, the local slope matters when calculating the local curvature via Monge gauge. But it *also* shows that if the local slope vanishes, then the second derivative really *is* the curvature.

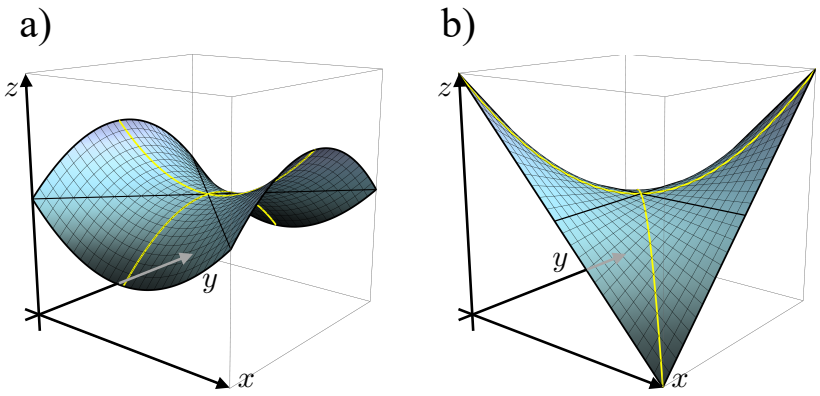


Figure 3.2 | The two functions defined in Eqn. (3.6). The left one in panel (a) is $f_1(x,y) = \frac{1}{2}(x^2 - y^2)$, the right one in panel (b) is $f_2(x,y) = xy$. The yellow curves illustrate cuts along the two orthogonal principal directions, of each surface, illustrating the lines of curvature. The black lines show that between those two there are also directions along which the curvature vanishes. This can of course only happen if the two principal curvatures have opposite sign (*i. e.*, one bends up, the other bends down). These so-called "asymptotic curves" generally do not intersect each other at right angles; that they do so here is simply due to the high symmetry.

So if we specialize to calculating the curvature at points where $\boldsymbol{\nabla} f = 0$ (we can always rotate the surface to make that happen), then maybe the Hessian $\mathbf{H}$ also describes the curvature. This is true, but we still need to figure out precisely in what way, because there sure are many curvatures at some point. What is "*the*" curvature?

It is no coincidence that the curvature state of a surface is indeed not described by a single number but by a matrix: at any point p, there are infinitely many direction along which we could ask how much the surface bends, and a matrix is what's needed to at least locally describe that. But we can do a bit better: notice that the Hessian in Eqn. (3.4) is *symmetric*, and as such it can be *diagonalized*. It has two eigenvectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ and two associated eigenvalues $\{c_1, c_2\}$. The eigenvectors point in what's called the two "principal directions," and their associated eigenvalues are the "principal curvatures" associated with these two directions. If they are different, the principal directions must further be orthogonal.

Let's look at two examples. Take the two functions

$$f_1(x,y) = \frac{1}{2}\big(x^2 - y^2\big) \qquad \text{and} \qquad f_2(x,y) = xy\ . \tag{3.6}$$

These are plotted in Fig. 3.2. Their gradient vanishes at the origin, so the Hessian suffices to characterize their curvature properties there.

Let's calculate it:

$$\mathbf{H}_{f_1} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \text{and} \qquad \mathbf{H}_{f_2} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} . \tag{3.7}$$

Obviously, $\mathbf{H}_{f_1}$ is already diagonal, showing that the x- and y-axis are the principal directions and the associated eigenvalues are $+1$ (the surface bends "up" along x) and -1 (the surface bends "down" along y. This checks out with Fig. 3.2.

In contrast, $\mathbf{H}_{f_2}$ is not diagonal, but we can easily guess two orthogonal eigenvectors: $\boldsymbol{v}_1 = (1,1)^\top$ and $\boldsymbol{v}_2 = (1,-1)^\top$. We evidently also have

$$c_1\boldsymbol{v}_1 = \mathbf{H}_{f_2}\boldsymbol{v}_1 \stackrel{!}{=} \boldsymbol{v}_1 \qquad \text{and} \qquad c_2\boldsymbol{v}_2 = \mathbf{H}_{f_2}\boldsymbol{v}_2 \stackrel{!}{=} -\boldsymbol{v}_2 . \tag{3.8}$$

This shows that the curvature along the $(1,1)$ direction is 1 and the curvature in the $(1,-1)$ direction is -1, which again tracks what we see in Fig. 3.2. Observe also that these are the same curvatures as those for the surface for f_1, just along directions that are rotated by 45°. Indeed, this is because the surface corresponding to f_2 is just the surface corresponding to f_1, rotated by 45 degree—as you are asked to confirm in problem 3 at the end of this chapter.

Once you know the principal directions and the principal curvatures, you know *all* curvatures along *all* directions. Say you care about the curvature along a direction that makes an angle φ with respect to principal direction $\boldsymbol{v}_1$. Then the curvature along that direction is

$$c(\varphi) = c_1 \cos^2\varphi + c_2 \sin^2\varphi . \tag{3.9}$$

This is called "Euler's Theorem," a name that is not particularly helpful when you search for it, because so much else in mathematics is also named after Euler. (He was a pretty sharp fellow.)

The curvature is related to the Hessian, and its eigenvalues are the principal curvatures. However, they by themselves are not yet good measures of the local surface curvature, because they still depend on the coordinates. Think about it: if we rotate the surfaces belonging to the functions f_1 or f_2 by 90°, we swap the principal directions and hence the eigenvalues, but nothing about the surfaces changes (other than their orientation in space). To get something that is independent from these remaining arbitrary coordinate choices, *i. e.*, *independent from the choice of a local basis*, we need to look at *invariants*. For instance, the sum of the eigenvalues is the *trace* of the Hessian, and we know from linear algebra that traces are invariant under basis changes

(*i. e.*, invertible linear coordinate transformations, such as for instance rotations). Another invariant of a matrix is its *determinant*, which is the product of the eigenvalues. These two invariants are extremely important and are given special names:

$$J := c_1 + c_2 \qquad \text{(total curvature)} \tag{3.10a}$$

$$K := \; c_1\, c_2 \qquad \text{(Gaussian curvature)}\ . \tag{3.10b}$$

I'd like to point out one little subtlety, related to the *sign* of the total curvature. As defined so far a surface such as $f(x,y) = \frac{1}{2}(x^2+y^2)$ would have positive total curvature $J = 2$ (check it!). Looking at the shape, this would mean that if you stand at the center of that parabola, and everywhere around you the floor bends up, that's positive curvature. Some like that convention, but others prefer that the curvature is positive when you stand on a hill, not when you stand at the bottom of a pit. But if you want hilltops such as $-\frac{1}{2}(x^2+y^2)$ to have positive curvature you need to flip the sign; meaning, you need to work with the eigenvalues of the *negative* Hessian. Moving forward, this is also the sign convention we will follow, such that for instance the total curvature at a horizontal point in Monge parametrization is given by

Heads up 5

Sign convention! This may be confusing, so be aware that not everyone agrees on how this specific sign is picked!

$$J \overset{\boldsymbol{\nabla} f=0}{=} \mathrm{Tr}[-\mathbf{H}] = \mathrm{Tr}\left[-\begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix}\right] \tag{3.11a}$$

$$= -\left(\frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2}\right) = -\Delta f\ . \tag{3.11b}$$

In words: the total curvature at horizontal points is the *negative* of the Laplacian of the height function.

If you were to do the calculation in full glory, even for a surface that is not horizontal at the point you care about, you instead find the much more complicated (since nonlinear!) expressions

$$J = -\boldsymbol{\nabla}\cdot\left(\frac{\boldsymbol{\nabla} f}{\sqrt{1+(\boldsymbol{\nabla} f)^2}}\right) = -\frac{(1+f_y^2)f_{xx} - 2f_x f_y f_{xy} + (1+f_x^2)f_{yy}}{\left[1+(\boldsymbol{\nabla} f)^2\right]^{3/2}}\ , \tag{3.12a}$$

$$K = \frac{\det(\mathbf{H})}{g^2} = \frac{f_{xx}f_{yy} - (f_{xy})^2}{\left[1+(\boldsymbol{\nabla} f)^2\right]^2}\ . \tag{3.12b}$$

In particular, Eqn. (3.12a) is the 2d generalization of what you (hopefully) worked out in 1d in problem 2 (up to possibly a sign that is—you now know—a bit of a convention.

While Monge gauge is intuitively straightforward, it nevertheless leads to pretty nonlinear expressions down the road—for instance because the occurrence of these square roots. However, there is a way to simplify the expressions for area and curvature by expanding them up to *quadratic* order in $\boldsymbol{\nabla} f$, because this, it turns out, will then later lead to *linear* shape equations. These approximate expressions are then referred to as "linear(ized) Monge gauge:"

$$\mathrm{d}A \approx \left[1 + \frac{1}{2}(\boldsymbol{\nabla} f)^2\right] \mathrm{d}x\, \mathrm{d}y \ , \tag{3.13a}$$

$$J \approx -\Delta f \ , \tag{3.13b}$$

$$K \approx f_{xx} f_{yy} - (f_{xy})^2 \ . \tag{3.13c}$$

What are the next order corrections to these results? For the area element and the Gaussian curvature this is pretty straightforward, since it merely amounts to Taylor expansions of expressions like $[1 + (\boldsymbol{\nabla} f)^2]^a$ with $a = \frac{1}{2}$ for the area element and $a = -2$ for the Gaussian curvature. The mean curvature is a bit more tricky. Some patient expansion work shows that the next term is *cubic* in f, and we get

$$J \approx -\Delta f + \boldsymbol{\nabla} f^\top \left[\mathbf{H} + \tfrac{1}{2}\mathrm{Tr}(\mathbf{H})\mathbb{1}\right] \boldsymbol{\nabla} f + (\text{quartic order}) \ , \tag{3.14}$$

Of course, $\mathrm{Tr}(\mathbf{H}) = \Delta f$, so the quadratic form is written a bit more fancily than it needs to be written. (But you gotta admit: it looks *so* cute that way!) At any rate, we again see that the corrections become sizable once $(\boldsymbol{\nabla} f)^2$ is no longer small compared to 1, which of course is the same condition we find for the next order expansion in the area element. All this is a bit easier to see in 1d, where the two conditions for area and curvature become

$$\mathrm{d}A \ = \ \left\{1 + \frac{1}{2}(f')^2\left[1 - \frac{1}{4}(f')^2 \pm \cdots\right]\right\} \mathrm{d}x\, \mathrm{d}y \ , \tag{3.15a}$$

$$J \ = \ -(f'')\left[1 - \frac{3}{2}(f')^2 \pm \cdots\right] . \tag{3.15b}$$

This also shows that, relatively speaking, the first correction in the curvature term is 6 times bigger than in the area term, suggesting that

curvature expressions will deteriorate more rapid in linear Monge than mere area expressions. This argument of course doesn't say whether the same is true in 2d.

3.2.2 A more general parametric surface representation

To show you one possible way for how to go beyond Monge gauge, and to take the first mini steps towards a more general differential geometric treatment, let us look at a parametrization that is a fair deal more general than Monge. Like essentially *any* parametrization, it will not free us from the need to explain what to do if we wish to describe a given patch of surface using different coordinates. In fact, this is a feature, not a bug, of every way of describing a surface (or, really, any geometric object): our ability to disentangle our discussion from any specific quirks of any given parametrization lies at the heart of doing proper geometry: coordinates are *always* arbitrary, and so we need to distill which bits and pieces are ultimately independent of them.

This section opens the door to a lot more differential geometry. But we will only be able to take a short peek, and we will not be able to look at all the many fun things that could be explored. We'll just whet the appetite. Those of you interested in more might find my review (Deserno, 2015) a helpful bridge between these notes and proper differential geometry books.

With that being said, consider the following way of describing a surface in 3d space:

$$(u^1,u^2)\mapsto \boldsymbol{X}(u^1,u^2)=\begin{pmatrix} X(u^1,u^2)\\ Y(u^1,u^2)\\ Z(u^2,u^2)\end{pmatrix}\in\mathbb{R}^3\ . \tag{3.16}$$

A first heads-up: *the superscripts "1" and "2" are not exponents*! They are the $(1,2)$ index pair that distinguish the two coordinates we use to describe a two-dimensional parametrization. You can legitimately ask: why don't we write this as u_1 and u_2, though? Answer: because it will turn out in the process of getting coordinate invariance right, that there will be two different types of two-index-objects showing up, which transform differently under coordinate changes, and the standard (and, it turns out, miraculously sleek) way of distinguishing them is by the placement of indices: upstairs versus downstairs. Of course, to sound less childish, we don't speak of "upstairs" and "downstairs" indices and components, but of "contravariant" and "covariant" components, respectively. More to follow. For now, look at Fig. 3.3 for two examples

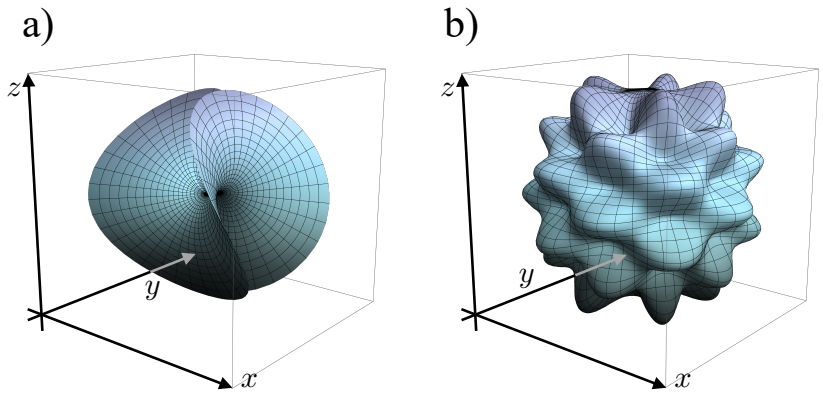


Figure 3.3 | Two examples of surfaces accessible via more general parametrizations. Panel (a) shows the "inner portion" of a surface called "Enneper's surface." It is a so-called "minimal surface" for which the total surface curvature J vanishes at every point. Observe that if we were to "continue" this surface beyond its current "rim," it would start to develop self-intersections—something which we cannot easily see from a local parametrization. Panel (b) shows an example of a surface plotted in "spherical Monge gauge". The specific example shows a surface whose wobbles follow the spherical harmonic $Y_{11}^5(\vartheta, \varphi)$.

of a surfaces that can be nicely expressed using this parametrization that we would struggle to get in conventional Monge gauge.

Let us make two examples that show how this works. As a first example, let us show that Monge gauge is actually a very special case of this parametrization:

$$(u^1, u^2) \mapsto \boldsymbol{X}(u^1, u^2) = \begin{pmatrix} u^1 \\ u^2 \\ f(u^2, u^2) \end{pmatrix} \in \mathbb{R}^3 \qquad \text{(Monge)} \ , \tag{3.17}$$

where we see that u^1 and u^2 can be viewed as the cartesian coordinates x and y of some base plane. Instead of three general functions $X(u^1, u^2)$, $Y(u^1, u^2)$, and $Z(u^1, u^2)$, we only have a single one, $Z(u^1, u^2) \equiv f(u^1, u^2)$, since we chose the other two to just be the identities for u^1 and u^2. This is why Monge gauge is not as general and has far more limitations.

Here's another example:

$$(u^1, u^2) \mapsto \boldsymbol{X}(u^1, u^2) = r \begin{pmatrix} \sin(u^1)\cos(u^2) \\ \sin(u^1)\sin(u^2) \\ \cos(u^1) \end{pmatrix} \in \mathbb{R}^3 \qquad \text{(sphere)} \ . \tag{3.18}$$

This is not a generic parametrization of a surface but a particular choice for a specific one: this describes the surface of a sphere of radius

r. I trust you will recognize this as *standard spherical polar coordinates* (but at *fixed* values of r), or at least when I tell you to rename $u^1 \to \vartheta$ and $u^2 \to \varphi$. You can make it more general by also letting r depend on u^1 and u^2, in which case you can look at a surface of spherical topology whose radius depends on the direction:

$$(u^1, u^2) \mapsto \boldsymbol{X}(u^1, u^2) = r(u^1, u^2) \begin{pmatrix} \sin(u^1)\cos(u^2) \\ \sin(u^1)\sin(u^2) \\ \cos(u^1) \end{pmatrix} \in \mathbb{R}^3 \qquad \text{(spherical Monge)} \,. \tag{3.19}$$

This is extremely useful when describing surfaces that only deviate weakly from spheres—in the same sense that Monge gauge is extremely useful for describing surfaces that deviate weakly from a plane. Fig. 3.3b is in fact an example of this parametrization.

Moving on. The two examples illustrated in Fig. 3.3 show a grid of curves plotted on the surface. What are these curves? Answer: these are curves of constant u^1 (or constant constant u^2) values. They are called "coordinate lines" or "coordinate curves." If you fix u^2 to some permissible value c and then only vary u^1, then you are walking on the coordinate curve called "$u^2 = c$". Notice that small variations in u^1 now define the "tangential direction" of that coordinate line. We can use this to set up a coordinate system for the local tangent plane to the surface by defining the two tangent vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$:

$$\boldsymbol{e}_1 = \frac{\partial \boldsymbol{X}}{\partial u^1} \qquad \text{and} \qquad \boldsymbol{e}_2 = \frac{\partial \boldsymbol{X}}{\partial u^2} \,, \tag{3.20}$$

or more compactly written:

$$\boxed{\boldsymbol{e}_i = \frac{\partial \boldsymbol{X}}{\partial u^i} \,.} \tag{3.21}$$

As you have undoubtedly noticed, we now use *downstairs* indices for the two different objects. This is because these objects genuinely behave differently under a change of coordinates than objects with their index upstairs. Problem 4 at the end of this section will give you a chance to explore this in more detail—if you're interested.

We can use the tangent vectors $\boldsymbol{e}_i$ as a basis for any other tangent vector $\boldsymbol{V}$ defined at that local point, thereby giving that tangent vector coordinates (V^1, V^2):

$$\boldsymbol{V} = V^1\boldsymbol{e}_1 + V^2\boldsymbol{e}_2 = \sum_{i=1}^{2} V^i \boldsymbol{e}_i \equiv V^i \boldsymbol{e}_i \,. \tag{3.22}$$

We dropped the summation sign in the last step—something that's called "Einstein summation convention:" whenever we see a repeated index pair, one upstairs and one downstairs,[2] then we sum over that pair over the permissible range (which here is of course just from 1 to 2). We'll introduce this here and *immediately* use it, because it's so convenient. As it turns out, differential geometry is a lot of local expansions in local coordinate systems, and so there's *a lot* of these sums. The notation massively brightens up once we drop all these sum symbols. The repeated index is called a "dummy index," and it is immaterial which letter we use for it—as long as it is different from any other index-letter already occurring in our expression.

Heads up 6

Attention, watershed moment: Einstein summation convention from here onwards!

Observe that these tangent vectors are generally not normalized: Their scalar products $\boldsymbol{e}_1 \cdot \boldsymbol{e}_1$ and $\boldsymbol{e}_2 \cdot \boldsymbol{e}_2$ are generally not 1. We will insist, though, that they have to be at least nonzero, because otherwise something singular and obviously bad has happened. If this is not the case, our parametrization is sick.

Observe *also* that the two tangent vectors are generally not orthogonal: $\boldsymbol{e}_1 \cdot \boldsymbol{e}_2$ does not have to vanish. If it does, the coordinates are called "orthogonal coordinates," but this is something we are generally not guaranteed.

Since all these scalar products are obviously interesting, it turns out to be useful to define an object that collects them all in one place. So here we go:

$$g_{ij} := \boldsymbol{e}_i \cdot \boldsymbol{e}_j \ . \tag{3.23}$$

Observe from the definition that g_{ij} is symmetric under exchange of i and j.

The object g_{ij} is called the "*metric tensor*," or just short the "*metric*." It is *fantastically* important in all of differential geometry. For instance, it is the object that is used to define scalar products of vectors. Say you have two tangent vector $\boldsymbol{V}$ and $\boldsymbol{W}$ on the surface expressed in the local tangent coordinate system. Their scalar product is

$$\boldsymbol{V} \cdot \boldsymbol{W} = (V^i \boldsymbol{e}_i) \cdot (W^j \boldsymbol{e}_j) = V^i W^j (\boldsymbol{e}_i \cdot \boldsymbol{e}_j) = V^i W^j g_{ij} \ . \tag{3.24}$$

And now that we have a scalar product, we can also define a *length*:

$$|\boldsymbol{V}| = \sqrt{\boldsymbol{V} \cdot \boldsymbol{V}} \quad \text{with} \quad \boldsymbol{V} \cdot \boldsymbol{V} = V^i V^j g_{ij} \ . \tag{3.25}$$

Moreover, it is extraordinary useful to define the "inverse metric", another tensor g^{ij}, with its indices upstairs rather than downstairs, basi-

[2] Sorry: *one contravariant and one covariant.*

cally as being the inverse matrix of g_{ij}, such that

$$g^{ij}g_{jk} = \delta^i_k = \begin{cases} 1 \text{ if } i = k \\ 0 \text{ if } i \neq k \end{cases} , \tag{3.26}$$

where the object δ^i_k is called the "Kronecker symbol" (which I trust you've seen before, maybe just not yet with its indices in upstairs-downstairs placement).

The tensors g^{ij} and g_{ij} can be used to raise and lower indices, respectively. For instance we have

$$V_j = g_{ij}V^j \qquad \text{and} \qquad \boldsymbol{e}^i = g^{ij}\boldsymbol{e}_j \ , \tag{3.27}$$

showing that if you "dislike" the placement of any index, you can easily change it. Observe that this is not a moot procedure: the *numbers* associated with (V^1, V^2) are obviously *not* the same as the numbers (V_1, V_2), because the two objects g_{ij} and g^{ij} are not just unit matrices. If we ever encounter an object that happens to have more than one index, we can do the same. In that case, though, we will have to be careful to say which index we raise or lower—meaning, whether it's the first, second, third, *etc.* index. So we need to make sure that the *horizontal* placement of indices is not affected. Here's a terrifying looking example:

$$g^{mp}T_{ijklmn} = T_{ijkl}{}^{p}{}_{n} \ . \tag{3.28}$$

With this index raising and lowering gymnastics we can also nicely rewrite scalar products:

$$\boldsymbol{V} \cdot \boldsymbol{W} = V^i W^j g_{ij} = V^i W_i = V_j W^j \tag{3.29a}$$

... or lengths:

$$\boldsymbol{V} \cdot \boldsymbol{V} = V^i V^j g_{ij} = V^i V_i \ . \tag{3.29b}$$

Knowing lengths, we can also calculate areas. The most important example would be the area of a tiny little coordinate patch—one of the small parallelograms you can see on the gridded surface in Fig. 3.3. Convince yourself that these patches are locally "spanned" by $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$. If we want to consider infinitesimal area elements that correspond to an infinitesimally finely sliced coordinate mesh, then their area can be expressed as

$$\mathrm{d}A = \left|\mathrm{d}\boldsymbol{X}_{\text{along }1} \times \mathrm{d}\boldsymbol{X}_{\text{along }2}\right| = \left|\boldsymbol{e}_1 \times \boldsymbol{e}_2\right| \mathrm{d}u^1 \, \mathrm{d}u^2 \ , \tag{3.30}$$

where we used the fact that the area of a parallelogram spanned by vectors $\boldsymbol{V}$ and $\boldsymbol{W}$ is $|\boldsymbol{V} \times \boldsymbol{W}|$. Using the well-known vector identity

$$(\boldsymbol{a} \times \boldsymbol{b}) \cdot (\boldsymbol{c} \times \boldsymbol{d}) = (\boldsymbol{a} \cdot \boldsymbol{c})(\boldsymbol{b} \cdot \boldsymbol{d}) - (\boldsymbol{a} \cdot \boldsymbol{d})(\boldsymbol{b} \cdot \boldsymbol{c}) \ , \tag{3.31}$$

we can now calculate the following:

$$\begin{aligned}\big|\boldsymbol{e}_1 \times \boldsymbol{e}_2\big|^2 &= (\boldsymbol{e}_1 \times \boldsymbol{e}_2)\cdot(\boldsymbol{e}_1 \times \boldsymbol{e}_2) \\ &= (\underbrace{\boldsymbol{e}_1 \cdot \boldsymbol{e}_1}_{g_{11}})(\underbrace{\boldsymbol{e}_2 \cdot \boldsymbol{e}_2}_{g_{22}}) - (\underbrace{\boldsymbol{e}_1 \cdot \boldsymbol{e}_2}_{g_{12}})(\underbrace{\boldsymbol{e}_2 \cdot \boldsymbol{e}_1}_{g_{21}}) \\ &= \det\big(g_{ij}\big) \;\equiv\; g \;, \end{aligned} \tag{3.32}$$

where $g \equiv \det(g_{ij})$ is the so-called *metric determinant*. Combining this with Eqn. (3.30), we arrive at the important result

$$\boxed{\mathrm{d}A = \sqrt{g}\,\mathrm{d}u^1\,\mathrm{d}u^2\;.} \tag{3.33}$$

The prefactor $\sqrt{g}$ is something you have seen before in another context—just under a different name: "Jacobian." Think for instance about spherical polar coordinates. The area element on the surface of a sphere is obviously not just $\mathrm{d}\vartheta\,\mathrm{d}\varphi$. That wouldn't even make sense dimensionally! Instead, you know you need to include the Jacobian for spherical polar coordinates, finding $\mathrm{d}A = R^2\sin\vartheta\,\mathrm{d}\vartheta\,\mathrm{d}\varphi$.

It is obviously useful to extend the local coordinate system for the tangent plane, $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ into a complete coordinate system for 3d space, which just happens to be locally pinned to the surface. All we need to do so is include the normal vector $\boldsymbol{n}$, which is given by

$$\boxed{\boldsymbol{n} = \frac{\boldsymbol{e}_1 \times \boldsymbol{e}_2}{\big|\boldsymbol{e}_1 \times \boldsymbol{e}_2\big|} = \frac{\boldsymbol{e}_1 \times \boldsymbol{e}_2}{\sqrt{g}}\;.} \tag{3.34}$$

Notice that unlike the two tangent vectors, the normal vector is normalized to length 1. It just turns out to be more useful that way.

Now that we "have" a normal vector, we can finally discuss the crucial notion of *curvature*. When is a surface curved? One way to say so is that if the normal vector changes direction as we move along the surface. Since $\boldsymbol{n}$ is, well, a vector, we still need to re-express this as something that is surface based, and the right thing to do is to project the change of the normal vector back onto the surface. Let us hence define another tensor, K_{ij} as follows:

$$\boxed{K_{ij} = \boldsymbol{e}_i \cdot \frac{\partial \boldsymbol{n}}{\partial u^j} = -\boldsymbol{n}\cdot\frac{\partial \boldsymbol{e}_i}{\partial u^j} = -\boldsymbol{n}\cdot\frac{\partial^2 \boldsymbol{X}}{\partial u^i\,\partial u^j}\;.} \tag{3.35}$$

The second identity follows from differentiating $\boldsymbol{n}\cdot\boldsymbol{e}_i = 0$, and it is often easier to work out, because it lacks the additional normalization factor

$1/\sqrt{g}$ tacked onto the normal vector that tends to make calculating its derivatives more cumbersome. The last expression shows that the curvature tensor K_{ij} is symmetric in its indices. It also shows that it is a second derivative of the embedding function $\boldsymbol{X}(u^1, u^2)$, which indeed smells a lot like curvature. However, it generally doesn't have the right units yet: If the coordinates u^i are dimensionless (think for instance of angles), and $\boldsymbol{X}$ is the position in space, which must have units of length, and the normal vector is by construction dimensionless, then K_{ij} does not have the dimension of curvature. Turns out, one thing is still missing: We need to raise one of the indices in order to turn the tensor K_{ij} into the matrix K_i^j:

$$K_i^j = K_{ik} g^{kj} \ , \tag{3.36}$$

and this object now has the right dimension. It is called the "*shape operator*," because it is a linear operator that maps tangent vectors to other tangent vectors in a way that's informative about the shape (more specifically: the local curvature). Of course, here, we only have it in components, but we see that we can write

$$K_i^j V^i = W^j \ , \tag{3.37}$$

so that the components V^i of the tangent vector $\boldsymbol{V} = V^i \boldsymbol{e}_i$ are mapped to the components W^j of the tangent vector $\boldsymbol{W} = W^j \boldsymbol{e}_j$. Being a linear map, we can ask what it's eigenvalues are—meaning which vectors $\boldsymbol{V}$ are mapped onto themselves, with at most a scaling of length? Answer: these vectors exactly correspond to the principal directions of curvature of the surface at that point, and the corresponding eigenvalues are the principal curvatures. Because of that, and recalling Eqns. (3.10a) and (3.10b), the trace and the determinant of the shape operator are the total and the Gaussian curvature, respectively:

$$J = \mathrm{Tr}\big(K_i^j\big) \tag{3.38a}$$

$$K = \det\big(K_i^j\big) \ . \tag{3.38b}$$

This means we now have a pretty general way to compute the key curvature invariants of a surface in a parametrization that's fairly general.

3.3 The elasticity of fluid membranes

3.3.1 Some big-picture words

Elasticity quantifies the ability of a body to reversibly resist deformations. You squish it, with some force, leading to some deformation,

and when you stop squishing, the body returns to its original shape. Elasticity is hence thought of as a property of *solids*, but whether some squishing is reversible or not might also depend on the type of deformation. For instance, the compression of gases can very well be done reversibly: think of increasing the pressure from $P = P_1$ to $P = P_2 > P_1$ and subsequently decreasing it back to P_1. Gases and liquids indeed have a bulk modulus. Of course, all the molecules of the gas will have arrived at a different position, but the macrostate is unchanged. In contrast, when we *shear* a fluid, no force is required for the deformation (if done slowly enough—we're not talking about dissipation due to a finite shear *rate*). But then, while we change the position of molecules upon shearing, we don't change the macrostate. So are we even *doing* anything?

Oh dear. It seems like we're running into semantic issues as to what we even mean by elasticity already in the first paragraph. Maybe we need some expert back-up. Landau and Lifshitz (1986) start their entire elasticity volume by saying "The mechanics of solid bodies, regarded as continuous media, forms the content of the *theory of elasticity*." OK, solids. Also: a continuum point of view. That's nice, but what about membranes? Haven't we said that the lipids are not chemically bound? So aren't membranes *fluid*? How can they also be *elastic*?

A better (but more formal) way to approach the problem is to listen to what Chaikin and Lubensky (1995) have to say in their book—specifically chapter 6 on "Generalized Elasticity." They take the very broad approach of reminding us that phase transitions break symmetries of disordered phases. For instance, going from a fluid to a crystalline solid breaks the translation symmetry of the system, because translations now only transform the system into an identical copy if they happen along a lattice vector. Things are really interesting when the broken symmetry is continuous, because even though some arbitrary but uniform continuous symmetry transformation θ generally doesn't map the system back onto itself (say, when you translate a crystalline lattice by *half* a lattice vector), it is still true that such a deformation doesn't change the free energy. However, what happens if the transformation is *non-uniform*, *i. e.*, what if $\theta(\boldsymbol{r})$ depends on position? What if we translate the system at each point, but a little bit more here and a little bit less there? If we do that, then we are obviously squishing the system, and that might cost an additional free energy. It is then reasonable to expect that we can expand the free energy density in gradients of the deformation, $\boldsymbol{\nabla}\theta(\boldsymbol{r})$, especially if these gradients are slowly varying compared to any microscopic structure of the system itself—which is what will give us the permission to describe this in a

continuum language. This type of gradient expansion of the free energy is what we mean by (general) elasticity theory. For every continuously broken symmetry we can then write down a suitable set of gradients that end up giving us scalar expressions (because the free energy density must be a scalar) and that still satisfy any remaining symmetry of the system. This definition has the advantage of permitting a much more fine-grained discussion of elastic responses. For instance, a fluid membrane might indeed have all its molecules diffusing in the membrane plane, but a nonuniform stretching or bending deformation will still cost energy, and hence there are deformations against which a fluid membrane elastically resists.

3.3.2 Membrane stretching

With those preliminaries out of the way, let's think about the easiest deformation we can do to membranes that should elicit an elastic response—the equivalent of a uniform stretching or compression that we could even do for liquids and gases: a uniform stretching or compression of the *area* of a membrane.

Upon second thought: let's be careful with compressing. As we can easily imagine, the absence of a bulk third dimension gives the membrane the opportunity to escape into the third dimension upon compression (something that's called "buckling"), and that doesn't look like a pure stretching deformation. So let's for now stick with pulling.

So consider a membrane that in its relaxed state (no pushing, no pulling) prefers a flat state of area A_0. What if we pull it to a slightly larger area $A > A_0$? The associated (dimensionless) strain is given by $s = (A - A_0)/A_0$, and we expect our free energy to be a function of that. It cannot be a linear function, since it must have a minimum at A_0, and so we know that, to lowest order, the energy density is quadratic in the strain:

$$e_{\text{stretch}} = \frac{1}{2} K_A \, s^2 + \cdots = \frac{1}{2} K_A \left(\frac{A - A_0}{A_0} \right)^2 + \cdots \ . \tag{3.39}$$

The prefactor $\frac{1}{2}$ is convention, and the material parameter K_A is called the stretching (or compression) modulus. Recall that for now we just think of this strain to be homogeneous across the entire membrane, and so if we want the total stretching energy of the membrane, all we need to do is to multiply it by the (relaxed, reference-state) membrane area A_0. At quadratic order, which is what physicist-level elasticity theory

typically restricts to, this gives the total stretching energy

$$E_{\text{stretch}} = \frac{1}{2} K_A \frac{(A - A_0)^2}{A_0} \ . \tag{3.40}$$

The value of the modulus K_A has been measured for many cases. It turns out that for most "standard" membranes its value is—within about 10%—somewhere in the ballpark of 240 mN/m (Rawicz et al., 2000). What does that mean? Imagine a spherical membrane vesicle of radius R whose area is perfectly relaxed. Now increase the radius by a small fraction p. The area increases by $(1+p)^2$ and so the strain is $(1+p)^2 - 1 = 2p + p^2 \approx 2p$. That means we accumulate the stretching energy $\Delta E_{\text{stretch}} = pK_A(4\pi R^2)$. For vesicles with a diameter of 100 nm, this gives $\Delta E_{\text{stretch}} \approx p \times 1.8 \times 10^6 k_{\text{B}}T$, meaning a change in radius by 0.1% increases the energy by about $1800\, k_{\text{B}}T$. This is quite a bit.

If you pull membranes a lot, then at some point they will rupture. When? Turns out, this depends on how quickly you increase the area strain. The faster you do this, the more strain you can put in before it fails. The reason is that the membrane needs some time for a fluctuation to open a pore, and if you're pulling quickly, you can keep increasing the strain while the system is waiting for that lethal fluctuation to happen (this is also the basic idea behind dynamic force spectroscopy). Applied to membranes, you find details in (Evans and Heinrich, 2003). A rough rule of thumb is that membranes will rupture under a stress in the few-to-ten mN/m range, which corresponds to a strain of a few percent—but details depend on the lipid and the conditions.

3.3.3 Membrane bending

Just like the free energy can depend on area strain, it can also depend on local curvature, and if we are in the business of writing free energies as expansions of gradients (or higher order derivatives), then we can take the scalar invariants we discovered in the previous section as legitimate ingredients for such an expansion. Up to quadratic order in the eigenvalues of the curvature tensor (or, equivalently, the shape operator), we get something like

$$e = c_0 + c_1 J + c_2 J^2 + c_3 K + \text{higher order terms} \ . \tag{3.41}$$

The first term is just a constant, and after integrating over the whole membrane, you can view it in a number of ways: surface tension, Lagrange multiplier that fixes the area, chemical potential of lipids. The

precise interpretation of this easiest-looking term is fraught with surprisingly many subtleties, but in this introduction we have to delegate them to a nice spot under the carpet.

The next term, linear in J, has the interesting property of changing sign if we flip the direction into which a membrane bends. The two height functions $f_\pm(x,y) = \pm\frac{1}{2}c(x^2+y^2)$ differ in the sign of their J. Is this reasonable? That depends on the membrane! If the membrane does not change its structure upon flipping it upside down—meaning, if it is up-down symmetric—we would not expect its energy to change either, and then we would not expect such a term to exist. In other words, the prefactor c_1 would have to be zero. However, for a membrane that is *not* up-down symmetric, no such argument holds, and so we generally expect that term to be present. This shows that the existence of this term depends on an important symmetry property of the membrane. Specifically, since lipid membranes are bilayers, we expect this term to be present if there is any physically relevant difference between the two leaflets.[3] The remaining two terms are both quadratic expressions in the principal curvatures, and as such we clearly expect them to express some genuine curvature elasticity.[4]

While the expansion in Eqn. (3.41) is perfectly fine, physicists have decided to rewrite it slightly differently and give the constants slightly different names. Of course, even what I'm writing here is not universally agreed upon:

$$e_{\text{bend}} = \sigma + \frac{1}{2}\kappa(J - J_{0\text{b}})^2 + \overline{\kappa}K \ . \tag{3.42}$$

Here, σ is the called the (bare) *surface tension* (which is related to c_0 from Eqn. (3.41) and hence inherits all the subtle caveats I mentioned and then refused to talk about); κ is called the *bending modulus*; $J_{0\text{b}}$ is called the *spontaneous bilayer curvature*; and $\overline{\kappa}$ is called the *Gaussian curvature modulus*. Notice that $J_{0\text{b}}$ is nonzero if and only if c_1 is nonzero, *i. e.*, if and only if there's something that breaks the up-down symmetry of the bilayer.

Since this is just the energy density, and its value depends on the

[3] In fact, to show signs of asymmetry it even suffices if there is a seemingly benign difference in the *embedding solvents* (Lipowsky and Döbereiner, 1998).

[4] You might wonder whether there are other quadratic invariants. We have $J^2 = (c_1+c_2)^2$ and $K = c_1c_2$, but what about $c_1^2+c_2^2$? We would first have to convince ourselves that this is actually an invariant. It is, because it's the trace of the square of the curvature tensor. But it is not independent, because evidently $c_1^2 + c_2^2 = J^2 - 2K$. More generally, one can prove that for any n-dimensional square matrix $\mathbf{M}$, there are only two quadratic expressions that are invariant (under all rotations, meaning under $\mathrm{SO}(n)$), and these are $[\mathrm{Tr}(\mathbf{M})]^2$ and $\mathrm{Tr}(\mathbf{M}^2)$. Every other quadratic invariant is a linear combination of those two.

position on the membrane (unless the membrane is a plane, a sphere, or a cylinder), we get the total energy by *integrating* over the whole membrane:

$$E_{\text{bend}}[\mathcal{S}] = \int_{\mathcal{S}} \mathrm{d}A \left\{ \sigma + \frac{1}{2}\kappa(J - J_{0\text{b}})^2 + \overline{\kappa} K \right\} . \tag{3.43}$$

Here "$\mathcal{S}$" is not just the surface area of the membrane, but the actual *surface—i. e.*, the thing that has a *shape*. This means that *the energy is a functional of the shape*, and hence a functional of a geometric object. Observe that "$\mathrm{d}A$" is the area element *on the surface*, not on some Monge-flavored base-plane. In terms of the differential geometric language we have developed in Sec. 3.2.2, it is area element given in Eqn. (3.33). Within the simpler Monge-gauge from Sec. 3.2.1, it is the one given in Eqn. (3.1)—which is *still* not equal to the base-plane area element $\mathrm{d}x\,\mathrm{d}y$.

This energy functional is typically referred to as the *Helfrich energy*, after the original massively influential paper by Wolfgang Helfrich (Helfrich, 1973) and a lot of groundbreakign follow-up from there; but people also sometimes add the names of Peter Canham (Canham, 1970) and Evan Evans (Evans, 1974), since they, too, clearly had been independently on the right track, and especially Evans has later contributed in many important ways to the experimental development of the field.

Let's count the number of "parameters" entering this expression: We have σ, $J_{0\text{b}}$, κ, and $\overline{\kappa}$. That's four numbers we need to know before we can calculate the energy of a particular shape. But in practice it's often less. The spontaneous curvature for instance often vanishes for symmetry reasons. And the Gaussian curvature modulus $\overline{\kappa}$ is often irrelevant (even though it by no means vanishes) due to some piece of magic coming from differential geometry, called the *Gauss Bonnet theorem*. This theorem states that the integral over the Gaussian curvature of some surface $\mathcal{S}$ can be taken apart into two contributions, a boundary term and a topological invariant (do Carmo, 1976):

$$\int_{\mathcal{S}} \mathrm{d}A\; K + \int_{\partial\mathcal{S}} \mathrm{d}s\; k_{\text{g}} = 2\pi\chi(\mathcal{S}) . \tag{3.44}$$

Here, k_{g} is the *geodesic curvature* at the boundary,[5] and $\chi(\mathcal{S})$ is the *Euler characteristic* of your surface—a topological invariant that does

[5] Picture driving along that boundary with a car such that your car stands flat on the surface, but super close to its edge. As you drive along the edge, aiming

not change under smooth deformations that do not rip things off or glue things together.[6]

What this remarkable theorem states is that the only way to make the Gaussian curvature term contribute nontrivially to the Helfrich energy is if your membrane either has a boundary that somehow changes, or if you change the topology of your surface (say, you fuse two vesicles into a single one). Since often neither happens, the Gaussian term is just a constant that can be ignored—and therefore very often is. This is good news and bad news. It's good news whenever you have the opportunity to drop a term that you know doesn't matter. It is bad news if you are facing one of these rare situations where it *does* matter, and now you not only have to worry about handling all the mathematics associated with it, but you also need to actually know the numerical value of the Gaussian modulus $\overline{\kappa}$. How would you measure it if you have to? Well, you have to consider experiments that change either the boundary or the topology of your membrane in order to be sensitive to it. But those tend to be difficult experiments, because it is very difficult to control an open edge of a membrane, and it is also difficult to monitor topological transformations in such a way that you can also measure the energy changes associated with them.

If you are lucky enough to be able to ignore Gaussian curvature shenanigans, and you have a symmetric membrane, then there are only two parameters left: the tension σ and the bending modulus κ. The tension is not a material parameter and instead controlled by the setup: how inflated are vesicles, what does a cell do to its plasma membrane, *etc.* It can for instance be measured in a variety of ways that push or pull the membrane, or monitor the pressures needed to inflate vesicles. The bending rigidity can also be measured in many ways, and it is a genuine material parameter (which depends on environmental conditions such as temperature, ionic strength, or pH). A ball-park number for its value is a few tens of $k_{\mathrm{B}}T$, often between $20\,k_{\mathrm{B}}T$ and $30\,k_{\mathrm{B}}T$. It depends quite a lot on the type of lipid, whether you have mixtures, or other solutes (such as small peptides or alcohols) in the membrane. It's a busy field, producing many numbers, not quite as many error bars, and surely some interesting controversies. Happy googling.

---

to stay stay super close, but also not to fall off, you will generally have to turn your steering wheel. The associated curvature with this turning is called the geodesic curvature. For instance, if the surface is a circular disk of radius R, then $k_{\mathrm{g}} = 1/R$. But if your surface is a hemispherical cap of radius R, then driving around what now basically amounts to the equator, you do not need to turn the steering wheel at all, and so $k_{\mathrm{g}} = 0$.

[6] For example, $\chi(\text{disk}) = 1$, $\chi(\text{sphere}) = 2$, $\chi(\text{torus}) = 0$.

Since we will subsequently do a lot of work in the linearized Monge gauge, it is worth finding out what the Helfrich energy looks like in this approximate parametrization. This means we need to expand everything to the right hand side of $\int$ up to quadratic order in f. Notice that I *didn't* say we need to expand "the integrand"; there's also a "$\mathrm{d}A$" we must expand! But we know how to do this—that's just Eqn. (3.13a). For the curvature we will take Eqn. (3.13b), which is admittedly only linear, not quadratic, but we know that the next order adds a cubic (see Eqn. (3.14)), and so we stick with the linear term. Actually, we're going to square it up, so we get something quadratic anyways. Well, but we have $(J - J_{0\mathrm{b}})^2$, which will also create linear and constant terms. We might be tempted to selectively multiply some quadratic terms from the area element into the constant $J_{0\mathrm{b}}^2$ term, which we would of course not have to do with the J^2 term, but people prefer to keep everything in the curvature term at the same order of expansion. Doing this, and for simplicity ignoring the Gaussian term by chanting "Gauss-Bonnet!", we find

$$\begin{aligned} E_{\mathrm{bend}} &\approx \int_{\mathcal{R}} \mathrm{d}x\,\mathrm{d}y\, \left[1 + \frac{1}{2}(\boldsymbol{\nabla} f)^2\right] \left\{\sigma + \frac{1}{2}\kappa(-\Delta f - J_{0\mathrm{b}})^2\right\} \\ &\approx \int_{\mathcal{R}} \mathrm{d}x\,\mathrm{d}y\, \left\{\sigma\left[1 + \frac{1}{2}(\boldsymbol{\nabla} f)^2\right] + \frac{1}{2}\kappa(-\Delta f - J_{0\mathrm{b}})^2\right\} . \end{aligned} \tag{3.45}$$

Observe now that the "1+" term in the tension part will just add a constant, namely σA_0, where A_0 is the area of the base plane region $\mathcal{R}$. This is typically dropped out, as a shift in the zero of the energy, so that the tension contribution really measures any *excess* energy accrued from the membrane *not* being just horizontal.

If we finally simplify to the frequent case $J_{0\mathrm{b}} = 0$, then the linearized Monge Helfrich functional looks as follows:

$$\boxed{E_{\mathrm{bend}} \approx \int_{\mathcal{R}} \mathrm{d}x\,\mathrm{d}y\, \left\{\frac{1}{2}\sigma(\boldsymbol{\nabla} f)^2 + \frac{1}{2}\kappa(\Delta f)^2\right\} .} \tag{3.46}$$

I suspect this is *by far* the most common form in which you will encounter the Helfrich expression for a membrane's bending energy. But please recall that we have dropped a whole bunch of things along the way. It is not too hard, though, to add them back, if for whatever reasons you do not want to set $J_{0\mathrm{b}} = 0$, or if you want to keep the Gaussian term in (Eqn (3.13c) will be your friend), or if you don't want to linearize. Of course, it's harder to work with the more general equations, especially if you don't linearize, but at least you know *in principle* what you'd have to write down.

3.3.4 Shape equations

An archetypal question in any Physics 101 course is the following: Here's a spring. Its spring constant is k. We pull it with a force F. How far does it extend?

We'll now *basically* ask same type of question, just for a membrane:

Here's a membrane. It has rigidities κ and $\overline{\kappa}$. We apply some boundary conditions to it. What shape does it take?

Of course, this is significantly harder, because we do not just have a single degree of freedom (the spring's elongation) but a whole surface. Yet, problems of this type are what variational calculus has been invented for. If you've never seen this before, I cannot adequately prepare you for the wonders that come with it, and you just have to valiantly hang in there for a wide eyed ride. If you know what this is, you can probably immediately write down the solution to this problem. I will opt for a presentation that will revisit some of the key steps, paying particular attention to something even most cognoscenti of this technique might not have seen before, namely, what to do with boundary terms.

I should also say that I will only show how this is done for linearized Monge gauge—specifically, Eqn. (3.46). It can, of course, also be done for the more general parametrizations, but then we really need to unpack a fair bit more differential geometry than I have time to develop. But it's OK. The gist of the idea will also become clear this way, and a lot of the work people do in this field is still done with the linear Monge shape equations. Rest assured these are definitely not just the baby versions of the real thing. They are used.

The basic idea is this. Assume we actually know the optimal shape that the surface would take—let's call it $f^\star(x, y)$. If we plug this into the Helfrich energy functional, it will minimize it. How do we know that it's minimized? Well, how do you know when you have a function $f(x)$, and someone claims there's a minimum at $x = x_0$, that this is actually true? You could shift x a tiny bit away from the minimum, $x \to x_0 + \delta x$, and if x_0 really was a minimum, then the value of $f(x)$ should increase. Even more basic: if you are at a differentiable minimum, then to first order the value of $f(x)$ should not change at all, because at a minimum the Taylor expansion of $f(x)$ should start with the quadratic term.

Our goal is to translate this thinking to our variational problem, but what we need to shift around is the whole surface. And by "shift around" we don't just mean moving it up and down; we mean smoothly deforming the surface a little bit at each point: $f(x, y) \to f^\star(x, y) +$

$\delta f(x,y)$. Pay close attention to a very important change in perspective: "f" switches meaning in an important way. In the motivational "minimize $f(x)$" story, f was the function to be minimized, and we wiggled its argument x. Now the function to be minimized is the Helfrich functional, $E[f]$, and *its argument is the entire function* $f(x,y)$.

Let's start modest and look at the tension term in the Helfrich energy from Eqn. (3.46), inserting our variation:

$$\begin{aligned} e_{\text{tension}} = \frac{1}{2}\sigma\big[\boldsymbol{\nabla} f\big]^2 &= \frac{1}{2}\sigma\big[\boldsymbol{\nabla}(f^\star + \delta f)\big]^2 \\ &= \frac{1}{2}\sigma\big[(\boldsymbol{\nabla} f^\star)^2 + 2\,\boldsymbol{\nabla} f^\star \cdot \boldsymbol{\nabla}\delta f + (\boldsymbol{\nabla}\delta f)^2\big] \ . \end{aligned} \tag{3.47}$$

The first term on the right hand side is the value which this tension expression takes at the supposed minimum. Furthermore, the last term is quadratic in δf, and our goal was to only expand the Helfrich functional to *linear* order in the variation δf, since this should be sufficient to identify the "no change at linear order" condition. That means, we can now work out, to linear order, the variation $\delta e_{\text{tension}}$ in the tension part of the energy:

$$\begin{aligned} \delta e_{\text{tension}} &= \frac{1}{2}\sigma\big[\boldsymbol{\nabla} f\big]^2 - \frac{1}{2}\sigma\big[\boldsymbol{\nabla} f^\star\big]^2 \\ &= \sigma\,\boldsymbol{\nabla} f^\star \cdot \boldsymbol{\nabla}\delta f + \mathcal{O}(\delta f^2) \ . \end{aligned} \tag{3.48a}$$

By exactly the same reasoning, we can also work out the variation of the bending part in the Helfrich energy:

$$\begin{aligned} \delta e_{\text{bending}} &= \frac{1}{2}\kappa\big[\Delta f\big]^2 - \frac{1}{2}\kappa\big[\Delta f^\star\big]^2 \\ &= \kappa\,\Delta f^\star\,\Delta\delta f + \mathcal{O}(\delta f^2) \ . \end{aligned} \tag{3.48b}$$

Let's insert this into the functional and see what we get:

$$\delta E_{\text{bend}} = \int_{\mathcal{R}} \mathrm{d}x\,\mathrm{d}y\,\Big\{\sigma\,\boldsymbol{\nabla} f^\star \cdot \boldsymbol{\nabla}\delta f + \kappa\,\Delta f^\star\,\Delta\delta f\Big\} \ . \tag{3.49}$$

What do we do with this? Our goal from here on is to rewrite the integrand in the form $[\cdots\text{some stuff}\cdots]\delta f$ and then argue as follows: we have enormous freedom in our choice of δf, and no matter what we pick, we want the variation of the energy to be zero (at first order in δf). This can only work out if the $[\cdots\text{some stuff}\cdots]$ contribution identically vanishes. This should give us some expression from which we can determine the equilibrium surface shape $f^\star$.

Notice that in Eqn. (3.49) δf does not occur by itself, but with one or two derivatives. How do we get rid of them? We integrate by parts! But how do we do this if we have gradients and stuff? Here's how. First, observe the following identity, which simply is a product rule:

$$\boldsymbol{\nabla} \cdot (\boldsymbol{v} g) = (\boldsymbol{\nabla} \cdot \boldsymbol{v}) g + \boldsymbol{v} \cdot (\boldsymbol{\nabla} g) \ . \tag{3.50}$$

Solving this for either one of the terms on the right hand side, and reordering just a tiny bit, we get

$$\boldsymbol{v} \cdot (\boldsymbol{\nabla} g) = \boldsymbol{\nabla} \cdot (\boldsymbol{v} g) - (\boldsymbol{\nabla} \cdot \boldsymbol{v}) g \ , \tag{3.51a}$$

$$g(\boldsymbol{\nabla} \cdot \boldsymbol{v}) = \boldsymbol{\nabla} \cdot (g \boldsymbol{v}) - (\boldsymbol{\nabla} g) \cdot \boldsymbol{v} \ . \tag{3.51b}$$

This shows that a terms of the form "vector times gradient of scalar" and "scalar times divergence of vector" can be rewritten as the difference between two terms: a total divergence, and a term in which the derivative got suitably "swapped" from one factor to the other.

Obviously, the first term in the integral (3.49) is exactly of the first form, Eqn. (3.51a), and the second term in the integral (3.49) is exactly of the second form, Eqn. (3.51b).[7] We can therefore rewrite the two parts in the integral (3.49) as follows:

$$\sigma \, \boldsymbol{\nabla} f^\star \cdot \boldsymbol{\nabla} \delta f = \sigma \boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} f^\star \, \delta f) - \sigma (\boldsymbol{\nabla} \cdot \boldsymbol{\nabla} f^\star) \delta f \tag{3.52a}$$

$$\kappa \, \Delta f^\star \, \Delta \delta f = \kappa \boldsymbol{\nabla} \cdot (\Delta f^\star \, \boldsymbol{\nabla} \delta f) - \kappa (\boldsymbol{\nabla} \Delta f^\star \cdot \boldsymbol{\nabla} \delta f) \ . \tag{3.52b}$$

Our goal was to "peel off" the derivatives from the δf terms, possibly at the expense of creating extra divergence terms. And we succeeded with the tension term! But we did not yet succeed with the bending term, unsurprisingly, because in that case *two* derivatives were hitting δf. Indeed, if you look at the second term on the right hand side of Eqn. (3.52b), you find an occurrence of $\boldsymbol{\nabla} \delta f$ that we still need to work on. But of course we now know how to: we use the helper formula (3.51a) a second time:

$$-\kappa (\boldsymbol{\nabla} \Delta f^\star \cdot \boldsymbol{\nabla} \delta f) = -\kappa \boldsymbol{\nabla} \cdot (\boldsymbol{\nabla} \Delta f^\star \, \delta f) + \kappa (\boldsymbol{\nabla} \cdot \boldsymbol{\nabla} \Delta f^\star) \delta f \ . \tag{3.52c}$$

Putting everything together, we have now rewritten the terms occurring in the variation of the Helfrich energy so that they take either of the following two forms: a total divergence, or a term that has no derivative left on the variation δf.

[7] If you don't see the latter, recall that $\Delta h = \boldsymbol{\nabla} \cdot \boldsymbol{\nabla} h$, because *a Laplacian is a divergence (of a gradient).*

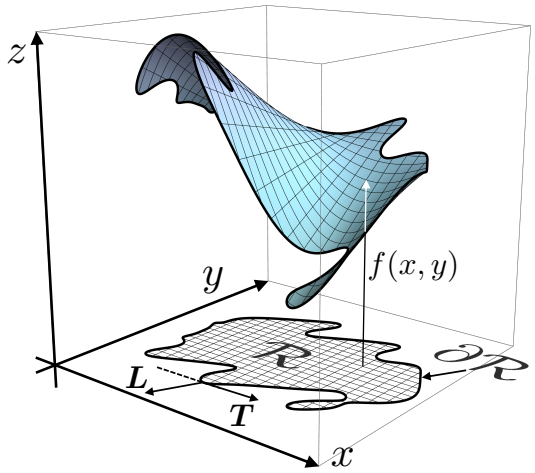


Figure 3.4 | A surface in Monge parametrization, $f(x, y)$, defined over some region $\mathcal{R}$ with boundary $\partial\mathcal{R}$ on the base plane. The boundary has a local normalized tangent vector $\boldsymbol{T}$ and an outward pointing unit normal $\boldsymbol{L}$.

What do we do with the divergence terms? That's the fun part: we use Gauss' theorem (in two dimensions) to push it to the boundary, using the following identity:

$$\int_{\mathcal{R}} \mathrm{d}x\,\mathrm{d}y\; \boldsymbol{\nabla}\cdot\boldsymbol{v}(x,y) = \oint_{\partial\mathcal{R}} \mathrm{d}s\; \boldsymbol{L}\cdot\boldsymbol{v}(x,y)\;. \tag{3.53}$$

In this expression, $\mathcal{R}$ is again the region on the base plane over which we integrate, $\partial\mathcal{R}$ is its boundary, $\mathrm{d}s$ is the line element on that boundary, and $\boldsymbol{L}$ is the outward-pointing unit vector that is normal to the boundary curve and lies in the xy base plane. Fig. 3.4 should help with an illustration.

If we insert all our rewritten terms from Eqns (3.52a), (3.52b), and (3.52c) into the variation Eqn. (3.49) we arrive at the expression

$$\delta E_{\text{bend}} = \int_{\mathcal{R}} \mathrm{d}x\,\mathrm{d}y \Big\{ \kappa\,\Delta\Delta f^\star - \sigma\,\Delta f^\star \Big\}\,\delta f \tag{3.54a}$$

$$+ \oint_{\partial\mathcal{R}} \mathrm{d}s\,\boldsymbol{L}\cdot\Big\{ \big[\sigma\,\boldsymbol{\nabla} f^\star - \kappa\boldsymbol{\nabla}\Delta f^\star\big]\delta f + \big[\kappa\,\Delta f^\star\big]\,\boldsymbol{\nabla}\delta f \Big\}\;. \tag{3.54b}$$

This is how the energy of our assumed optimal shape $f^\star(x, y)$ changes if we wiggle it a little bit out of shape with some smooth but otherwise arbitrary function $\delta f(x, y)$. Now remember: *if $f^\star$ really is the optimal shape, then δE_{bend} must vanish*! Since δE_{bend} consists of two contributions—one that "lives" in the interior $\mathcal{R}$ and another one that

lives at the boundary $\partial\mathcal{R}$—and we can vary these contributions independently, *the first and the second line of Eqn. (3.54) must both independently be zero.*

Let's start with the first line, Eqn. (3.54a). It is exactly of the form we hoped to achieve: an integral over $[\cdots\text{some stuff}\cdots]\delta f$. Since this integral has to vanish for any choice of δf, the $[\cdots\text{some stuff}\cdots]$ term must be zero. It is easy to see why: Assume it doesn't—meaning, there is some point (x_0, y_0) where the term in square brackets is nonzero. Without loss of generality, let's say it's positive. Because we will insist on our solution $f^\star(x, y)$ to be sufficiently smooth, we can be sure that the expression in square brackets must therefore also be positive in a small region $\mathcal{R}_0$ around (x_0, y_0). Let's now pick a particular variation $\delta f_0(x, y)$ which vanishes everywhere outside of $\mathcal{R}_0$ but smoothly rises to some positive value inside $\mathcal{R}_0$. With this choice of δf, our integral must be positive. This is a contradiction, and thus our assumption that we can find a point (x_0, y_0) where the term in square brackets does not vanish must be wrong. □

Ponder now what's actually inside that square brackets: a bunch of derivatives hitting our optimal solution $f^\star(x, y)$. The fact that this must be zero therefore amounts to a *differential equation* our optimal solution must satisfy—the so-called *Euler–Lagrange equation* of the variational problem:

$$\boxed{\kappa\,\Delta\Delta f^\star(x,y) - \sigma\,\Delta f^\star(x,y) \;=\; 0\;.} \tag{3.55}$$

This is the shape equation we must solve in order to find the optimal shape for our surface, the one that will minimize elastic costs due to bending or an externally applied tension. Three things are worth noticing here:

1. this is a *partial* differential equation in two variables—x and y;
2. it is of *fourth order* in derivatives of the shape function $f^\star(x, y)$;
3. it is *linear* in the shape function $f^\star(x, y)$.

It turns out that (1) and (2) hold more generally, even if we go to more fancy parametrizations. This is true because surfaces are two-dimensional, and so we must get partial differential equations in two variables—(u^1, u^2) in our more general parametrization; and the variation of a curvature-squared functional requires two integration by parts cycles to peel off the two derivatives from our shape variation, which will dump two more derivatives onto a curvature term, leading to a fourth order equation.

Lamentably, (3) is not generally true. It came out like this because we used the quadratic expansion of Monge gauge, not the actual Monge gauge with all its awkward nonlinearities. This means that, generally, the shape equation is a *nonlinear* fourth order partial differential equation. Such a beast is *exceptionally* hard to solve. And because of that, it is *really* nice if we can get away with the linearized version (3.55), for which an enormous arsenal of solution strategies exists.

It is instructive to slightly rewrite the differential equation (3.55) in a way that makes it more transparent what's going on. First of all, realize that the two parameters κ and σ differ in dimensions, and that their ratio is a squared length. Let's in fact give that length a symbol:

$$\lambda := \sqrt{\frac{\kappa}{\sigma}} \,. \tag{3.56}$$

Using this so-called "elasto-capillary length," and "factoring things out" a bit, we can write

$$\boxed{\left(\Delta - \lambda^{-2}\right) \Delta \, f^\star(x,y) \;=\; 0 \,.} \tag{3.57}$$

Any function for which $\Delta f^\star = 0$ will obviously solve this equation. Hence, harmonic functions (*i. e.*, eigenfunctions of the Laplacian with eigenvalue 0) are a subset of all solutions. Next, the two operators $\Delta - \lambda^{-2}$ and Δ clearly commute, and so we could have swapped them when writing the equation. But then solutions of $\Delta f^\star = \lambda^{-2} f^\star$ are *also* solutions to this equation, and those are eigenfunctions of the Laplacian with eigenvalue λ^{-2}. This in fact exhausts the possibilities, because we've arrived at a four-dimensional solution space. Miraculously, what might have *looked* like a fourth order differential equation can actually be factored into two second order differential equations, and that makes things a bit easier. So the upshot is: *the solutions to the linearized shape equation (3.55) or (3.57) are the eigenfunctions of the Laplacian with eigenvalues* 0 *and* λ^{-2}.

Speaking of solutions to this differential equation: we will need boundary conditions to pick the correct solution out of the four-dimensional solution space. Where do they come from? *They come from the second line, Eqn. (3.54b)*! This expression "lives" on the boundary, and we need to make it vanish, and to do so, some conditions must hold at that boundary—the *boundary conditions*.

Permit me to vent just a little bit: when I was a budding physicist I surely have seen many instances of variational problems, discussed in courses, lectures, textbooks, and whatnot. In virtually all cases the

instructor presenting the material was obviously vastly more interested in deriving the Euler–Lagrange equation, and so they short-circuited the discussion of the boundary terms that invariably arise from the integration by parts step, mumbling magic incantations like "we'll assume they vanish," or "if we push the boundaries to infinity, then these terms of course vanish"—or any other such rubbish. *Boundary terms do not vanish automatically.* You must *make* them vanish, and this is where the boundary conditions come from. *Where else would they come from*?

OK, stepping off my soap box, let's look at the boundary terms and see how, indeed, we can make them vanish. To slightly clean up our notation a little bit more, we first define the directional derivative

$$\nabla_\perp := \boldsymbol{L} \cdot \boldsymbol{\nabla} \ , \tag{3.58}$$

which is the derivative (in the base plane) along the outward pointing unit normal $\boldsymbol{L}$. Using this, we can ever so slightly simplify the boundary term as follows:

$$\delta E_{\text{bend}}\Big|_{\text{boundary}} = \oint_{\partial\mathcal{R}} \mathrm{d}s \left\{ \underbrace{\Big[\sigma\, \nabla_\perp f^\star - \kappa \nabla_\perp \Delta f^\star\Big] \delta f}_{\text{position}} + \underbrace{\Big[\kappa\, \Delta f^\star\Big]\, \nabla_\perp \delta f}_{\text{slope}} \right\} .$$

The integrand contains two contributions: one that is proportional to δf (*i. e.*, the variation of the *position* of the surface anywhere on the boundary $\partial\mathcal{R}$) and one that is proportional to $\nabla_\perp \delta f$ (*i. e.*, the variation of (basically) the *slope* of the surface anywhere on the boundary $\partial\mathcal{R}$—more precisely, the derivative along the local normal direction). Observe that at the edge of a surface we can clearly vary position and slope independently, so these two terms are in fact separate contributions that will give rise to *two separate boundary conditions*.

Let's begin with the first one. How can we make it vanish? This term is a product of two factors, and setting either one of them to zero will do. We could for instance demand that $\delta f = 0$—*but what does this mean*? It means that we do not permit the position of the surface to vary at the boundary. What could stop us? The physical situation might, if for whatever reason we happen to know that the surface must approach some specific position at the boundary. Say we know that at the boundary the surface must reach the value $f = 0$, that it's *pinned* there. If so, all the shapes we're trying out when testing out all possible variations must have in common that $f = 0$ at the boundary. But if we're not allowed to change the position at the boundary, then $\delta f = 0$, and then—hey presto!—this boundary term vanishes.

But not all physical situations enforce a specific position at the boundary of a surface. Picture a piece of paper lying on your desk in such a way that it partly dangles over the edge. What shape does the hanging portion take? Calculating it is exactly the type of bending energy minimization problem we should be able to tackle with the theory we're developing here.[8] What boundary condition should we prescribe at the paper edge that hangs over the table? We do not know how far the paper droops, so we have no way to prescribe this. It might droop more, it might droop less. We should try out how much drooping is best, energetically speaking. This means we want to vary the paper's position at the boundary, and then $\delta f \neq 0$ at this edge. To nevertheless wipe out the boundary term, we must demand that the term in square brackets multiplying the δf vanishes instead. *And that's our alternative boundary condition*! Sure, it looks more complicated, involving a first directional derivative of the shape, and another first directional derivative of the Laplacian of the shape. But so what, that stuff is zero, and it will help us to determine our integration constants. One thing is clear, though: this does not look like a boundary condition we would have easily guessed. So a big shout-out to those boundary terms that had the grace to pop out for free out of the functional variation.

We're only half way done, though. There is a second term, the "slope term," that also needs to be zero. But now we know how to think about it. This term is again the product of two factors. One of them, $\nabla_\perp \delta f$ is the derivative of the variation along the outward pointing normal. It vanishes if we have decided not to change that derivative at the edge, meaning, if for some reason the slope of the membrane at the edge is fixed. The proper lingo here is: "the membrane is clamped." It could for instance be adhering to a substrate of given orientation, and the slope of the substrate sets the slope of the membrane. Alternatively, if the slope of the membrane is not fixed, and so we should try out different slopes, then the factor that multiplies $\nabla_\perp \delta f$ must vanish at the edge, and that is $\Delta f^\star$.

In analogy to a similar case many of you are familiar with from electrostatics (*i. e.*, solving Poisson's equation $\Delta\phi(\boldsymbol{r}) = -\rho(\boldsymbol{r})/\varepsilon_0$), we can refer to the conditions that relate to the vanishing of δf or $\nabla_\perp \delta f$ as *Dirichlet boundary conditions* and those that hold when δf or $\nabla_\perp \delta f$ are free and so the prefactors in square brackets needed to vanish as *Neumann boundary conditions*.[9] Table 3.1 summarizes your options.

[8] Strictly, a piece of paper is a solid membrane—it has a shear resistance. That makes a crucial difference, *but not if we only bend along one direction.*

[9] Fun fact: Neumann boundary conditions are *not* named after the famous mathematician and physicist John von Neumann, but rather after the slightly less

	Dirichlet condition	Neumann condition
first condition	fix f	require $(\kappa\Delta - \sigma)\nabla_\perp f = 0$
second condition	fix $\nabla_\perp f$	require $\Delta f = 0$

Table 3.1 | Possible choices of boundary conditions for the differential equation (3.55). For each of the two rows *one* condition needs to be fulfilled, but it is not necessary that the same condition is picked over the entire boundary.

Everywhere on the boundary you must enforce two conditions, and for each of these you can choose either the Dirichlet or the Neumann variant. You can even mix them—say, on some boundary stretch Dirichlet for the δf condition, on some other stretch the Neumann conditions corresponding to the δf condition. It's all fine—as long as everywhere a choice is made for each of the two independent conditions.

Example: This is all very nice, but let's look at a worked-out specific case to see how to crank this machine.

Assume we want to calculate the shape of a membrane that smoothly covers a step-edge of height h_0 and touches the lower level a distance L away, as illustrated in Fig. 3.5. Since the membrane shape changes only in x-direction, we have a one-dimensional problem, *i. e.*, a function $f(x)$ to find, and the Laplacian Δ is simply equal to the second derivative $\mathrm{d}^2/\mathrm{d}x^2$. The shape equation (3.55) becomes $f''''(x) - f''(x)/\lambda^2 = 0$. Two independent eigenfunctions belonging to the eigenvalue 0 are 1 and x, and for the eigenvalue λ^{-2} we conveniently take $\cosh(x/\lambda)$ and $\sinh(x/\lambda)$. Defining the scaled variables $\tilde{x} := x/\lambda$ and $\ell := L/\lambda$, we can write the general solution as

$$f(x) = A + B\tilde{x} + C\,\cosh(\tilde{x}) + D\,\sinh(\tilde{x})\ , \tag{3.59}$$

where the integration constants $A \ldots D$ are determined by the four obvious boundary conditions

$$h_0 = f(0) = A + C\ , \tag{3.60a}$$
$$0 = \lambda\, f'(0) = B + D\ , \tag{3.60b}$$
$$0 = f(L) = A + B\ell + C\,\cosh(\ell) + D\,\sinh(\ell)\ , \tag{3.60c}$$
$$\text{and} \quad 0 = \lambda\, f'(L) = B + C\,\sinh(\ell) + D\,\cosh(\ell)\ . \tag{3.60d}$$

From Eqn. (3.60a) follows $A = h_0 - C$, and from Eqn. (3.60b) follows $B = -D$. Inserting this into the remaining two equations (3.60c) and

famous mathematician *Carl Neumann*, who was born 5/7/1832 in Königsberg (Prussia) and died 3/27/1925 in Leipzig.

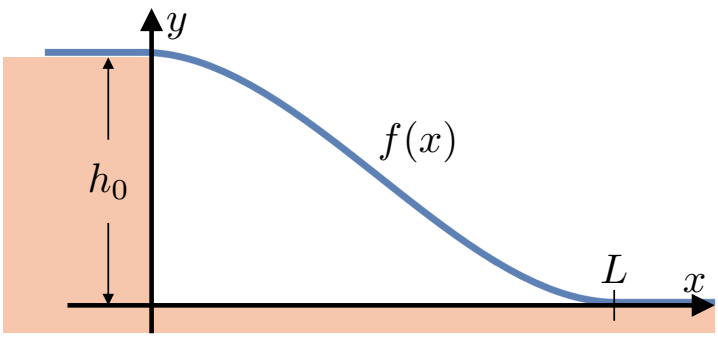


Figure 3.5 | A membrane traverses a step-edge of height h_0 and attaches to the lower lying substrate a distance L away from the step. What shape does it take?

(3.60d) yields a simple matrix equation for C and D,

$$\begin{pmatrix} \cosh(\ell)-1 & \sinh(\ell)-\ell \\ \sinh(\ell) & \cosh(\ell)-1 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} -h_0 \\ 0 \end{pmatrix}, \tag{3.61}$$

which can be readily solved by matrix inversion. We thus find the solution of our shape problem. It can be expressed in the following way—not fully simplified, but this way it's a bit more revealing:

$$\frac{f(x)}{h_0} = 1 - \frac{\big[\cosh(\ell)-1\big]\big[\cosh(\tilde{x})-1\big] - \sinh(\ell)\big[\sinh(\tilde{x})-\tilde{x}\big]}{\big[\cosh(\ell)-1\big]\big[\cosh(\ell)-1\big] - \sinh(\ell)\big[\sinh(\ell)-\ell\big]} \tag{3.62a}$$

$$\overset{\sigma\to 0}{=} 1 - 3\left(\frac{x}{L}\right)^2 + 2\left(\frac{x}{L}\right)^3 . \tag{3.62b}$$

For nonzero tension, $\sigma > 0$, there are two characteristic length scales in the problem: λ and L, and the shape looks slightly different depending on which of these two is the bigger one (*i. e.*, whether ℓ is small or large compared to 1), as the interested reader might want to check.

What determines L? In the simplest case the substrate becomes "sticky" a distance L away from the step and pins the membrane there. A more complicated situation arises when the substrate has a uniform adhesion energy w per area, and the membrane can *decide* at which distance L to detach. For small L much adhesion energy will be gained, but the membrane has to bend a lot. Conversely, if L is chosen very large, bending will be weak, but a lot of adhesion energy is sacrificed. At some optimal distance the energy is minimal. It can be shown[10] that this leads to another boundary condition—this time for the *moving*

[10] This is actually a practice problem in (Landau and Lifshitz, 1986), namely §12, prob. 6. But the question what type of boundary conditions ought to be applied

boundary L—that in the present situation reads $f''(L) = 1/\rho_c$, where $\rho_c = \sqrt{\kappa/2w}$ is the contact radius of curvature. Using Eqn. (3.62b) this results in the transcendental equation $\ell \coth \frac{\ell}{2} - 2 = h_0\rho_c/\lambda^2$ for ℓ, whose solution is easily determined numerically. For $\sigma \to 0$ (*i. e.* $\lambda \to \infty$) it can be solved exactly: $L = \sqrt{6h_0\rho_c}$.

3.3.5 Membrane fluctuations

I opened these notes by reminding you that the central energy scale of soft matter physics is k_BT. It sets the strength of the interactions that give rise to self assembly, and together with the (possibly emergent) characteristic length scales it sets the typical magnitude of moduli—hence "*soft*." It should therefore not be a surprise that thermal fluctuations also reign supreme in this field. They are often not just small corrections to the fluctuation-free energetic side of the story; they may rule the relevant physics.

In the context of membranes, my goal is to show you how to treat fluctuations in the continuum language we have developed so far, restricting to the case where this is actually doable. Like always in Statistical Thermodynamics, we will have to calculate thermal averages, which means that one way or another we need to sum over some sort of phase space. But what are the degrees of freedom of a continuum surface? This is not too hard when we consider small fluctuations around a simple reference state, but in general this can be quite tricky. I will therefore restrict to the simple case.

The problem we will now consider is of significant interest both for experiment and for simulation: what is the power spectrum of thermal undulations of a membrane that is on average flat? For ease of calculation it is convenient to consider this membrane to be subject to periodic boundary conditions—which is actually exactly right for the type of computer simulations commonly done. Since we will only be concerned with small fluctuations (and we can—and should—later check that this assumption is self-consistent), we will be "allowed" to work in linear Monge gauge.

Consider therefore a membrane above the square base plane patch $\square \equiv [0, L] \times [0, L]$, defined by the height function $f(x, y)$, as illustrated in Fig. 3.6. We will assume periodicity, *i. e.*, $f(x+nL, y+mL) = f(x, y)$

when membranes adhere to substrates (or other membranes), in the absence of any symmetry, is a really interesting one, and it can be solved rather elegantly using differential geometric techniques only slightly more refined than those we started to develop here (the material in (Deserno, 2015) is definitely sufficient). The reader will find this worked out in (Deserno et al., 2007).

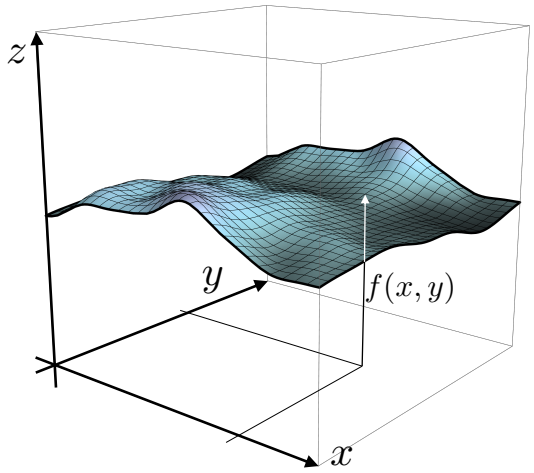


Figure 3.6 | Snapshot of a fluctuating membrane under periodic boundary conditions. The magnitude of fluctuations is exaggerated for visual clarity—typically, it is considerably smaller, as we will soon see.

for any integer values n and m. Given this, it seems natural to Fourier expand:[11]

$$f(\boldsymbol{r}) = \sum_{\boldsymbol{q}} \tilde{f}_{\boldsymbol{q}}\, \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{r}} \qquad \text{with} \quad \boldsymbol{q} = \frac{2\pi}{L}\begin{pmatrix} n_x \\ n_y \end{pmatrix} \quad \text{and} \quad (n_x, n_y) \in \mathbb{Z}^2 \ . \tag{3.63}$$

A few comments are in order

1. The permissible $\boldsymbol{q}$-vectors come from a discrete lattice, because the function we wish to represent is periodic with period L in both x- and y-direction.

2. At this point, theory does not force any obvious large-$|\boldsymbol{q}|$ ("ultraviolet") cutoff on us, since *theoretically* we can imagine the membrane wiggling on smaller and smaller scales. Yet, we know that in reality wavelengths shorter than the nanometer scale make no sense. We could decide to set the small length cutoff such that the total number of Fourier modes matches the total number of degrees of freedom (as counted at the *particle* level), but it turns out that the continuum approximation is not valid down to that scale. In practice, if we need to pick a large-$|\boldsymbol{q}|$ cutoff, we'd rather let it demarcate the edge of continuum-land.

[11] Of course, there are tons of opportunities for a basis expansion, even if we insist on periodicity. What makes the Fourier expansion special? I'm glad you asked. Hang in there, it will become clear very soon!

3. We use a *complex* expansion, since this is mathematically easier to handle. But of course the membrane itself needs to be real. You should verify that this forces the following condition on the Fourier coefficients:
$$\tilde{f}_{-\boldsymbol{q}} = \tilde{f}^*_{\boldsymbol{q}} \,. \tag{3.64}$$

4. Admittedly, mode counting is a bit of a drag in complex notation. We know that $\sin(x)$ and $\cos(x)$ are independent Fourier modes with the same frequency, but here we only have a single function $\mathrm{e}^{\mathrm{i}x}$. However, it is complex. More importantly the expansion coefficients for its amplitude are also complex, and thus have a real and an imaginary part. So two degrees of freedom per complex mode. Except, they also need to satisfy Eqn. (3.64). I will add a little clean-up note for those interested in it after we've done the calculation.

The energy we will work with is the Helfrich Hamiltonian in linear Monge gauge, as written in Eqn. (3.46). It contains integrals over $(\boldsymbol{\nabla} f)^2$ and $(\Delta f)^2$—let's see what happens to those after we insert the Fourier transform. Start with the first one:

$$\int_\square \mathrm{d}^2 r\, (\boldsymbol{\nabla} f)^2 = \int_\square \mathrm{d}^2 r \left[\boldsymbol{\nabla} \sum_{\boldsymbol{q}} \tilde{f}_{\boldsymbol{q}}\, \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{r}}\right] \cdot \left[\boldsymbol{\nabla} \sum_{\boldsymbol{q}'} \tilde{f}_{\boldsymbol{q}'}\, \mathrm{e}^{\mathrm{i}\boldsymbol{q}'\cdot\boldsymbol{r}}\right] \tag{3.65a}$$

$$= \int_\square \mathrm{d}^2 r \left[\sum_{\boldsymbol{q}} \mathrm{i}\boldsymbol{q} \tilde{f}_{\boldsymbol{q}}\, \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{r}}\right] \cdot \left[\sum_{\boldsymbol{q}'} \mathrm{i}\boldsymbol{q}' \tilde{f}_{\boldsymbol{q}'}\, \mathrm{e}^{\mathrm{i}\boldsymbol{q}'\cdot\boldsymbol{r}}\right] \tag{3.65b}$$

$$= \sum_{\boldsymbol{q},\boldsymbol{q}'} (-\boldsymbol{q}\cdot\boldsymbol{q}') \tilde{f}_{\boldsymbol{q}} \tilde{f}_{\boldsymbol{q}'} \int_\square \mathrm{d}^2 r\, \mathrm{e}^{\mathrm{i}(\boldsymbol{q}+\boldsymbol{q}')\cdot\boldsymbol{r}} \,. \tag{3.65c}$$

Ponder the integral: if $\boldsymbol{q} = -\boldsymbol{q}'$, then the exponent of the exponential vanishes, and the integral simply picks up the box area, yielding L^2. But if $\boldsymbol{q} \neq -\boldsymbol{q}'$, then we integrate some integer number of oscillations, which cancels to zero. So the integral is equal to $L^2\, \delta_{\boldsymbol{q},-\boldsymbol{q}'}$.

OK, maybe that's a bit too fast for some. Let's do it by hand:

$$\begin{aligned}
\int_\square \mathrm{d}^2 r\, \mathrm{e}^{\mathrm{i}(\boldsymbol{q}+\boldsymbol{q}')\cdot\boldsymbol{r}} &= \int_0^L \mathrm{d}x \int_0^L \mathrm{d}y\, \exp\left\{\frac{2\pi\mathrm{i}}{L}\begin{pmatrix} n_x + n'_x \\ n_y + n'_y \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}\right\} \\
&= L^2 \int_0^1 \mathrm{d}x' \int_0^1 \mathrm{d}y'\, \exp\left\{2\pi\mathrm{i}\begin{pmatrix} n_x + n'_x \\ n_y + n'_y \end{pmatrix} \cdot \begin{pmatrix} x' \\ y' \end{pmatrix}\right\} \\
&= L^2 \underbrace{\int_0^1 \mathrm{d}x'\, \mathrm{e}^{2\pi\mathrm{i}(n_x+n'_x)\,x}}_{\delta_{n_x,-n'_x}} \underbrace{\int_0^1 \mathrm{d}y'\, \mathrm{e}^{2\pi\mathrm{i}(n_y+n'_y)\,y}}_{\delta_{n_y,-n'_y}} \\
&= L^2\, \delta_{\boldsymbol{q},-\boldsymbol{q}'} \,.
\end{aligned} \tag{3.66}$$

In the second step we substituted $x' = x/L$ and $y' = y/L$, and the "underbraced identities" follow because when $n_x + n'_x$ is a nonzero integer n, then the integral covers exactly n oscillations of the complex exponential, which average to zero. □

Putting this result into Eqn. (3.67), we get

$$\int_{\square} \mathrm{d}^2 r \; (\boldsymbol{\nabla} f)^2 = L^2 \sum_{\boldsymbol{q},\boldsymbol{q}'} (-\boldsymbol{q}\cdot\boldsymbol{q}') \tilde{f}_{\boldsymbol{q}} \tilde{f}_{\boldsymbol{q}'} \, \delta_{\boldsymbol{q},-\boldsymbol{q}'} = L^2 \sum_{\boldsymbol{q}} q^2 |\tilde{f}_{\boldsymbol{q}}|^2 \; , \quad (3.67)$$

where in the last step we used $\tilde{f}_{-\boldsymbol{q}} = \tilde{f}^*_{\boldsymbol{q}}$ and thus $\tilde{f}_{\boldsymbol{q}} \tilde{f}^*_{\boldsymbol{q}} = |\tilde{f}_{\boldsymbol{q}}|^2$.

The calculation for the second $(\Delta f)^2$ term in Eqn. (3.46) proceeds identically. Putting things together, we now have the linearized Monge Helfrich energy expressed in Fourier space:

$$E_{\mathrm{bend}} = \frac{1}{2} L^2 \sum_{\boldsymbol{q}} |\tilde{f}_{\boldsymbol{q}}|^2 (\sigma q^2 + \kappa q^4) \; . \quad (3.68)$$

This is a pretty nice result! Why did this work so well? The key reason is some piece of magic that happened in the step from Eqn. (3.65a) to Eqn. (3.65b): $\boldsymbol{\nabla} \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{r}} = \mathrm{i}\boldsymbol{q}\, \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{r}}$ (or the equivalent magic that would happen during the rewriting of the $(\Delta f)^2$ term). We could replace the differential operator by a simple prefactor, *because the function $e^{i\boldsymbol{q}\cdot\boldsymbol{r}}$ is an eigenfunction of that differential operator*. And this explains why expanding our surface in Fourier modes is actually a good idea: it is an expansion in the eigenfunctions of the differential operators $\boldsymbol{\nabla}$ and Δ that feature in our energy expression! We often just "autopilot" into Fourier transforming things, because it often works; but the reason for that is that gradient operators and Laplacians are ubiquitous. If we ever come across a problem where a different operator characterizes the physics, we should remember why certain types of magic incantations work and maybe search for eigenfunctions of our new operator instead of Fourier expanding things as we "normally" do.

Eqn. (3.68) is also exceptionally nice from a Statistical Thermodynamics point of view: if we think of the set of Fourier modes, labeled by $\{\boldsymbol{q}\}$, as our degrees of freedom, then we have discovered that these transformed degrees of freedom *decouple*: none of the amplitudes $\tilde{f}_{\boldsymbol{q}}$ for some mode $\boldsymbol{q}$ couples to any other mode $\boldsymbol{q}'$! Since furthermore each amplitude enters quadratically ("$|\tilde{f}_{\boldsymbol{q}}|^2$"), each degree of freedom is just a simple harmonic oscillator with some $\boldsymbol{q}$-dependent spring constant. In other words, *a fluctuating Helfrich membrane is a sum of independent harmonic oscillators in Fourier space*!

We know a lot about the thermodynamics of simple harmonic oscillators; so much in fact that we do not even have to do a partition

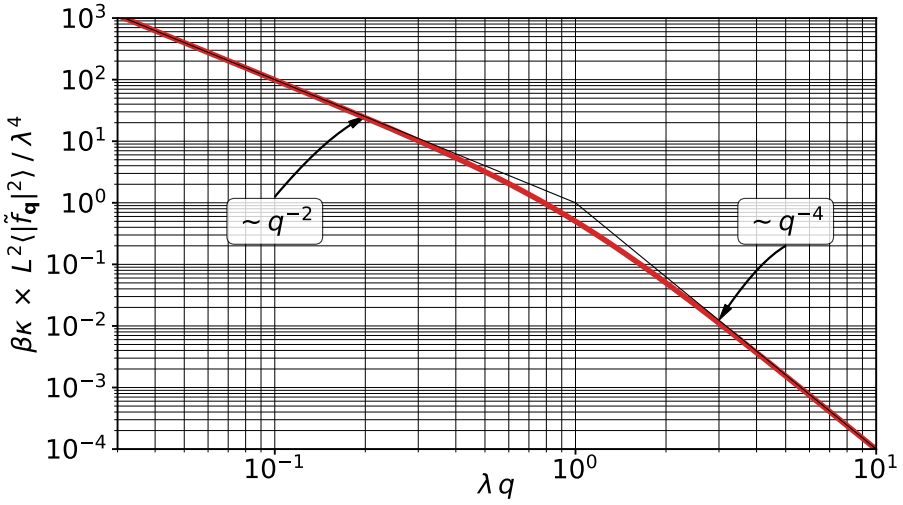


Figure 3.7 | Log-log plot of the power spectrum of membrane undulations, plotted in the rescaled fashion suggested in Eqn. (3.71). There is a crossover from tension-dominated fluctuations at large length scales (*i. e.*, small q) and bending dominated fluctuations at short scales. The crossover happens when the $q\lambda = 1$, where $\lambda = \sqrt{\kappa/\sigma}$ is the elasto-capillary length from Eqn. (3.56). In both experimental and computational applications it is usually advantageous to make the tension as small as possible, so that the region in q-space where the bending energy sets the undulation spectrum is as large as possible.

function to calculate what we are now interested in, namely, the power spectrum. All we need is a special case of the equipartition theorem, which says that the average energy of every quadratic degree of freedom is $\frac{1}{2}k_{\mathrm{B}}T$.[12] Using angular brackets $\langle\cdots\rangle$ to denote thermal averages, we find for each undulation mode

$$\frac{1}{2}k_{\mathrm{B}}T \overset{!}{=} \left\langle \frac{1}{2}L^2|\tilde{f}_{\boldsymbol{q}}|^2(\sigma q^2 + \kappa q^4)\right\rangle = \frac{1}{2}\underbrace{L^2(\sigma q^2 + \kappa q^4)}_{\text{“spring constant”}}\left\langle|\tilde{f}_{\boldsymbol{q}}|^2\right\rangle , \quad (3.69)$$

from which we get the classical Helfrich power spectrum

$$\boxed{\left\langle|\tilde{f}_{\boldsymbol{q}}|^2\right\rangle = \frac{k_{\mathrm{B}}T}{L^2(\sigma q^2 + \kappa q^4)} .} \quad (3.70)$$

Since the fluctuations are excited by temperature, it is not surprising that they scale with $k_{\mathrm{B}}T$. Observe also that they decrease with q,

[12] A simple example as a reminder: say the degree of freedom is x, and it enters the Hamiltonian in the form $\frac{1}{2}Ax^2$. Then the equipartition theorem says that $\frac{1}{2}k_{\mathrm{B}}T = \langle\frac{1}{2}Ax^2\rangle = \frac{1}{2}A\langle x^2\rangle$, from which we immediately get the mean squared amplitude $\langle x^2\rangle = k_{\mathrm{B}}T/A$.

meaning higher spatial frequencies are rapidly suppressed. The different scaling of tension and bending parts can be illustrated quite well by rescaling the wave number by the characteristic elasto-capillary length introduced in Eqn. (3.56), and we find

$$\beta\kappa \, \frac{L^2 \left\langle |\tilde{f}_{\boldsymbol{q}}|^2 \right\rangle}{\lambda^4} \;=\; \frac{1}{(\lambda q)^2 + (\lambda q)^4} \;, \tag{3.71}$$

where $\beta = 1/k_{\mathrm{B}}T$ is again the inverse thermal energy. This scaled version of the spectrum is shown in Fig. 3.7.

Sidenote for the curious about mode counting: There's a little bit of magic going on under the hood that is worthwhile bringing into the light. The constraint $\tilde{f}_{-\boldsymbol{q}} = \tilde{f}^*_{\boldsymbol{q}}$ (which we needed to ensured that $f(x, y) \in \mathbb{R}$) implies that the amplitudes $\tilde{f}_{\boldsymbol{q}}$ and $\tilde{f}_{-\boldsymbol{q}}$ are not independent degrees of freedom. In fact, they describe the *same* degree of freedom. Are we falling victim to a counting problem here?

We've glossed over this, but here's a way to do it properly: evidently only *half* the $\boldsymbol{q}$-vectors in $\boldsymbol{q}$-space describe independent modes—say, those with a positive q_y. Or more precisely, the set of $\boldsymbol{q}$-vectors that's the union of the sets $\{q_x > 0 \,\wedge\, q_y = 0\}$ and $\{q_y > 0\}$, illustrated in Fig. 3.8. Let's in a slight abuse of notation call this the set "$\boldsymbol{q} > \boldsymbol{0}$." The Hamiltonian in $\boldsymbol{q}$-space can then more properly be written as

$$\frac{1}{2} L^2 \sum_{\boldsymbol{q}} |\tilde{f}_{\boldsymbol{q}}|^2 (\sigma q^2 + \kappa q^4) \qquad \longrightarrow \qquad L^2 \sum_{\boldsymbol{q}>\boldsymbol{0}} |\tilde{f}_{\boldsymbol{q}}|^2 (\sigma q^2 + \kappa q^4) \;. \tag{3.72}$$

Notice that when we dropped half the $\boldsymbol{q}$-vectors, we also dropped the prefactor $\frac{1}{2}$.

Something else is going on, though: $\tilde{f}_{\boldsymbol{q}}$ isn't just a single degree of freedom. *It's two degrees of freedom*, stored in the real and imaginary part of the complex amplitude. All this trouble arises from our use of a convenient but subtle complex notation. Had we instead decided to stay real, we could have expanded the shape into sine- and cosine-modes, showing that for every wave vector there are *two* independent modes.

Let us write $R_{\boldsymbol{q}} := \Re[\tilde{f}_{\boldsymbol{q}}]$ for the real part and $I_{\boldsymbol{q}} := \Im[\tilde{f}_{\boldsymbol{q}}]$ for the imaginary part of our complex amplitude $\tilde{f}_{\boldsymbol{q}}$. The Hamiltonian can then be written as

$$E[\tilde{f}_{\boldsymbol{q}}] \;=\; L^2 \sum_{\boldsymbol{q}>\boldsymbol{0}} \left(R_{\boldsymbol{q}}^2 + I_{\boldsymbol{q}}^2\right)(\sigma q^2 + \kappa q^4) \;. \tag{3.73}$$

For each $\boldsymbol{q} > \boldsymbol{0}$ we then get two independent quadratic modes, and by the equipartition theorem each has a thermal average of $\frac{1}{2}k_{\mathrm{B}}T$. This leads to

$$\langle R_{\boldsymbol{q}}^2 \rangle = \frac{k_{\mathrm{B}}T}{2\,L^2(\sigma q^2 + \kappa q^4)} \qquad \text{and} \qquad \langle I_{\boldsymbol{q}}^2 \rangle = \frac{k_{\mathrm{B}}T}{2\,L^2(\sigma q^2 + \kappa q^4)} \;. \tag{3.74}$$

Pay attention to the crucial difference with respect to Eqn. (3.70), though: there's an extra factor of 2 in the denominator on the right hand sides, because the prefactor $\frac{1}{2}$ in the expectation value no longer cancels with a matching prefactor $\frac{1}{2}$ in the Hamiltonian—we got rid of the latter when we realized we should not overcount the modes.

Let's now come back to the complex amplitude, and discover that after unearthing two forgotten factors of 2, both of them conveniently cancel:

$$\langle |\tilde{f}_{\boldsymbol{q}}|^2 \rangle = \langle R_{\boldsymbol{q}}^2 \rangle + \langle I_{\boldsymbol{q}}^2 \rangle = \frac{k_{\mathrm{B}}T}{2\,L^2(\sigma q^2 + \kappa q^4)} + \frac{k_{\mathrm{B}}T}{2\,L^2(\sigma q^2 + \kappa q^4)} = \frac{k_{\mathrm{B}}T}{L^2(\sigma q^2 + \kappa q^4)} \;, \tag{3.75}$$

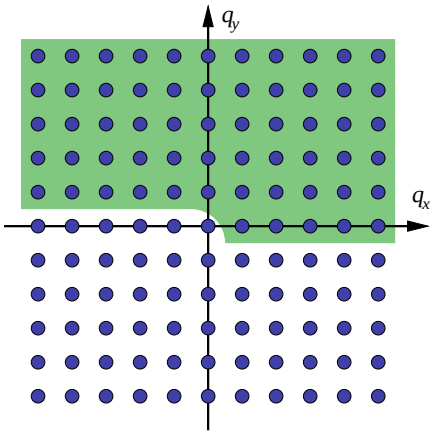


Figure 3.8 | The blue dots represent a set of discrete wave vectors $\boldsymbol{q} = (q_x, q_y)$ in two-dimensional Fourier space. Since a real functions $f(\boldsymbol{r})$ has Fournier components that satisfy $\tilde{f}_{-\boldsymbol{q}} = \tilde{f}^*_{\boldsymbol{q}}$, the wave vectors $\boldsymbol{q}$ and $-\boldsymbol{q}$ do not label independent modes, and so it suffices to pick one of them. The green shaded region is a possible choice for how to do that: pick the union of the sets $\{q_x > 0 \ \wedge \ q_y = 0\}$ and $\{q_y > 0\}$. Observe that we exclude $\boldsymbol{q} = \boldsymbol{0}$, since the average height of the surface is fixed and not a degree of freedom. We will sloppily denote the green region as "$\boldsymbol{q} > \boldsymbol{0}$".

and so *Eqn. (3.70) is correct after all*!

This all boils down to the following: If we use a complex notation, then the real and the imaginary part of the amplitude are independent degrees of freedom. But on the other hand, we must take the condition $\tilde{f}_{-\boldsymbol{q}} = \tilde{f}^*_{\boldsymbol{q}}$ seriously and only sum over half of $\boldsymbol{q}$-space. It now turns out that both from a mode-counting and from a final-prefactor point of view, we get the same answer if we stick with a sum over the entire $\boldsymbol{q}$-space and *pretend* that the complex amplitude $\tilde{f}_{\boldsymbol{q}}$ really only describes a single degree of freedom. We have just explicitly seen that this comes down to the same thing—*at least for the particular calculation we did here*. Care should be taken once other types of calculations are done, just to be sure.

So we now know the (mean squared) amplitude as a function of wavenumber in Fourier space. What is the mean squared amplitude in *real* space, $\langle f^2(\boldsymbol{r})\rangle$? Due to translation symmetry it cannot depend on position, but it turns out to be convenient for the calculation to nevertheless write the result as an average over the box:

$$\langle f^2\rangle = \frac{1}{L^2}\int_\square \mathrm{d}^2 r\, \langle f^2(\boldsymbol{r})\rangle = \frac{1}{L^2}\int_\square \mathrm{d}^2 r\, \Big\langle \Big(\sum_{\boldsymbol{q}} \tilde{f}_{\boldsymbol{q}}\, \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{r}}\Big)\Big(\sum_{\boldsymbol{q}'} \tilde{f}_{\boldsymbol{q}'}\, \mathrm{e}^{\mathrm{i}\boldsymbol{q}'\cdot\boldsymbol{r}}\Big)\Big\rangle$$

$$= \sum_{\boldsymbol{q},\boldsymbol{q}'} \langle \tilde{f}_{\boldsymbol{q}} \tilde{f}_{\boldsymbol{q}'}\rangle \underbrace{\frac{1}{L^2}\int_\square \mathrm{d}^2 r\, \mathrm{e}^{\mathrm{i}(\boldsymbol{q}+\boldsymbol{q}')\cdot\boldsymbol{r}}}_{\text{again: } =\delta_{\boldsymbol{q},-\boldsymbol{q}'}} = \sum_{\boldsymbol{q}} \langle |\tilde{f}_{\boldsymbol{q}}|^2\rangle = \sum_{\boldsymbol{q}} \frac{k_\mathrm{B}T}{L^2(\sigma q^2 + \kappa q^4)}\ . \tag{3.76}$$

The pleasantly clean formula arising as the second-to-last step is maybe

not quite unexpected: it's high-dimensional Pythagoras!

How do we evaluate such a sum? Difficult. A quite common workaround is to replace the sum by an integral, using effectively a smeared-out density of states in Fourier space: $D(\boldsymbol{q}) = (L/2\pi)^2$. Furthermore introducing a minimum and a maximum wave number, we get

$$\begin{aligned}\langle f^2\rangle &\approx \int_{q_\mathrm{min}}^{q_\mathrm{max}} \mathrm{d}q\; 2\pi q \left(\frac{L}{2\pi}\right)^2 \frac{k_\mathrm{B}T}{L^2(\sigma q^2+\kappa q^4)} \\ &= \frac{k_\mathrm{B}T}{2\pi\sigma}\int_{q_\mathrm{min}}^{q_\mathrm{max}} \frac{\mathrm{d}q\; q}{q^2+\lambda^2 q^4} \;=\; \frac{k_\mathrm{B}T}{2\pi\sigma}\int_{\lambda q_\mathrm{min}}^{\lambda q_\mathrm{max}} \frac{\mathrm{d}x}{x+x^3} \\ &= \frac{k_\mathrm{B}T}{2\pi\sigma}\left[\log\frac{x}{\sqrt{1+x^2}}\right]_{\lambda q_\mathrm{min}}^{\lambda q_\mathrm{max}} . \end{aligned} \tag{3.77}$$

An easier special case is the limit of vanishing tension, in which $\lambda \to \infty$, and so we need to evaluate the term in the square brackets at $x \to \infty$, where it approaches $-1/2x^2$. Setting $q_\mathrm{max} \to \infty$ will make that contribution vanish entirely, and setting $q_\mathrm{min} \to 2\pi/L$ yields

$$\langle f^2\rangle \xrightarrow{\sigma\to 0} \frac{1}{16\pi^3}\times\frac{k_\mathrm{B}T}{\kappa}\times L^2 \approx 0.002\frac{k_\mathrm{B}T}{\kappa}\times L^2 , \tag{3.78}$$

showing that the root mean square surface roughness $\langle f^2\rangle^{1/2}$ scales with system size:

$$\frac{\langle f^2\rangle^{1/2}}{L} \approx \sqrt{0.002\,\frac{k_\mathrm{B}T}{\kappa}} \approx 1\% , \tag{3.79}$$

where in the last step we assumed typical values of κ in the range of $20\,k_\mathrm{B}T \ldots 30\,k_\mathrm{B}T$. This finally explains why the caption in Fig. 3.6 warned you that the undulations are a bit overdone.

Incidentally, the zero-tension case can also be done numerically, meaning, without the need to approximate the sum with an integral. Since $q_i = \frac{2\pi}{L}n_i$, we find

$$\begin{aligned}\langle f^2\rangle_{\mathrm{exact},\sigma=0} &= \frac{k_\mathrm{B}T}{L^2\kappa}\sum_{\boldsymbol{q}}\frac{1}{q^4} = \frac{k_\mathrm{B}T}{L^2\kappa}\left(\frac{L}{2\pi}\right)^4 {\sum_{n_x,n_y}}' \frac{1}{(n_x^2+n_y^2)^2} \\ &= \frac{k_\mathrm{B}T}{L^2\kappa}\left(\frac{L}{2\pi}\right)^4 \underbrace{\sum_{n_x=1}^{\infty}\sum_{n_y=0}^{\infty}\frac{4}{(n_x^2+n_y^2)^2}}_{\approx 6.026812} \\ &= \frac{k_\mathrm{B}T}{16\pi^3\kappa}L^2 \times \frac{6.026812}{\pi} = \langle f^2\rangle_\mathrm{continuum}\times 1.92 , \end{aligned} \tag{3.80}$$

showing that the exact value is about twice as big as the approximate one coming from the $\sum \to \int$ approximation.[13] In other words, the rms surface roughness of a typical lipid membrane under vanishing lateral tension is about 2% of its lateral dimension.

3.4 Things to think about

1. Consider a smooth function $f(x)$. Given two close points x and $x + \mathrm{d}x$, as well as their two associated function values $f(x)$ and $f(x+\mathrm{d}x)$, find a simple heuristic argument for calculating the arc length $\mathrm{d}s$ which the function covers between these two points, and from there write down an expression that mirrors Eqn. (3.1) in this simpler one-dimensional case!

2. Consider again a smooth function $f(x)$, and construct the *local normal* at some point $(x_0, f(x_0))$. Now also construct the normal at the slightly displaced point $(x_0 + \delta x, f(x_0 + \delta x))$. These two lines (generally) intersect at some point $(x_\mathrm{m}, y_\mathrm{m})$. In the limit $\delta x \to 0$ this point becomes the center of a circle that, if it touches the curve at the original point $(x_0, f(x_0))$, does not merely match the *slope* but also the *curvature*, and indeed the inverse radius of that circle is called the curvature of the curve at that point. Use this reasoning to calculate the curvature of $f(x)$ at a given point. Compare against the exact 2d result from Eqn. (3.12a).

3. The principal curvatures for the two surfaces described by the functions f_1 and f_2 from Eqn. (3.6) turned out to be identical, and the two principal directions were rotated by 45°. I claimed this happend because the *entire surfaces* are identical, just rotated with respect to one another by 45°. Show that this is true.

4. The upstairs-downstairs business with coordinates seems a bit odd. This problem strives to demystify this situation by explaining that two types of single-index objects exist which are fundamentally different in the way they transform under coordinate changes. And while this sounds like a big nuisance, this is re-

[13] Totally irrelevant fun fact: the lattice sum can actually be done *analytically*. The value 6.026812… is equal to $4\zeta(2)\beta(2)$ (where $\zeta(z)$ is Riemann's Zeta-function and $\beta(z)$ is Dirichlet's Beta function), which further evaluates to $\frac{2}{3}\pi^2 G$, where G is Catalan's constant. Who knew.[†]

[†] The internet did: https://math.stackexchange.com/questions/197496/series-involving-catalan-and-zeta.

ally awesome, because it means we can construct *scalars* that are completely independent of the choice of coordinates.

a) Say we have chosen a set of local coordinates $\{u^i\}$. Now we get second thoughts and rather take a different set, $\{\overline{u}^j\}$. To be able to translate back and forth between these choices, we must require there to be a relation of the form $\overline{u}^j = \overline{u}^j(u^i)$ (*i. e.*, the new coordinates are functions of the old ones), and also an inverse, $u^i = u^i(\overline{u}^j)$ (*i. e.*, the old coordinates can also be written as functions of the new ones). Moreover, these functions have to be sufficiently smooth, so that higher derivatives, if needed, also translate properly. (Technically, they are *diffeomorphisms*.) Now, these functions can be scarily nonlinear, but at any given point on our surface, we can expand these coordinates and recognize that the *local* relation is linear. To see this, simply look at the differential:

$$\mathrm{d}\overline{u}^j = \frac{\partial \overline{u}^j}{\partial u^i}\mathrm{d}u^i = M^j_i \mathrm{d}u^i \ . \tag{3.81}$$

Now here comes the fun bit: the differential $\mathrm{d}u^i$ is an object with an upstairs index, and it happens to transform from old to new coordinates with a matrix that describes the derivative of the new coordinates with respect to the old ones. But check out the tangent vectors $\boldsymbol{e}_i$. Show that when we transform those, *we get the inverse transformation behavior*, *i. e.*, a transformation with the *inverse* matrix: the derivative of the old coordinates with respect to the new ones!

b) Let's now say we have an object that arises as the contraction of one upstairs index with one downstairs index. Something like $U^i V_i$. Show that if we change our coordinates, such that $U^i \to \overline{U}^i$ and $V_i \to \overline{V}_i$, then the contracted object $\overline{U}^i \overline{V}_i$ has numerically the same value as the old one, *i. e.*, *it remains invariant under a change of coordinates*. In other words, it describes a piece of geometry that is finally free of the idiosyncrasies of a specific parametrization.
Hint: use what you have just discovered about the transformation matrix for objects with an upstairs vs. *a downstairs index. Observe in particular that the entire magic depends on the existence of the two transformation laws, and that it is hence* essential *that when we create these index contractions via the Einstein summation, one index is upstairs and the other one is downstairs. For instance, an object*

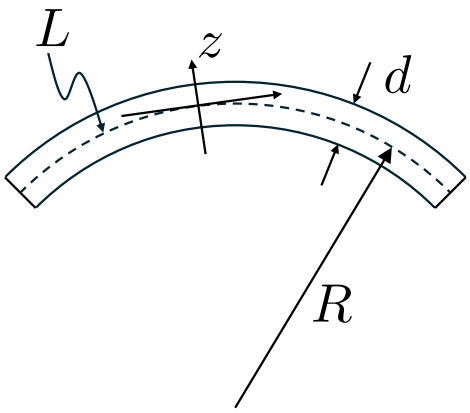


Figure 3.9 | Illustration of a thin plate bent along the "L-direction" into a circular arc of radius R. Set up a local coordinate system that measures the radial displacement from the mid-surface of the plate, but its origin lies at that mid-surface.

such as $\sum_i U^i V^i$ *would* not *be invariant under coordinate transformations.*

5. Let's get some practice with the more general $\boldsymbol{X}(u^1, u^2)$ parametrization!

   a) Recall that equation (3.17) is Monge gauge, just in the more fancy language. Use the definitions you find in Sec. 3.2.2 to work out the tangent vectors $\boldsymbol{e}_i$, metric g_{ij}, inverse metric g^{ij}, metric determinant g, normal vector $\boldsymbol{n}$, curvature tensor K_{ij} (which will look nice) and the shape operator K_i^j (which will unfortunately look less nice). By patiently working out the derivatives in Eqn. (3.12a), show that this boils down to the same expression as the trace of the matrix K_i^j you just evaluated.

   b) Same game, but now for the spherical surface in Eqn. (3.18). Find $\boldsymbol{e}_i$, and g_{ij}. Observe that g_{ij} is diagonal and hence spherical coordinates are orthogonal. Calculate g and recognize $\sqrt{g}$ as the spherical Jacobian. Explicitly calculate $\boldsymbol{n}$ and see that, as hopefully expected, it is equal to $\boldsymbol{X}/|\boldsymbol{X}|$. Finally, calculate g^{ij}, K_{ij}, and K_i^j. When the dust has settled, you should find that $K_i^j = \delta_i^j/r$, showing that the two principal curvatures are identical and always equal to $1/r$, which we would of course expect for a sphere of radius r!

6. Consider a thin elastic solid plate of length L, width w, and thickness d. Imagine bending it, as illustrated in Fig. 3.9, such that its

mid-plane now assumes a constant curvature of radius R. This will cost bending energy. In this problem our goal is to work out how this *bending* energy is related to the *stretching* energy of little volumes inside the plate because, clearly, pieces of the plate on the "outer side" of the mid-plane are stretched, while pieces on the "inner side" are compressed.

a) Let's set up a radial coordinate system such that $z = 0$ sits at the plate's mid-plane. Slice up the plate into lots of thin concentric shells of thickness $\mathrm{d}z$ and let us make two critical assumptions: first, the mid-plane does *not* change its length along the "L-direction"; second, stretching any one of these thin slices will not change its thickness or width.[14] Given that, what is the length change of a thin slice that is a distance z away from the mid-plane?

b) If the resulting volume change of any such infinitesimal slice costs an infinitesimal elastic energy of the form

$$\mathrm{d}E = \frac{1}{2} Y \frac{(\mathrm{d}V - \mathrm{d}V_0)^2}{\mathrm{d}V_0} \ , \tag{3.82}$$

where Y is the so-called "Young modulus" (this is quite analogous to Eqn. (3.40)), calculate the total elastic deformation energy of the bent plate by integrating up the infinitesimal costs of all slices.

c) Show that your result from (b) can be written as a bending energy density, $e = \frac{1}{2}\kappa(1/R)^2$, where the bending constant is given by

$$\kappa(\text{thin plate}) = \frac{Y d^3}{12} = \frac{K_A d^2}{12} \ , \tag{3.83}$$

where $K_A = Y d$ simply defines a 2d stretching modulus from the 3d Young modulus Y.

d) If you picture a lipid bilayer of thickness d as two elastic plates of thickness $d/2$ and use the bending energy results you've derived so far for these individual leaflets, show that the bending constant κ of a bilayer relates to the bilayer stretching constant K_A in the following way:

$$\kappa(\text{bilayer}) = \frac{K_A \, d^2}{48} \ . \tag{3.84}$$

[14] Technically, this means we assume that the *Poisson ratio* of the material is zero.

The number "48" in the denominator is not quantitatively correct in real life. People have derived alternatives based on better models (including Poisson ratio corrections, accounting for the polymeric nature of lipid tails, *etc.*). Overall, my feeling is that this argument is too crude to be trusted quantitatively, and efforts to make it more precise are maybe misguided; but it gives you a good order-of-magnitude idea.

7. Let's think about the surface of water. It has a surface tension σ, and we already know how to account for this. But water also has a density ρ, and raising or lowering some small packet of water above or below the resting level $f = 0$ under a gravitational acceleration g will cost energy (at sufficiently large scales). Convince me that within linear Monge gauge this can be accounted for by the energy

$$E = \int \mathrm{d}^2 r \left\{ \frac{1}{2}\sigma \big(\boldsymbol{\nabla} f\big)^2 + \frac{1}{2}\rho g f^2 \right\} = \frac{\sigma}{2} \int \mathrm{d}^2 r \Big\{ \big(\boldsymbol{\nabla} f\big)^2 + \big(f/\ell_\mathrm{c}\big)^2 \Big\} \,, \tag{3.85}$$

where $\ell_\mathrm{c} = \sqrt{\sigma/\rho g}$ is the so-called *capillary length*.

   a) What is the Euler-Lagrange equation satisfied by $f(\boldsymbol{r})$? What are its solutions in one dimension (*i. e.*, for functions $f(x)$)?

   b) Calculate the numerical value of ℓ_c for water and convince yourself that this, combined with your finding from part (a), makes sense in light of everyday experience.

   c) In analogy to Eqn. (3.70), calculate the power spectrum $\langle |f_{\boldsymbol{q}}|^2 \rangle$ for the thermal fluctuations of a water surface.

   d) Transform your result into real space and calculate the root-mean-square roughness $\langle f^2 \rangle^{1/2}$ of such a water surface. You may use a continuum approximation for the sum over $\boldsymbol{q}$-modes, but you will need to make some reasonable assumptions about the largest and smallest q-value.

8. Consider a rectangular paper strip of width w lying on a table such that a piece of length L reaches over the edge and sags under its own weight—see Fig. 3.10. Our goal in this problem is to calculate the shape of that strip in the limit of small sag, predict how much it sags, and from there determine the paper's bending rigidity.

   a) If κ is the bending rigidity of the paper, ρ its mass per unit area, and g is the gravitational acceleration, argue that the

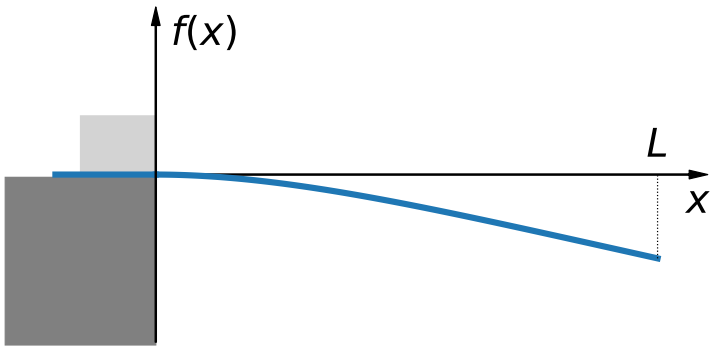


Figure 3.10 | Illustration of a piece of paper extending over the edge of a table, "clamped down" such that it initially extends horizontally over the edge but then begins sagging under its own weight.

energy of a paper described by the shape $f(x)$ is given by

$$E = w \int_0^L \mathrm{d}x \, \left\{ \frac{1}{2}\kappa \big[f''(x)\big]^2 + \rho g f(x) \right\} \, . \tag{3.86}$$

There are a few approximations going on here—list them carefully and argue why they are permitted in the limit of small sag. (What in fact *is* "small sag" in the first place?

b) Perform the functional variation and show that the differential equation from which we will get $f(x)$ is

$$f''''(x) + \ell^{-3} = 0 \tag{3.87}$$

with some characteristic length ℓ. What is ℓ in terms of our given parameters?

c) What are the boundary conditions we need to apply at the two ends? The ones at $x = 0$ are fairly obvious, but the ones at $x = L$ are not, and you need to revisit our more careful thinking from Sec. 3.3.4.

d) Now solve the differential equation—finding, as usual, a general solution and a particular solution. Show that in terms of the scaled variables $\tilde{x} = x/L$, $\tilde{\ell} = \ell/L$, and $\tilde{f} = f/L$, the solution can be written as the *universal shape*

$$\tilde{f}(\tilde{x}) = -\frac{\tilde{x}^4 - 4\tilde{x}^3 + 6\tilde{x}^2}{24\,\tilde{\ell}^3} \, . \tag{3.88}$$

e) Calculate the sag $s = |f(L)|$ as a function of all parameters, and solve this equation to determine κ in terms of parameters

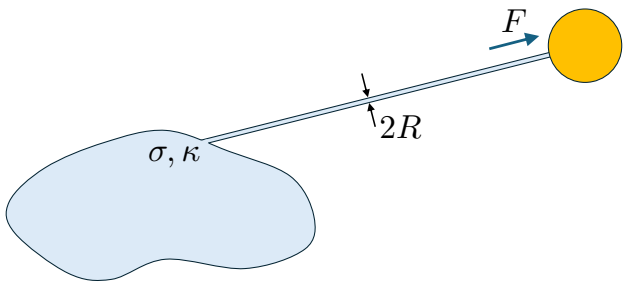


Figure 3.11 | An experimentalist pulls a thin long tether of radius R out of a cellular membrane; holding it requires the force F. Suitably set up, such an experiment can be used to measure the tension σ in the membrane, but a lot more is possible when you pull ingenious extra tricks.

in the problem, as well as s. Now do a literal "table top" experiment: measure s and determine κ for some paper.

f) Making some reasonable estimate about the thickness of your piece of paper, and using Eqn. (3.83), estimate its stretching modulus K_A. Does its magnitude strike you as "large"? (Well, it has dimensions, so what would be a good point of comparison?)

9. A famous biophysical measurement on cells is to determine their surface tension. This works by attaching a microbead to them (somehow...) and then pull that bead away (many microns) in such a way that we can measure the force—see Fig. 3.11. Let us assume we can describe the cell's biomembrane with the Helfrich model, meaning we ascribe a tension σ and a bending rigidity κ to it. We now wish to learn how these parameters relate to one another, and how we can use that to measure interesting things.

   a) How much energy E is needed to pull a tether of radius R and length L out of a cell?
   (Hint: the calculation will show, and experimentalists know very well, that $R \ll L$, and so you do not need to worry about the ends of the tether.)

   b) The experimentalist controls L, but not R. The tether radius will adjust such that the energy of the whole tether is minimal. From this, find an equation that gives R in terms of σ and κ. Using typical values for those that I have hidden in these notes, calculate typical R values. Could you optically resolve them?

   c) The force to hold the tether is the derivative of the energy

with respect to the length (at constant value of R). Show that it is given by

$$F = 2\pi\sqrt{2\kappa\sigma}\ . \tag{3.89}$$

That means: if we know F and κ, we can determine σ without having to know R. What are typical values for F? Google and convince yourself that this is exactly the right range for optical tweezers.

4 Membrane asymmetry

As a relatively straightforward (but also relatively recent) application to the elasticity considerations we have explored in Sec. 3, I would now like to explore one specific aspect of biomembranes which—at least at first sight—seems to have little to do with elasticity and more with individual lipids. But we will quickly see how it rather strongly affects almost every aspect of a membrane's physical behavior, including geometry and elasticity: membrane asymmetry. I have opined on some aspects of this topic in a recent short review (Deserno, 2024), which addresses a good fraction of this chapter's content (but more tersely).

4.1 Background: asymmetry preliminaries

4.1.1 Discovery and simple facts

Many practicioners develop a sense of history for the field they are working in. The urge to regale others with the origins of present day thinking, and certainly with anectodes from the days of yore, definitely increases with age. (I'm sorry if I'm telling on myself here.) At any rate: biomembrane history is fascinating, and it is at times shocking how much of the foundational work is also surprisingly old. For instance, the "Helfrich Hamiltonian," which we developed in the last chapter, and which really underlies almost anything in continuum membrane theory, is more than 50 years old (Helfrich, 1973). Even one year older is the so-called "fluid mosaic model" (Singer and Nicolson, 1972), which argued that experimental evidence points towards what we today take as obvious: that a biomembrane is a lipid bilayer with inserted transmembrane and attached peripheral proteins.[1]

Maybe even more remarkable: just 12 days after the fluid mosaic model got published, another important paper saw the light of day, in

[1] This looks slightly less mundane if one recalls what types of models were the competing view at the time, for instance the one by Danielli and Davson (1935), later refined by Robertson (1959), which imagined a pristine lipid bilayer sandwiched between two protein layers, or the model by Benson (1966), in which proteins "soak up" lipids to form lipoprotein condensates, and these then laterally aggregate into membranes.

which Bretscher (1972) suggested that the lipid content in human erythrocytes is unevenly distributed between the inner ("cytosolic") and the outer ("exoplasmic") leaflet. It was well appreciated that biomembranes contain many types of lipids, but these were assumed to partition evenly between the two leaflets. Bretscher instead argued that the majority of what's called phosphatidylcholine and sphingomylelin lipids reside on the outside, while the majority of phosphatidylethanolamine and phosphatidylserine (a negatively charged lipid) reside on the inside. A year later, Verkleij et al. (1973) put some tentative numbers on the extent of this asymmetry. Three years later similar numbers were published for human platelets (Schick et al., 1976), and before the end of the decade for fully nucleated cells (mouse LM fibroblasts) (Sandra and Pagano, 1978).[2]

Bretscher's basic conjecture, first quantified by Verkleij *et al.*, has held up remarkably well, as much more modern work shows: Lorent et al. (2020) have recently revisited the asymmetry of the human red blood cell and—while adding an incredible amount of detail—have confirmed all the basic points. Importantly, they also added a (technically not very arduous) bioinformatics study in which they looked at all transmembrane proteins held in place by a single alpha-helical anchor and showed that the area of the inner leaflet portion of that helical anchor is statistically bigger than that of the outer portion. Why is this interesting? Because they repeated this analysis for several organisms all across the phylogenetic tree—fungi, plants, insects, worms, reptiles, and amphibia—and found the same imbalance. This strongly suggests that membrane asymmetry is evolutionarily conserved across all of eukarya. *Something* about it seems to be *really* important to all those eukaryotes. Wouldn't it be nice to know *what*?

4.1.2 Passive flip-flop and active flippases

How is lipid membrane asymmetry even possible, you might ask? We have learned that biomembranes are self-assembled structures in which lipids are not held in place by chemical bonds. A lipid that sits in one leaflet might well transition into the other one—a process slightly facetiously called "flip-flop." Nothing prevents this *in principle*, except

[2] It is worth reminding that as far as lipidomics is concerned, human red blood cells are far easier to work with than most other cells, because the only lipid membrane of a red blood cells is in fact the plasma membrane. Trying to learn anything about the plasma membrane composition of cells which have a rich endomembrane system seriously limits the kind of experiments you can do. Anything that would for instance lyse the cell risks mixing up plasma membrane content with any of the other membranes.

it would involve intermediate states that have a somewhat larger free energy (for instance because a hydrophilic head group will have to pass through the hydrophobic bilayer interior, which results in a free energy penalty that will presumably affect the rate with which such processes happen.

Let's look at this. Say, there are $N_\pm$ lipids in the $+$ or $-$ leaflet, and a lipid in the $\pm$ leaflet can flip out of it into the other one with a rate $r_\pm$. This gives rise to the following set of two coupled differential rate equations:

$$\dot{N}_+ = -r_+ N_+ + r_- N_- \ , \tag{4.1a}$$

$$\dot{N}_- = \ \ r_+ N_+ - r_- N_- \ . \tag{4.1b}$$

Obviously, $N_+ + N_- \equiv N$ is conserved, and hence $N_\mp = N - N_\pm$. This permits us to decouple the set:

$$\dot{N}_\pm = -2\bar{r} N_\pm + r_\mp N \ , \tag{4.2}$$

with the average rate $\bar{r} = (r_+ + r_-)/2$. This has the solution

$$N_\pm(t) = \underbrace{\left[N_\pm(t=0) - \frac{r_\mp}{2\bar{r}} N\right] \mathrm{e}^{-2\bar{r}t}}_{\text{vanishes as } t \to \infty} + \underbrace{\frac{r_\mp}{2\bar{r}} N}_{\text{limit}} \ , \tag{4.3}$$

which among many other things shows that random lipid identity scrambling due to lipid flip-flop happens with the rate $2\bar{r} = r_+ + r_-$. As it turns out, typical flip-flop times for common phospholipids under most ordinary circumstances is in the hours-to-days regime (Sperotto and Ferrarini, 2017), which poses the question why cells, which live longer than that, are in fact asymmetric.

The answer is: because they *actively* pump certain types of lipids towards a specific side, counteracting the random scrambling. Let us work out a very simple model for that. Assume we have M "flippases" in the system which with a rate of F_+ can flip a particular type of lipid (which we'll count with the variable P) into the $+$ leaflet. Such flippases really exist, and they turn out to be slow. This in particular means they are not limited by the available "substrate," and so their transport rate is not proportional to how many flippable lipids they find in the $-$ leaflet. Our specific lipid species will therefore satisfy the following flippase-corrected set of differential equations:

$$\dot{P}_+ = -r_+ P_+ + r_- P_- + M F_+ \ , \tag{4.4a}$$

$$\dot{P}_- = \ \ r_+ P_+ - r_- P_- - M F_+ \ . \tag{4.4b}$$

We again have $P_+ + P_- = P = \text{const.}$, and the solution is once more easy to work out:

$$P_\pm(t) = \underbrace{\left[P_\pm(t=0) - \frac{r_\mp P \pm M F_+}{2\bar{r}}\right] \mathrm{e}^{-2rt}}_{\text{vanishes as } t \to \infty} + \underbrace{\frac{r_\mp P \pm M F_+}{2\bar{r}}}_{\text{limit}} , \tag{4.5}$$

where the "$\pm\, MF_+$" terms are new compared to Eqn. (4.3). Observe that their impact is to ensure that the + leaflet will get more lipids and the − leaflet will get less. In particular, the fraction of special lipids in the $\pm$ leaflet in the limit $t \to \infty$ is

$$\lim_{t\to\infty} \frac{P_\pm(t)}{P} = \frac{r_\mp P \pm M F_+}{2\bar{r}\, P} = \frac{r_\mp}{2\bar{r}} \pm \frac{M}{P} \times \frac{F_+}{2\bar{r}} \, . \tag{4.6}$$

Notice that once $F_+ \to r_+ \times (P/M)$, the sorting becomes perfect: all lipids are in the + leaflet.[3] Without taking our model too seriously, a very simple high-level conclusion is that flippases can clean up scrambling if they are faster than the spontaneous flip-flop rate of individual lipids, and the factor by which they need to be better is the extent to which they are outnumbered by the lipids they are responsible for. For instance, if they are 1000 times less abundant than those lipids, they have to operate a 1000 times faster than those lipids flip in order to clean up. Are they?

It seems that typical flippases operate at the glacial-seeming speed of about 10 lipids per minute (Shukla and Baumgart, 2021). And yet, this is still around 4 orders of magnitude faster than the spontaneous flip-flop rate. In other words: only one sluggish flippase per 10 000 lipids can housekeep its substrate lipid's asymmetry, because spontaneous flip-flop is simply so very slow. It's for that reason that lipidomic biomembrane asymmetry is even a thing.

4.2 Differential stress

4.2.1 Setting the stage

This is probably more than enough background on "classical asymmetry" for us. We have seen that the two leaflets of a bilayer can harbor a different lipidome, a situation which nature creates by spending energy on certain types of transport proteins that sort lipids across the

[3] This is of course a limit in which our theory breaks down: once the flippable lipids become that rare in the − leaflet, our flippase becomes substrate limited, unlike what we assumed when we wrote Eqns. (4.4).

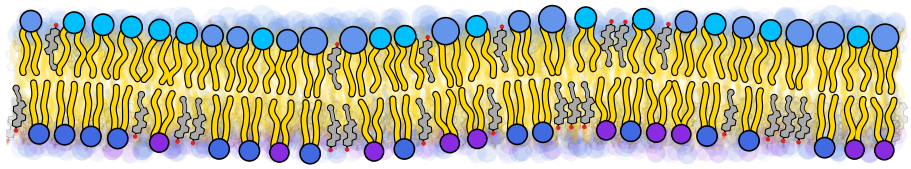


Figure 4.1 | Simple illustration of an asymmetric membrane. Different head group species are indicated by different colors, while different tail saturations are indicated by how wiggly the tails are (more wiggly, more double bonds). The gray beasts are cholesterol molecules, about which we will learn a bit more in Sec. 4.3.

leaflets. The resulting membranes may then look something like what's shown in Fig. 4.1, where the colors of the head groups and the different wigglienesses of the tails are cheap ways to distinguish lipids of different build. Pictures of this type do a great job in reminding us of the obvious, but also a terrible job of helping us see the less obvious, namely: *there are many ways for how to be asymmetric that go beyond mere lipid type.*

After all, a bilayer consists of two leaflets, and therefore any observable that is defined on a leaflet level could be different on the two sides. Many examples come to mind. Structurally, we have area per lipid, thickness, or charge density; mechanically, we should consider elastic moduli (bending, stretching, tilting, twisting, ...), spontaneous curvature, and leaflet tension; and dynamically, one could think of diffusion constants, membrane viscosity, and any number of relaxation rates.

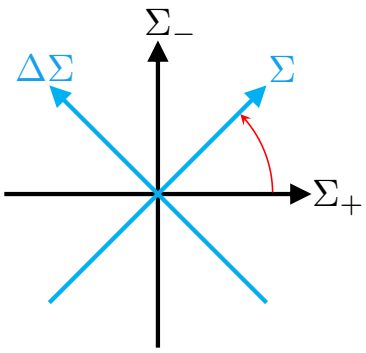
Furthermore: it is reasonable to expect that all of these observables somehow "talk" to one another—in the sense that if we change one, this will likely have implications for the others. But if so, then this also means that if we make one of these observables asymmetric, this might very well lead to asymmetries in other observables, too. A very physicsy way of saying is: *if you break one symmetry, others will likely break as well.*

My goal in this chapter is to get you curious about one specific observable: the difference between the mechanical tensions in the two leaflets, something my group has proposed to call *differential stress* (Hossein and Deserno, 2020; Foley et al., 2023).

4.2.2 An innocent change of variables

Let us hence consider a membrane, maybe like the one in Fig. 4.1, and define the lateral mechanical tensions Σ_+ and Σ_- of its two leaflets. How would we measure those. Ooh, that's really difficult in experi-

ment, and as best as I know there's no reliable method yet. But *operationally*, we would have to imagine *somehow* grabbing only one of the two leaflets and measuring how much it laterally pulls or pushes against us holding it at some given area strain. In a simulation this is far easier: we basically just need to measure a spatially resolved stress tensor, from it the excess lateral tension, and then only integrate it over the leaflet of interest. This is not really important here, so let me gloss over it with the appropriate air of nonchalance.[4] The point is that for various reasons it is useful to not just work with the individual leaflet tensions but instead with their symmetric and antisymmetric linear combinations, defined as



$$\left.\begin{aligned} \text{net tension:} \quad \Sigma &= \Sigma_+ + \Sigma_- \\ \text{differential stress:} \quad \Delta\Sigma &= \Sigma_+ - \Sigma_- \end{aligned}\right\} \longleftrightarrow \Sigma_\pm = \frac{1}{2}\big(\Sigma \pm \Delta\Sigma\big) \ . \tag{4.7}$$

The reason is that, as I will soon argue, we often are in a situation in which $\Sigma = 0$ or at least very small compared to $\Delta\Sigma$. The two leaflet tensions are then equal (or close to equal) in magnitude, but opposite in sign, and much of the bilayer's response is then driven by $\Delta\Sigma$. Another way of saying this is: formulas that describe the way in which the stresses affect the system will then only depend on $\Delta\Sigma$, not individually on Σ_+ and Σ_-.

Just a few lines up I have admitted that at the moment we do not really know how to measure leaflet tensions. Hence, we also cannot explicitly check whether they are different. It seems therefore quite bold to claim that $\Delta\Sigma \neq 0$ is an interesting scenario to ponder. Why should we care if we cannot measure it?

"Not yet measured" and "not yet measurable" are no reasons to not look at something theoretically. The really poisonous situation is if something is *in principle* not measurable, but this is not the case here.[5] $\Delta\Sigma$ is very much measurable. In fact, it turns out that some simple physical considerations not only suggest that, often, $\Delta\Sigma \neq 0$,

[4] OK, if you insist: check out the following references for a bit more detail: Różycki and Lipowsky (2015); Hossein and Deserno (2020); Foley et al. (2023); Lipowsky (2024); Foley and Deserno (2024).

[5] Think of gravitational waves: Einstein predicted them in 1916, and we measured them in 2015. Theoretical physicists did not stop working on the subject just because it was really hard to find them. In contrast, quantum mechanics insists that the values of two non-commuting observables cannot be simultaneously measured precisely, and now it would be quite bad to pretend that they have some definite but unknown values, because this opens the door to quantum paradoxes.

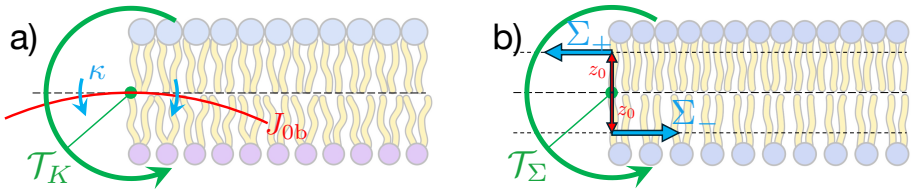


Figure 4.2 | There exist (at least) two conceptually different mechanisms that create a torque (density) in a lipid membrane. (a) If the bilayer prefers a spontaneous curvature J_{0b} because the lipids have a shape that is happier in that arrangement, then $-\kappa J_{0b}$ is a "bending-derived" torque. (b) Alternatively, if the two leaflets have different leaflet tensions $\Sigma_+ \neq \Sigma_-$, and hence a differential stress $\Delta\Sigma$, then $\Delta\Sigma\, z_0$ is a tension-derived torque, in which the lever arm z_0 is the position of a suitable reference surface in the leaflet. Illustration adapted from (Deserno, 2024).

but that the value can be quite large. I will give you two examples for this. One we can do right now, the other one I need to postpone until we have covered cholesterol—very much towards the end of these notes, in Sec. 4.3.2.

The first example refers to something called "torque balance". A Helfrich membrane with a spontaneous curvature wants to bend. There is an internal driving force that compels it to assume a curved state. The better word for "internal driving force to bend" is *torque*. In a slightly simplified way,[6] we can define the bending torque (density) as

$$\mathcal{T}_K = \frac{\partial e_{\text{bend}}}{\partial J} \overset{(3.42)}{=} \kappa(J - J_{0b}) \;, \tag{4.8}$$

where we ignored the Gaussian curvature term. For instance, if such a membrane is flat, then the torque is $-\kappa J_{0b}$, and if it already has the curvature $J = J_{0b}$, then the torque vanishes. I tried to illustrate this in Fig. 4.2a. Notice the (maybe slightly awkward) sign convention that follows from this: a *positive* J_{0b} bends the membrane *down*, and that gives rise to a *negative* torque.

Observe now that a bending-derived spontaneous curvature is not the only way a torque can arise in a bilayer. Consider the situation illustrated in Fig. 4.2b: the upper leaflet is under compressive stress and pushes outward. The inner leaflet is under tensile stress and pulls inward. Where does it push? Well, across its entire thickness. But

[6] Torque is really more complicated, since in a funnily shaped membrane it might locally act with different strength in different directions. What does "different strength in different directions" even mean? To make this precise, we'd need to talk about a *torque tensor*, which is the curving partner in crime of a *stress tensor*. For the case of membranes the reader will find this sussed out in considerable detail in (Deserno, 2015).

in order to work with such situations more easily, it is useful to create the notion of a ("Gibbs") reference surface, picked such that force and torque balance work just the same if we pretend that a delocalized force density that acts all across the thickness of a leaflet (conceivably even with different strength) is subsumed into a single force at a single distance. There are subtleties involved with the choice of such surfaces,[7] but these are often not all that important. Let us therefore just pretend that we have such a surface, and it has a distance z_0 from the bilayer midsurface. In our present context, we need a surface that's typically called the "neutral surface."[8] As the illustration shows, this quite literally results in a force couple that creates the torque

$$\mathcal{T}_\Sigma = \Sigma_+ z_0 - \Sigma_- z_0 = \Delta\Sigma\, z_0 \ . \tag{4.9}$$

Let's make sure the sign convention works out: we previously discovered that in order to bend the membrane *down* we need a *negative* torque. This expression is negative if the differential stress is negative. Since tensions pull and hence have the opposite sign of forces, down-bending means the upper leaflet needs to be under *compression* and hence *negative tension*, which in turn implies a negative differential stress, as required. Phew.

With these definitions under our belt, let us consider the following experimental fact: we know that scientists can create asymmetric membranes that have two different types of lipids on the two sides.[9] For instance, Elani et al. (2015) produced asymmetric giant unilamellar vesicles (GUVs) in which one leaflet contains a lipid called "POPC" and the other side contains a lipid called "DOPC." Because we know a fair amount about spontaneous lipid curvatures (and how bilayer curvatures arise from those of their underlying leaflets), we know that such a bilayer has a bending derived spontaneous curvature $J_{0\mathrm{b}} \approx 0.017\,\mathrm{nm}^{-1}$. The spherical curvature at which such a vesicle would have its torque relaxed is $R_0 = 2/J_{0\mathrm{b}} \approx 120\,\mathrm{nm}$. And yet, the vesicles which these authors created had radii around $20\,\mu\mathrm{m}$, which is about two orders of

[7] There's more than one way to do this, depending on what you want your reference surface to do. I'm afraid I'll have to skip that here, too.

[8] Definition: the reference surface where at quadratic order bending and stretching decouple, so that we do not get a bilinear term of the form "stretch times bend."

[9] Highly nontrivial statement! For the longest time, scientists did *not* know how to do this. Luckily, the state of affairs has vastly changed in the past couple of years. In fact, Krompers and Heerklotz (2023) have recently published a review in which they list approximately 70 individual experimental protocols, grouped into 4 general strategies. Without doubt, the ability to reconstitute clean artificial asymmetric model membranes is a major reason for the renaissance of membrane asymmetry.

magnitude larger than the relaxed R_0 value we just calculated. This raises the urgent question: why are these GUVs stable? Why don't they fall apart, for instance by spitting thin tubules out (or, maybe, *in*, depending on which way the POPC/DOPC asymmetry points). A possible answer is that the *bending* torque might well want to do that, but it's not the only torque in town. If these vesicles were *also* under differential stress, there would be a second source of torque, and these two torques could balance in such a way that the much weaker curvature of the GUV is stable.

Let us work out the required differential stress. Assuming, for simplicity, that torque balance is perfect, and taking $J = 0$ as a good proxy for these huge GUVs, we get

$$0 \stackrel{!}{=} \mathcal{T} = \mathcal{T}_K + \mathcal{T}_\Sigma = -\kappa J_{0\mathrm{b}} + \Delta\Sigma\, z_0 \quad \Rightarrow \quad \Delta\Sigma = \frac{\kappa J_{0\mathrm{b}}}{z_0}\,. \tag{4.10}$$

Observe that the sign makes sense: if the spontaneous bending-related curvature $J_{0\mathrm{b}}$ is *positive*, so the + leaflet bends like the outside of a sphere, we need a tension in the outer leaflet to pull it back flat, and this means a positive differential stress.

We're ready for the punchline: inserting numbers! Everything on the right hand side of Eqn. (4.10) is known. A typical bending rigidity for either POPC or DOPC membranes is about $30\,k_\mathrm{B}T$, the spontaneous curvature was $J_{0\mathrm{b}} \approx 0.017\,\mathrm{nm}^{-1}$, and a very typical value for the neutral surface is $z_0 \approx 1\,\mathrm{nm}$. Plugging in these numbers, we find $\Delta\Sigma \approx 2\,\mathrm{mN/m}$.

How much is $2\,\mathrm{mN/m}$? Here are two possible points of comparison. Sec. 3.3.2 mentioned that the net bilayer tension needed to rupture membranes tends to be in the few-to-ten mN/m range. Hence, the differential stress we just estimated is getting uncomfortably close to *bilayer* tensions that might destroy membranes. Second, typical bilayer tensions for relaxed biomembranes are much lower—fractions of mN/m (Morris and Homann, 2001). This shows that even if a membrane might look fairly relaxed (it's measured tension is "quite small"), it could actually experience sizable internal stress. This matters for anything that would notice this internal stress, such as the phase behavior in each leaflet, the distribution of cholesterol, the diffusion of lipids, or the functioning of transmembrane proteins whose mode of action requires any type of depth-dependent area change (mechanosensitive channels or transporters would be prime examples to watch out for).

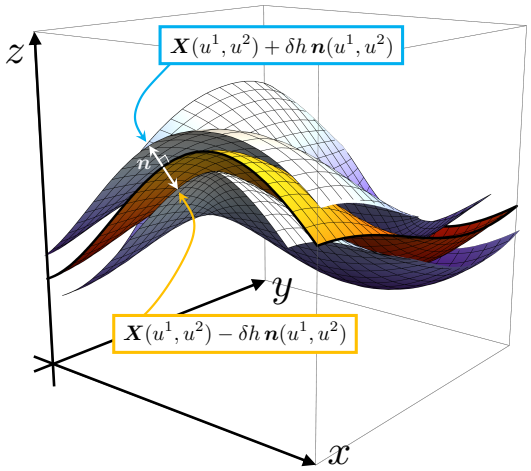


Figure 4.3 | Illustration of the notion of a *parallel surface*. A "central" surface is defined via the parametrization $\boldsymbol{X}(u^1, u^2)$ (see Sec. 3.2.2). At each point on that surface we have a normal vector, $\boldsymbol{n}(u^1, u^2)$, and we can therefore define a so-called parallel surface by locally displacing the central reference surface by some amount δh along $\boldsymbol{n}(u^1, u^2)$ that does not depend on the coordinates $\{u^1, u^2\}$. (The two examples shown here are an "up" and a "down" displacement.) Fun fact: this creates a *local* coordinate system $(u^1, u^2, \delta h)$ of 3d space near our central surface, but this usually does not extend to arbitrary values of δh, because in concave regions of the original surface the normal vectors all come together and can create cusps.

4.2.3 Differential stress as preferred curvature

A difference in lipid packing density on the two sides of a leaflet can give rise to differential stress. This, as we have seen, creates a torque which wants to bend the membrane. Conversely, if we bend the membrane such as to reduce the differential stress, then there should be a special curvature at which the packing density difference relaxes, because the "outside" has gained enough area, and the "inside" has lost enough area, such as to exactly accommodate for the imposed area difference. This is true, and it offers a way to rephrase the notion of differential stress (or strain) in terms of curvature. It's not necessary to do so, but it provides some nice intuition, and it might help to express different concepts in the same framework. To do so here, I first need to tell you about a very nifty result from differential geometry, called the "parallel surface theorem."

Consider Fig. 4.3, which looks like a stack of three surfaces. What we look at here is a central surface $\boldsymbol{X}(u^1, u^2)$ (using the more general parametrization discussed in Sec. 3.2.2) sandwiched between two cousins, called "parallel surfaces." These are defined by locally displacing our initial reference surface by some fixed amount δh along the local

normal vector:

$$\boldsymbol{X}'(u^1, y^2) = \boldsymbol{X}(u^1, u^2) + \delta h\, \boldsymbol{n}(u^1, u^2)\ . \tag{4.11}$$

For small enough δh, the new surface $\boldsymbol{X}'(u^1, u^2)$ is again nice and smooth, and so we can define everything for it that we can also define for $\boldsymbol{X}(u^1, u^2)$, such as area elements, normal vectors, curvatures, *etc.* In fact, we can express these objects as changes of the corresponding objects on the reference surface. Three particularly remarkable formulas show how area element, total curvature, and Gaussian curvature on the primed surface change into

$$\mathrm{d}A' = \mathrm{d}A\, \left[1 + J\,\delta h + K\,\delta h^2\right]\ , \tag{4.12a}$$

$$J' = \frac{J + 2K\,\delta h}{1 + J\,\delta h + K\,\delta h^2}\ , \tag{4.12b}$$

$$K' = \frac{K}{1 + J\,\delta h + K\,\delta h^2}\ . \tag{4.12c}$$

You might think these are maybe expansions in δh, but they are not. *These expressions are exact*! Proving them is an exercise of playing around withe more general surface parametrization from Sec. 3.2.2. Indeed, do Carmo (1976) poses it as an exercise. Absent a proof, you might at least want to convince you that this gives the correct result for spheres and cylinders.[10]

With these preliminaries settled, let us now look at a curved piece of membrane, such as the one illustrated in Fig. 4.4. It has an "outside" and an "inside" leaflet, with an area that's larger or smaller than the bilayer midsurface, respectively. And by "area" we mean the area as measured some reference distance z_{0+} displaced towards the outside and z_{0-} towards the inside (the numbers might differ since, after all, we assume our membrane to be asymmetric). This is obviously where the parallel surface theorem comes in. We would like to know what these displaced areas are, because this will tell us how many lipids we can fit onto the two sides—with or without extra stress. Making use of Eqn. (4.12a), we immediately find

$$\mathrm{d}A'_{\pm} = \mathrm{d}A\, \left[1 \pm J\, z_{0\pm} + K\, z_{0\pm}^2\right]\ . \tag{4.13}$$

[10] To be fair, this theorem is a lot more than what we need right now, but in my career I have found it to be useful on *so* many occasions that I felt it doesn't hurt if I tell you about this particularly beautiful spell.

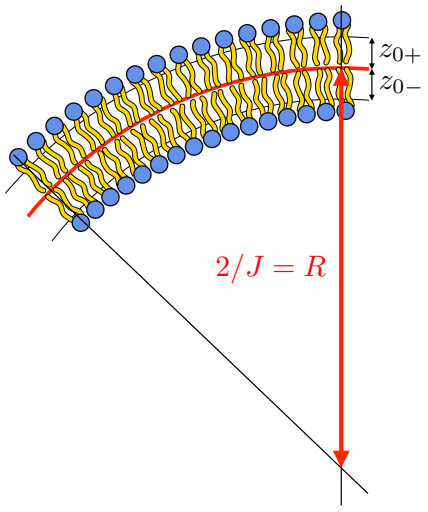


Figure 4.4 | Illustration of a small piece of a curved membrane, which has a midsurface of curvature J, corresponding to a radius of curvature $R = 2/J$. Each individual leaflet has its own reference surface (say, the *neutral surface* we encountered in the previous section), which is *parallel displaced* from the bilayer's midsurface by some amount $\pm z_{0\pm}$. The parallel surface theorem quickly tells us the areas of these two neutral surfaces.

Given this, we can now calculate the *area strain* $s_\pm$ on the two surfaces:

$$s_\pm = \frac{\mathrm{d}A'_\pm - \mathrm{d}A}{\mathrm{d}A} = \pm J\, z_{0\pm} + K\, z_{0\pm}^2 \ . \tag{4.14}$$

If we have a membrane which is packed in such a way that it experiences no area strain in either leaflet when it is flat, then such a strain will build up upon bending. Conversely, if we have a membrane that does experience an area strain, we can reduce it by suitably bending it away from the more crowded side. In fact, at linear order in $z_{0\pm}$ we are guaranteed that there is a J at which $\Delta s = s_+ - s_-$ takes whatever values is needed to relax a pre-existing area strain difference. At quadratic order this is also possible, because we can choose J and K any way we want, even though now the specific choice for J and K might not be unique. This doesn't matter, though, because it will turn out later that all we need is expressions that are quadratic in the strain, which to lowest order turns out to be quadratic in $z_{0\pm}$, and for such terms it is actually enough to take Eqn. (4.14) up to linear order in $z_{0\pm}$.

Let's take one step after the other, though. Assume we have a bilayer that is unevenly packed on the two sides such that there is an area strain. We know we can pick (possibly not unique) values $J_{0\mathrm{s}}$ and $K_{0\mathrm{s}}$ for total and Gaussian curvature, respectively, such that this strain is canceled. However, at any *other* values a strain remains, which integrated over the whole membrane surface is given by

$$\langle s_\pm \rangle = \frac{1}{A} \int_{\mathcal{S}} \mathrm{d}A \left[\pm (J - J_{0\mathrm{s}})\, z_{0\pm} + (K - K_{0\mathrm{s}})\, z_{0\pm}^2 \right] \tag{4.15a}$$

$$= \pm \big(\langle J \rangle - J_{0\mathrm{s}} \big)\, z_{0\pm} + \big(\langle K \rangle - K_{0\mathrm{s}} \big)\, z_{0\pm}^2 \ , \tag{4.15b}$$

where A is the area of the whole membrane patch on its midsurface, and where $\langle J\rangle$ and $\langle K\rangle$ are the total and Gaussian curvature averaged over the patch. These area strains in the two leaflets of course come with an elastic stretching energy, following an energy expression à la Eqn. (3.40):

$$E_{\text{stretch}} = A\left[\frac{1}{2}K_{A\text{m}+}s_+^2 + \frac{1}{2}K_{A\text{m}-}s_-^2\right] \tag{4.16a}$$

$$= A\left[\frac{1}{2}\big(K_{A\text{m}+}z_{0+}^2 + K_{A\text{m}-}z_{0-}^2\big)\big(\langle J\rangle - J_{0\text{s}}\big)^2 + \mathcal{O}(z_{0\pm}^3)\right], \tag{4.16b}$$

where $K_{A\text{m}\pm}$ are the area expansion moduli in the $\pm$ leaflets and where we expanded up to quadratic order in $z_{0\pm}$ (and hence also up to quadratic order in curvature (differences). As mentioned earlier, we realize that at that order the quadratic term in Eqn. (4.14) is indeed irrelevant, and we do not need to know $K_{0\text{s}}$ after all.

This is a bit of a curious energy. At first sight, it *looks* like a bending energy, just like the κ-term in Eqn. (3.42). But notice that since we needed to *first* integrate over the whole membrane in order to get the total area strain, and only *then* did we square up to get an elastic energy, we ended up squaring the *average* curvature of the patch. In contrast, the usual bending energy arises from squaring up *local* curvatures and then integrating over the whole membrane. This swapping of "squaring" and "integrating" creates an expression that is genuinely *nonlocal*: we need to know something about the curvature everywhere on the membrane to calculate $\langle J\rangle$, and only then do we enter this averaged expression into our quadratic elastic energy.

Having done so, we can then make the expression at least *superficially* look like a bending energy, namely by renaming the assemblage of prefactors out front (and dropping the higher order $\mathcal{O}(z_0^3)$ terms):

$$\frac{E_{\text{stretch}}}{A} = \frac{1}{2}\kappa_{\text{nl}}\big(\langle J\rangle - J_{0\text{s}}\big)^2 \quad \text{with} \quad \kappa_{\text{nl}} = K_{A\text{m}+}z_{0+}^2 + K_{A\text{m}-}z_{0-}^2\,, \tag{4.17}$$

where κ_{nl} can be viewed as a *nonlocal bending modulus*. Expressions like this have in fact been proposed numerous times before (Evans, 1974; Helfrich, 1974b; Evans, 1980; Svetina et al., 1985; Seifert et al., 1992; Miao et al., 1994).

Besides the local $\rightarrow$ global nuisance, we have succeeded in rewriting the stretching energy that is associated with uneven leaflet packing, and how it changes when we curve the surface more or less, into an effective nonlocal bending energy. To do so, we re-expressed the area strain by some effective curvature $J_{0\text{s}}$. For completeness, let us also

calculate what that curvature is in terms of the areas $A_\pm$ which the two sides would prefer to take individually in order to get rid of area strain, and how these two areas related to the midsurface area A. Since to linear order in $z_{0\pm}$ we have $A_\pm = A(1 \pm z_{0\pm} J_{0\mathrm{s}})$, we can divide these two equations to eliminate A and then solve the result for $J_{0\mathrm{s}}$. We can then subtract them, insert whatever we just found for $J_{0\mathrm{s}}$, and then solve for A. Doing so yields

$$J_{0\mathrm{s}} = \frac{A_+ - A_-}{A_+ z_{0-} + A_- z_{0+}} \qquad \text{and} \qquad A = \frac{A_+ z_{0-} + A_- z_{0+}}{z_{0+} + z_{0-}} \,. \tag{4.18}$$

We would now like to reap the benefits of our rewriting. But in order to keep things manageable, we want to get rid of the local/global nuisance. We can rather cheaply do so by agreeing to only talk about **surfaces of constant total curvature**. This might seem like a serious restriction, but it does cover the by far most common cases of planes, spheres, and cylinders.

Heads up 7

Important restriction happening! You've been warned.

We have at this point encountered two elastic energies: a genuine bending energy, Eqn. (3.42), and a stretching energy that arises from packing differences in the two leaflets, the associated stress this causes, and how it changes upon additional bending—Eqn. (4.17). For a general elastic description, both of these terms must be present, since they account for subtly different physics. In the case of constant total curvature surfaces, we can write this combined energy density as

$$\boxed{e^\star_{\mathrm{bend}} = \frac{1}{2}\kappa(J - J_{0\mathrm{b}})^2 + \frac{1}{2}\kappa_{\mathrm{nl}}(J - J_{0\mathrm{s}})^2 \,,} \tag{4.19}$$

where for simplicity we ignore the net tension and any Gaussian terms. This expression contains two bending moduli (a local and a nonlocal one) as well as two spontaneous curvatures: one due to lipid shape (*i. e.*, a material parameter), and one due to a pre-existing area strain (*i. e.*, a condition created when the membrane was made).

This generalized curvature-stretching elastic energy permits us to account for the two torque contributions we have discussed in the previous sections in one go. All we need to do is differentiate this energy with respect to the curvature to get the torque density:

$$\mathcal{T} = \frac{\partial e^\star_{\mathrm{bend}}}{\partial K} = \underbrace{\kappa(J - J_{0\mathrm{b}})}_{\mathcal{T}_K} + \underbrace{\kappa_{\mathrm{nl}}(J - J_{0\mathrm{s}})}_{\mathcal{T}_\Sigma} \,. \tag{4.20}$$

The first term is of course again the torque density from Eqn. (4.8); the second term does not look like the stretching-based torque density from

Eqn. (4.9), but of course has to be equal to it. We see that it is—taking $J = 0$ as the reference state—by inserting the known expressions for $\kappa_{\rm nl}$ (Eqn. (4.17)) and $J_{0\rm s}$ (Eqn. (4.18)), respectively. If for simplicity we take $z_{0+} = z_{0-} = z_0$, we get

$$\begin{aligned}\mathcal{T}_\Sigma = -\kappa_{\rm nl} J_{0\rm s} &= -K_A z_0^2 \frac{A_+ - A_-}{A_+ z_0 + A_- z_0} = -K_A z_0 \frac{A_+ - A_-}{A_+ + A_-} \\ &= -2K_{A\rm m}\left[\frac{A_+ - A}{2A} + \frac{A - A_-}{2A}\right] z_0 \stackrel{*}{=} -K_{A\rm m}(-s_+ + s_-) z_0 \\ &= \big(\Sigma_+ z_0 - \Sigma_- z_0\big) = \Delta\Sigma\, z_0 \ , \end{aligned} \tag{4.21}$$

which is exactly the answer we expected from Eqn. (4.9). The only nontrivial thing happened at $*$, where we strangely pretended that $(A_+ - A)/A$ is the negative of the strain in the $+$ leaflet. Why? Because we decided to take $J = 0$ as the reference state, not $J_{0\rm s}$. If $J_{0\rm s} > 0$, then the surface wants a positive curvature and so its outer surface A_+ must be bigger than its inner surface A_-, or their suitably defined average A—exactly in line with Eqn. (4.18). But that also means that in the flat state, $J = 0$, the upper leaflet is compressed, even though $(A_+ - A)/A$ is positive—showing that it's the *negative* of that value which is the actual strain present in the leaflet (and similarly for the other side).

If you get the feeling that keeping track of the signs of torques is tricky, you might be on to something. But then, this is also a (minor) reason why rewriting the combined bending-stretching energy as two curvature terms is helpful: our visual intuition about what the signs of $J_{0\rm s}$ and $J_{0\rm b}$ mean is clear, and we're comparing apples to apples here, so the possibility of confusion is minimal.

Now that we have the total torque density in Eqn. (4.20) we can ask at what special curvature it will vanish—meaning, what (constant curvature) shape the membrane wishes to take. Setting $\mathcal{T} = 0$ and solving for J leads to:

$$\boxed{J_0^\star = \frac{\kappa J_{0\rm b} + \kappa_{\rm nl} J_{0\rm s}}{\kappa + \kappa_{\rm nl}} \ ,} \tag{4.22}$$

showing that the relaxed curvature is the *rigidity-weighted average* of our two characteristic curvatures, $J_{0\rm b}$ and $J_{0\rm s}$, each weighted with their respective modulus.

There's something conceivably unexpected about a membrane in its relaxed state with curvature $J_0^\star$: *it is generally under differential stress.*

You might think that in the final equilibrium state all forces vanish, but that is not so. In fact, if you look at the energy in Eqn. (4.19), you can picture this as a sum of two (harmonic) springs (in fact, two springs *in parallel*), each of which has a rest length. *But these rest lengths are not the same*! This means that both springs have to come to a compromise, and $J_0^\star$ is exactly that compromise. Obviously, in the resulting state *neither* spring is perfectly relaxed; in other words, neither the bending torque nor the tension torque vanish, their nonzero values just balance one another. And since the tension torque goes along with a differential stress, that differential stress cannot be zero.

Since to linear order in z_0 the leaflet strain is $s_\pm = \pm(J - J_{0\mathrm{s}})z_{0\pm}$ (see Eqn. (4.15b)), the differential stress is

$$\Delta\Sigma = K_{A\mathrm{m}+} z_{0+}(J - J_{0\mathrm{s}}) - \Big[- K_{A\mathrm{m}+} z_{0-}(J - J_{0\mathrm{s}})\Big] \ . \tag{4.23}$$

Taking the simple case where $z_{0+} = z_{0-} = z_0$, we find that the differential stress in the torque-balanced state is

$$\boxed{\Delta\Sigma^\star = \frac{\kappa_{\mathrm{nl}}}{z_0}(J_0^\star - J_{0\mathrm{s}}) = -\frac{\kappa}{z_0}(J_0^\star - J_{0\mathrm{b}}) \ ,} \tag{4.24}$$

where the second equation follows from using Eqn. (4.22). Multiplying the first expression by κ and the second by κ_{nl} and adding them up, we can eliminate $J_0^\star$ and write the answer purely in terms of our original characteristic spontaneous curvatures:

$$\boxed{\Delta\Sigma^\star = \frac{\kappa_{\mathrm{harm}}}{2z_0}\left(J_{0\mathrm{b}} - J_{0\mathrm{s}}\right) \ ,} \tag{4.25}$$

where $\kappa_{\mathrm{harm}} = 2\kappa\kappa_{\mathrm{nl}}/(\kappa+\kappa_{\mathrm{nl}})$ is the *harmonic mean* between κ and κ_{nl}. This expression shows rather vividly that if the bending term and the leaflet packing term do not agree on the same relaxed curvature, then the bilayer will be under differential stress. This is yet another reason for why we should generally expect such a stress to exist.

4.3 Cholesterol

Among the many lipids contained in biomembranes, cholesterol is a particularly strange fellow (Mouritsen and Zuckermann, 2004; Nes, 2011). Unlike most "normal" lipids, which contain some hydrophilic head group and usually two aliphatic chains, cholesterol is a rigid ring

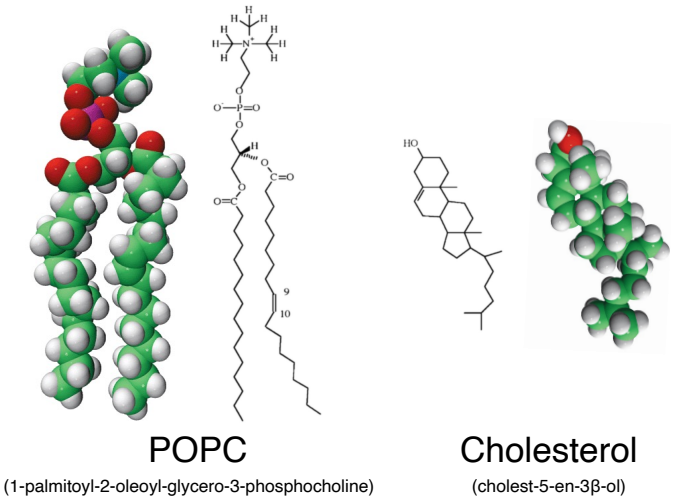


Figure 4.5 | Illustration of two very common lipids in mammalian plasma membranes: POPC and Cholesterol. Observe that POPC has the "typical" structure we associate with lipids: a hydrophilic head group (here: phosphatidylcholine) attached via a glycerol backbone via ester bonds to two fatty acid chains (here: palmitic acid and oleic acid), one of which happens to be "unsaturated" (*i. e.*, it has a double bond). In contrast, cholesterol is a complicated ring system and has as its hydrophilic moiety a single hydroxyl group. [The chemical structures were taken from (Cheng et al., 2006), the 3D models come from the webpage of Avanti Polar Lipids, https://avantilipids.com/.]

system that is a fair bit smaller than the size of a typical phospholipid, and its only hydrophilic moiety is a small hydroxyl group—see Fig. 4.5. Cholesterol is contained in the cell membranes of all higher animals and reaches its largest concentration in the plasma membrane, where it typically amounts to 20 – 40 mol%. Many other organisms contain similar "sterol" molecules, which the unaided (*i. e.*, non-chemist) eye cannot easily distinguish from cholesterol; for instance, fungi contain ergosterol,[11] plants contain phytosterol.

4.3.1 Cholesterol flip-flop

Cholesterol is responsible for a huge set of cellular phenomena in general, and membrane phenomena in particular. Too much to go into now. For the purpose of these notes I will largely restrict to a first

[11] Fun fact: sterols are clearly important for these organisms, and taking them away is lethal. Since fungi use a different sterol (ergosterol) than humans, whose biosynthesis relies on proteins which we humans do not have, those proteins are a primary target for antifungal drugs, because they are then less likely to interfere with human biochemistry. Clotrimazole and fluconazole are great examples.

discussion on how cholesterol impacts the phenomenon of lipid membrane asymmetry. Everything starts with the observation that, being such a small molecule with such a miniscule hydrophilic moiety, it has a much larger flip-flop rate than phospholipids. While (as usual) precise values depend on the specific conditions, flip-flop rates seem to be so fast that the inferred results tend to be limited by the time resolution of the experimental technique that was applied. Leventis and Silvius (2001) conclude that the flip-flop time is shorter than a minute, Steck et al. (2002) underbid this by claiming it is shorter than a second, and Hamilton (2003) argues it can be as fast as milliseconds. Much larger times (200 min at 50 °C) have been claimed by Garg et al. (2011) and subsequently been criticized by Steck and Lange (2012). My (a theorist's...) impression is that such large times are not widely accepted in the community. If we take the between-seconds-and-milliseconds range, then it's fair to say that cholesterol's flip-flop rate appears to be between 5 and 6 orders of magnitude faster than that of phospholipids. Why does this matter?

In Sec. 4.1.2 I had argued that the flip-flop rate of phospholipids is *slow enough* for cellular flippases to sort them between the bilayer's leaflet to the cell's liking. The asymmetric lipid distribution is a non-equilibrium steady state, whose very existence is only possible because even "sluggish" flippases are fast enough to out-shuffle random flip-flop. But this sorting is no longer possible for a fast-flipping species like cholesterol. The product of abundance and shuffling rate of any hypothetical cholesterol flippase would have to be those 5 to 6 orders of magnitude larger, and that just seems biochemically impossible. Steck and Lange (2018) in fact point out that such a machinery would completely exhaust a cell's energy budget.

The upshot is: cholesterol cannot be placed in some particular asymmetric arrangement as part of a cell's overall lipidomic asymmetry. It goes where it wants to go. And since (at least in the plasma membrane) the mole fraction of cholesterol is fairly sizable, this must affect everything we have so far discussed in terms of lipid packing, area strain, and differential stress. How do we need to adjust our thinking?

One conceivable deduction is that cholesterol, if sufficiently abundant, will eliminate differential stress. The reasoning being, if there is a crowded leaflet and a depleted leaflet, giving rise to compressive and tensile strain, respectively, which in turn creates elastic energy due to the resulting differential stress, then we can lower that energy by shuffling cholesterol from the side that is crowded to the side that still has space. The lowest energy state we can achieve this way is the one where the differential stress vanishes. Indeed, it has been claimed that bilayer

membranes with frequent flip-flops have tensionless leaflets—which is in fact the title of a paper by Miettinen and Lipowsky (2019). Besides the brief argument I just outlined, these authors further bolster their findings with Molecular Dynamics simulations (coarse-grained, MARTINI level), in which the differential stress in an asymmetric membrane relaxes to values indistinguishable from zero after replacing 10 POPC lipids on each side with cholesterol molecules.

Hence, considering the abundance of cholesterol in cellular membranes, it would seem that everything I have said in Sec. 4.2 about differential stress only applies to artificial cholesterol-free model systems, and you might get worried I'm wasting your time with conceptual niceties that are irrelevant in the real world. Have I?

Fear not—I haven't. And the reason is that, like virtually anything concerning cholesterol, things are always just a little bit more subtle. For a start: if I had asked you how a fast flipping species would distribute between the two leaflets of a bilayer *before* you even knew anything about asymmetry, you would most likely have said: 50:50. Why? *Because of entropy.* One of the archetypal thought experiments I am certain you all have encountered at the beginning of your first course in Statistical Physics is the distribution of gas molecules between two compartments of equal volume. The 50:50 answer simply comes about because this yields the largest number of microstates (the logarithm of which is proportional to the entropy). Here, the compartments are the two leaflets, and the entropy is largest if we put half of our cholesterol into one, and the other half into the other.

Which part of that argument becomes wrong in the presence of asymmetry? Clearly, thermodynamics teaches us that entropy maximization is a far more general idea—it basically underlies the extremalization principles of all thermodynamic potentials. But we learn that additional *energetic* considerations can compete with entropy (which is of course why we introduce all those other thermodynamic potentials in the first place). And indeed, differential stress is such a competing effect: it is energetically better to shuffle cholesterol around such as to lower the elastic *energy*, but it will come at the cost of also lowering the *entropy*. So what will happen? A compromise, generally. The stress will be *somewhat* reduced, and the entropy will be *somewhat* lowered, but neither side fully gets its way. Slightly more precise: from energy E and entropy S we can define a free energy $G = E - TS$, and rather than exclusively minimizing E or exclusively maximizing S, we will compromise by minimizing G. This obviously means that a differential stress will remain.

How in fact *do* we minimize G? Imagine that each leaflet has its

free energy—say, G_+ and G_-. and what we're doing now is to shuffle cholesterol around—*i. e.*, change its number in each leaflet—until any further shuffline no longer changes G (to first order). As an equation:

$$0 = \frac{\partial G}{\partial C_+} = \frac{\partial (G_+ + G_-)}{\partial C_+} = \frac{\partial G_+}{\partial C_+} + \frac{\partial G_-}{\partial C_+} \stackrel{*}{=} \frac{\partial G_+}{\partial C_+} - \frac{\partial G_-}{\partial C_-} \ . \tag{4.26}$$

$C_\pm$ is the number of cholesterol molecules in each leaflet and at "$*$" we used $\mathrm{d}C_+ = -\mathrm{d}C_-$, which follows from $C_+ + C_- = C = \text{const}$. Since the derivative of the free energy with respect to particle number is the *chemical potential*, we have arrived at the proper thermodynamic equilibrium condition that must hold for a fast flipping species:

$$\boxed{\mu_+ = \mu_- \ .} \tag{4.27}$$

In some sense, this of course also holds for all the other lipids—for sufficiently long times and when we switch the non-equilibrium flippases off. But surely on time scales larger than the flip-flop time of cholesterol and shorter than time scales over which even dead model membranes equilibrate their phospholipids, the cholesterol distribution is determined by an equilibration of cholesterol's chemical potential between the leaflets.

Notice that this condition is not badly in conflict with the original idea by Miettinen and Lipowsky: it is not *wrong* that differential stress is a driver of the cholesterol distribution. It's just not the *only* driver. Rephrased in thermodynamic language: there are multiple contributions to the chemical potential of cholesterol, *one of them* being differential stress. It's just not the only one.

Our musings have actually paved the way to being more quantitative about this situation and trying to predict the cholesterol distribution. We need to find its drivers and account for them in a suitably constructed free energy of the situation. As usual, the gold standard here would be to write down a Hamiltonian for this situation and then do the partition function. And also as usual, this is completely impractical.[12] The far more common approach is to patch together plausible terms using a variety of well understood model calculations. While not necessarily quantitatively accurate (and sometimes not even qualitatively...), this almost always provides a deeper conceptual insight into this situation. Towards the end of these lectures, let me show how one might do this for the case in point here.

[12] Except: this is *exactly* what Molecular Dynamics simulations do. Admittedly, though, this is not an attempt at an analytical answer.

4.3.2 A simple model for cholesterol's trans-leaflet distribution

The model I'm about to describe has been developed by Malavika Varma and myself (Varma and Deserno, 2022). It accounts for the two drivers already mentioned—entropy and differential stress—as well as a third important one: preferential partitioning. It is well known that cholesterol's free energy of insertion into a membrane depends on the lipid composition of that membrane. More specifically, it prefers to associate with more saturated phospho- and sphingolipids and less so with unsaturated lipids,[13] as is known from both experiment (Silvius, 2003) and simulation (Bennett et al., 2009).

The model I will show you has the redeeming quality of being simple, and so its predictions can be physically interpreted. It does not account for the complete physical situation, though, because it neglects quite a number of things. Two rather notable disinvites to the party are curvature and condensation effects. The former means we will not account for curvature elastic energies that arise—even in flat membranes!—because cholesterol can change the spontaneous curvature of the phase it embeds in, which has important energetic effects as recently examined by Allender et al. (2019). The hard part here is not so much including a curvature term but to figure out how cholesterol changes the spontaneous curvature, which is not just a matter of contributing its own intrinsic curvature to that of the lipids it mingles with (Sodt et al., 2016). This is also related to the second effect, condensation. It is generally true that the area of a lipid membrane is not just the

[13] "Saturation" is a measure for the absence of double bonds in the aliphatic chain of a lipid. The term "saturation" is rather widely used in organic chemistry to denote a compound (or part of a larger chemical structure) that resists addition reactions, such as hydrogenation (addition of hydrogen atoms). A hydrocarbon chain that only consists of single bonds is "saturated," since there is no place where we could easily add anything. Conversely, a double bond is more reactive and lends itself to a variety of additions. These double bonds interrupt what could be a regular zig-zag configuration of the aliphatic chain, thus making it more difficult for them to pack and order. As a consequence, such unsaturated fatty acids "melt" at lower temperature than their "high-melting" saturated counterparts. For instance, the major fatty acid in olive oil is oleic acid, which contains a double bond between carbon atom 9 and 10 (it's exactly the right "O" chain in the POPC lipid shown in Fig. 4.5), while the most common fatty acid in butter is palmitic acid (the left "P" chain in the POPC lipid shown in Fig. 4.5), followed by myristic acid and stearic acid—all of which are saturated. Hence, butter is solid at room temperature, while olive oil is a liquid. Moreover, the more reactive double bond in olive oil makes it more prone to oxidation, especially in the presence of light ("photo-oxidation"), which is why olive oil comes in darkened bottles.

sum of the molecular areas of its individual lipids,[14] and cholesterol is a particularly notable outlier in this regard: its planar polycyclic ring structure can order neighboring lipids so much that they pack better, leading to an overall area that can even be *smaller* than what it was before cholesterol was added, giving cholesterol effectively a negative differential area. This was especially vividly demonstrated in recent simulations by Leeb and Maibaum (2018).

OK, let's get going before you run out of steam. The three drivers we wish to include come in two different flavors, elasticity and mixing. The first we can write down as follows: Take a membrane of area A, whose individual leaflets rather wish to have areas $A_\pm$. The elastic part of the free energy is then of the form

$$G_\mathrm{el} = \frac{1}{2} K_{A\mathrm{m}+} \frac{(A - A_+)^2}{A_+} + \frac{1}{2} K_{A\mathrm{m}-} \frac{(A - A_-)^2}{A_-} \,, \tag{4.28}$$

which I trust we will by now feel immediately comfortable with. If we were to only account for this term, the membrane tension would be given by $\Sigma = \partial(G_\mathrm{el}/\partial A)$, which we can solve for the equilibrium bilayer area:

$$\frac{1}{A_\mathrm{eq}} = \frac{\alpha_+}{A_+} + \frac{\alpha_-}{A_-} \qquad \text{with} \quad \alpha_\pm = \frac{K_{A\mathrm{m}\pm}}{K_A + \Sigma} \,, \tag{4.29}$$

leading to the area-relaxed but tension-dependent elastic free energy

$$\boxed{G_\mathrm{el}(\Sigma) = \frac{1}{2} \frac{(A_+ - A_-)^2}{\frac{A_+}{K_{A\mathrm{m}+}} + \frac{A_-}{K_{A\mathrm{m}-}}} + \frac{1}{2} \frac{\Sigma^2}{\frac{K_{A\mathrm{m}+}}{A_+} + \frac{K_{A\mathrm{m}-}}{A_-}} \,.} \tag{4.30}$$

Notice that this is a quadratic that is minimal when $\Sigma = 0$ and when the imposed differential area strain vanishes, *i. e.*, if $A_+ = A_-$. Moreover, even if the net bilayer tension vanishes, the leaflet tensions are not zero as long as a differential area strain remains:

$$\Sigma_{\pm,0} = \left(\frac{\partial G_\mathrm{el}}{\partial A} \right)_{\substack{A = A_\mathrm{eq} \\ \Sigma = 0}} = \mp K_{A\mathrm{m}\pm} \frac{A_+ - A_-}{A_+ + A_-} \,. \tag{4.31}$$

We now need to determine how these areas are supposed to be connected to the lipid content in the leaflets. This we will do in the simplest possible way: just assuming area additivity. In this spirit, let us

[14] This is true far beyond membranes: molecular areas or volumes are fraught with failed intuitions, because what really matters is how molecules locally pack, and that depends on the neighbors they are packing with. A well known illustration of this is that 500 ml of water and 500 ml of (anhydrous!) ethanol do not add up to 1 l of an ethanol-water mixture but to about 10% less in volume.

then define the following variables:

$$\text{relaxed leaflet areas :} \qquad A_\pm = L_\pm a_\pm + C_\pm a_\mathrm{c} \ , \tag{4.32a}$$

$$\text{cholesterol area fraction :} \qquad \phi_\mathrm{a} = \frac{C a_\mathrm{c}}{L_+ a_+ + L_- a_- + C a_\mathrm{c}} \ , \tag{4.32b}$$

$$\text{phospholipid area difference :} \qquad \Delta A_- = L_+ a_+ - L_- a_- \ , \tag{4.32c}$$

$$\text{cholesterol asymmetry :} \qquad \delta c = \frac{C_+ - C_-}{C_+ + C_-} \ , \tag{4.32d}$$

where $L_\pm$ and $C_\pm$ are the number of phospho**L**ipids and **C**holesterol in the two leaflets (and $C = C_+ + C_-$), and where $a_\pm$ are the areas of the phospholipids and a_c is the cholesterol area. Using these definitions, and the leaflet stresses defined from Eqn. (4.31), we find the differential stress at zero net tension

$$\boxed{\Delta\Sigma_0 = \Sigma_{+,0} - \Sigma_{-,0} = -K_A \phi_\mathrm{a} \left[\frac{\Delta A}{C a_\mathrm{c}} + \delta c \right] \ .} \tag{4.33}$$

This nicely shows the two effects contributing to it: the total phospholipid leaflet areas can differ, $\Delta A \neq 0$, or cholesterol can distribute unevenly between the leaflets, $\delta c \neq 0$.

So much for the elastic part. How about mixing? For this, we use an approximate description that accounts for an ideal gas entropy of mixing and a non-ideal interaction term that is proportional to the product of the two concentration of the species that mix. This is moderately straightforward, because in each leaflet we only have two components: whatever phospholipid species was put there in the first place, and whatever amount of cholesterol happens to end up there. If for such binary mixtures (component $i \in \{1, 2\}$) we call the molecular numbers N_i, areas a_i, total area $A_{12} = N_1 a_1 + N_2 a_2$, area fractions $\phi_i = N_i a_i / A_{12}$, the non-ideal solution free energy can be expressed as

$$\boxed{\beta G_\mathrm{sol} = N_1 \log \phi_1 + N_2 \log \phi_2 + \chi \, N_1 N_2 \frac{\sqrt{a_1 a_2}}{A_{12}} \ ,} \tag{4.34}$$

where the dimensionless coupling constant "χ" multiplying the non-ideal part is usually called the "Flory-Huggins mixing parameter." In our situation, we will end up with two such parameters, $\chi_\pm$, one for the nonideal mixing between cholesterol and the phospholipids "native" to each leaflet. It will then be useful to consider a situation where we

describe the extent to which the leaflets are different by how strongly their χ-parameters differ. This is most naturally done by defining

$$\left.\begin{aligned} \bar{\chi} &= \tfrac{1}{2}(\chi_+ + \chi_-) \\ \delta\chi &= \chi_+ - \chi_- \end{aligned}\right\} \quad \longleftrightarrow \quad \chi_\pm = \bar{\chi} \pm \frac{1}{2}\delta\chi \ . \tag{4.35}$$

The next steps are conceptually easy but algebraically very tedious. The complete free energy is simply the sum of the elastic term (4.30) and a mixing term of the form (4.34) for each leaflet. It is already area minimized, so the only other equilibrium condition we need to impose is the equality of cholesterol's chemical potential between the leaflets, à la Eqn. (4.26). Sadly, though, the presence of the log-terms makes the resulting algebra terribly unwieldy, and in order to come up with manageable formulas that we can stare at and say "aha," we need to expand around small deviations from the symmetric state. What follows in (Varma and Deserno, 2022) is some tedious gymnastics that is straightforward, boring, and unfortunately tends to hide the simplicity of the underlying idea. I will skip it entirely. Instead, I will show you a few highlights you find if you have the patience to slog through it by hand or, more advisable, let MATHEMATICA® help you with all those tedious expansions.

In fact, I will make one more approximation, which is not very good, but massively cleans up all formulas. It gives rise to results that are almost surely *quantitatively off* in the real world but still capture most qualitative effects—which shall be good enough for now. This approximation is the "equal area assumption," which takes the area of all lipids to be the same: $a_+ = a_- = a_\mathrm{c} \equiv a$. It is not too bad for the phospholipids, but evidently highly dubious for cholesterol. Recall, though, that we have already been quite slap-dash with cholesterol's area, assuming additivity, which we know to be wrong. Still, I will put the reminder "e.a." above equations relying on this approximation, just to be extra clear.

If areas are the same, some of the definitions in Eqns. (4.32) become simpler, and it also makes sense to define two others:

$$\text{cholesterol mole fraction :} \qquad \phi = \frac{C}{L_+ + L_- + C} \ , \tag{4.36a}$$

$$\text{phospholipid asymmetry :} \qquad \delta\ell = \frac{L_+ - L_-}{L_+ + L_-} \ , \tag{4.36b}$$

$$\text{scaled stretching modulus :} \qquad \tilde{K}_A = \beta K_A a \ , \tag{4.36c}$$

$$\text{scaled differential stress :} \qquad \Delta\tilde{\Sigma} = \Delta\Sigma / K_A \ . \tag{4.36d}$$

Notice that if we again take $K_A \approx 240\,\mathrm{mN/m}$ (Rawicz et al., 2000) and pick $a \approx 0.55\,\mathrm{nm}^2$ as a "compromise" for all lipid areas, we find that $\tilde{K}_A \approx 32$.

One of the key questions we wish to answer is: *where is the cholesterol*? In terms of our newly defined variables: how does the cholesterol asymmetry δc depend on the imposed phospholipid asymmetry $\delta\ell$ and a possible cholesterol partitioning bias $\delta\chi$ between the leaflets? After some serious cranking and approximating, the answer turns out to be

$$\delta c(\delta\chi, \delta\ell) \overset{\text{e.a.}}{\approx} -\frac{(1-\phi)^2\delta\chi + (\tilde{K}_A - 2)(1-\phi)\delta\ell}{2 + (\tilde{K}_A - 2)\phi} \,. \tag{4.37}$$

Increasing $\delta\chi$ (which yanks cholesterol into the $-$ leaflet) and increasing $\delta\ell$ (which puts more phospholipids into the $+$ leaflet) both pushes cholesterol out of the $+$ leaflet. Both effects are linear in the drivers, which is a consequence of all the expansions leading up to this point. The two effects can compensate if $\delta\chi$ and $\delta\ell$ have opposite signs, and they compensate "fully" (meaning: cholesterol now distributes evenly) when $\delta\chi/\delta\ell = -(\tilde{K}_A - 2)/(1-\phi)$. At 25% cholesterol this means an extra $0.4\,k_\mathrm{B}T$ free energy difference for cholesterol between the sides for every extra percent of phospholipid asymmetry. Also notice that this equation has a very simple limit in the case where elasticity completely outcompetes the other two drivers, *i. e.*, where differential stress is the only thing that matters: all we need to do is to set $\tilde{K}_A \to \infty$ and find $\delta c = (\phi^{-1} - 1)\delta\ell$, which you can easily check corresponds to $L_+ + C_+ = L_- + C_-$. This means that lipids will distribute such that both leaflets take the same area, and hence the differential stress vanishes.

The second key question to ask is: what is the differential stress in the system, at a given phospholipid asymmetry and partitioning bias? Here, the answer turns out to be

$$\Delta\Sigma(\delta\chi, \delta\ell) \overset{\text{e.a.}}{\approx} K_A \frac{\frac{1}{2}\phi(1-\phi)\delta\chi - \delta\ell}{1 + \frac{1}{2}\tilde{K}_A \frac{\phi}{1-\phi}} \,. \tag{4.38}$$

Just like the cholesterol asymmetry, the differential stress again depends linearly on the two drivers $\delta\chi$ and $\delta\ell$, and the directions make sense. The stress vanishes when $\delta\chi/\delta\ell = 2/[\phi(1-\phi)]$, which unlike the asymmetry diverges at $\phi = 0$ because we have so little cholesterol to repair the situation.

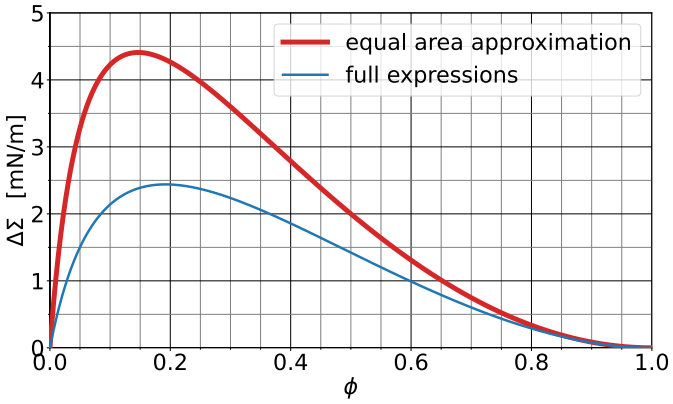


Figure 4.6 | Differential stress in a membrane with no area imbalance ($\Delta A = 0$) as a function of cholesterol mole fraction ϕ. We assume $K_A = 240\,\mathrm{mN/m}$ and $\delta\chi = 1$ (*i. e.*, a $1\,k_\mathrm{B}T$ preference for cholesterol to go into the lower leaflet). The thick red curve is the equal area approximation from Eqn. (4.38), taking $\delta\ell = 0$ as "no area imbalance" and an average lipid area of $a = 0.55\,\mathrm{nm}^2$, which gives $\tilde{K}_A = 32$. The thinner blue curve instead uses the more precise answer using the unequal lipid areas $a_+ = a_- = 0.6\,\mathrm{nm}^2$ and $a_\mathrm{c} = 0.35\,\mathrm{nm}^2$.

An interesting prediction is that even in the case of no packing imbalance, $\delta\ell = 0$, when the differential stress vanishes in the absence of cholesterol, the *addition* of cholesterol can *create* differential stress if there is a partitioning bias $\delta\chi \neq 0$. This is literally the opposite of the original expectation that adding cholesterol would always cancel differential stress; here it creates stress that wasn't there to begin with. How big is this effect? Fig. 4.6 plots this for the parameter choices we have been working with all along. At $\phi = 0$ the differential stress indeed vanishes, but the moment we add cholesterol, it shoots up rather rapidly. In the limit $\phi \to 1$ the stress again goes back to zero.[15] The interesting thing is how much it shoots up in between. At 15% it reaches a remarkably large value in excess of $4\,\mathrm{mN/m}$, even if there's just a $1\,k_\mathrm{B}T$ difference in cholesterol's free energy of partitioning between the leaflets. If we do the calculation a bit more carefully and forgo the equal area approximation, the effect is a bit smaller (because the cholesterol molecules we yank from one leaflet into the other one are smaller and hence create less strain), but it is still in the few mN/m

[15] This is a bit unphysical, because membranes cannot be made from 100% cholesterol; but the theory doesn't know about this, and we are allowed to be happy about the fact that it makes a reasonable extrapolation to what *would* happen if such membranes were stable (or, maybe more realistically, what if one particular lipid species would be able to flip flop rapidly, possibly with the assistance of some enzymes ("scramblases")).

ballpark. And, obviously, it gets proportionally larger if $\delta\chi$ increases. This, at long last, is my other argument I promised for why we should expect differential stress to be nonzero.

Heads up 8

Promise fulfilled! Another reason for why $\Delta\Sigma$ generally has sizable values.

4.4 Things to think about

1. A bending energy of the form $\frac{1}{2}\kappa(J - J_{0\mathrm{b}})^2$ can also be defined on the *leaflet* level, with individual monolayer bending rigidities $\kappa_{\mathrm{m}\pm}$ and spontaneous monolayer curvatures $J_{0\mathrm{m}\pm}$. Unlike $J_{0\mathrm{b}}$, the monolayer values $J_{0\mathrm{m}\pm}$ are generally nonzero, because there's no up-down symmetry in a single leaflet that would enforce this. Assuming that the two leaflets can slide past each other when the bilayer is bent, and ignoring stretching issues for this pure bending question, show that the bending-associated spontaneous curvature $J_{0\mathrm{b}}$ can be written as a function of the monolayer elastic parameters!
   (Hints: (1) The subtle difference in curvature a distance z_0 out or in from the bilayer midplane will turn into a higher order effect, so you can ignore it for this question. (Or you can check that this is in fact so.) (2) There is a subtle minus sign hiding in this question. As a control: do you get the expected answer when $J_{0\mathrm{m}+} = J_{0\mathrm{m}-}$?)

2. One of the hallmarks of membrane asymmetry are the two measures of spontaneous curvature appearing in Eqn. (4.19), $J_{0\mathrm{b}}$ and $J_{0\mathrm{s}}$, and the fact that they can be different from one another. Show that, sadly, both of them drop out of the physics in linear Monge gauge under periodic boundary conditions. This for instance means that a standard fluctuation analysis would not be able to learn anything about either of these values.

3. Consider an asymmetric membrane which has the lipid POPC on one side and the lipid POPE on the other. From the computer simulations of Venable et al. (2015) we take that they have virtually identical monolayer bending moduli $\kappa_{\mathrm{m}} \approx 15\,k_{\mathrm{B}}T$ but very different leaflet spontaneous curvatures: $J_{0\mathrm{m}}(\mathrm{POPC}) = -0.032\,\mathrm{nm}^{-1}$ and $J_{0\mathrm{m}}(\mathrm{POPE}) = -0.213\,\mathrm{nm}^{-1}$. Using the formula you might have derived in problem 1, this gives a bilayer curvature of $J_{0\mathrm{b}} = 0.09\,\mathrm{nm}^{-1}$ using the convention that POPC is "up" or "outside". Let's think a bit about vesicles we could make from such strongly asymmetric membranes.

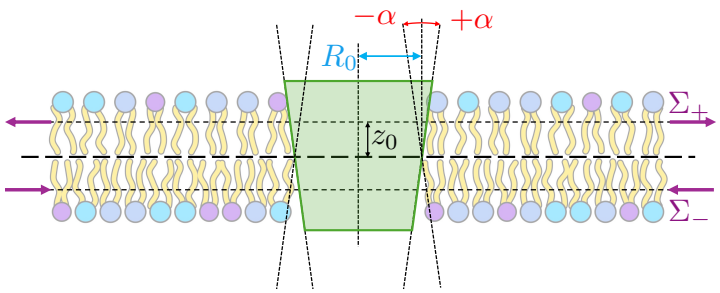


Figure 4.7 | Transmembrane protein embedded in a differentially stressed membrane. The indicated conformational change, which flips the truncated cone angle from $+\alpha$ to $-\alpha$, requires work to be done against the differential stress.

a) What's the radius of a vesicle in which both the overall torque and the differential stress are relaxed? On which side is POPE?

b) Eicher et al. (2018) succeeded in making vesicles with this type of lipidomic asymmetry, but with radii of $R \approx 60\,\mathrm{nm}$. In fact, they made them in both variants: $\mathrm{POPC^{out}/POPE^{in}}$ as well as $\mathrm{POPE^{out}/POPC^{in}}$. If we assume that the formation process selects the number of lipids on both side such that the overall bending rigidity from Eqn. (4.19) is minimal, what would be the differential stress for these two types of vesicles? Which side is under tension?

4. Some transmembrane proteins (such as ABC transporters) undergo conformational transitions that change their cross-sectional area in a way that differs between the two leaflets. As a toy example, consider the situation illustrated in Fig. 4.7: a protein shaped like a truncated cone "pivots" its orientation at the bilayer midplane in such a way that the cone angle changes from $+\alpha$ to $-\alpha$. If the membrane is under a differential stress $\Delta\Sigma$, what work needs to be done (or is done) due to this motion? Pick numbers for all the relevant parameters of the problem that strike you as biophysically plausible and express this free energy change in units of $k_\mathrm{B}T$. Do you think this could be relevant?

Bibliography

D. W. Allender, A. J. Sodt, and M. Schick. Cholesterol-dependent bending energy is important in cholesterol distribution of the plasma membrane. *Biophys. J.*, 116(12):2356–2366, 2019.

W. F. D. Bennett, J. L. MacCallum, M. J. Hinner, S. J. Marrink, and D. P. Tieleman. Molecular view of cholesterol flip-flop and chemical potential in different membrane environments. *J. Am. Chem. Soc.*, 131(35):12714–12720, 2009. doi: 10.1021/ja903529f.

A. A. Benson. On the orientation of lipids in chloroplast and cell membranes. *J. Am. Oil Chem.' Soc.*, 43(5):265–270, 1966.

O. Berger, O. Edholm, and F. Jähnig. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.*, 72(5):2002–2013, 1997.

M. S. Bretscher. Asymmetrical lipid bilayer structure for biological membranes. *Nat. New Biol.*, 236(61):11–12, 1972.

P. B. Canham. The minimum energy of bending as a possible explanation of the biconcave shape of the human red blood cell. *J. Theor. Biol.*, 26(1):61–81, 1970.

P. M. Chaikin and T. C. Lubensky. *Principles of condensed matter physics*, volume 10. Cambridge university press Cambridge, 1995.

K. H. Cheng, B. Cannon, J. Metze, A. Lewis, J. Huang, M. W. Vaughn, Q. Zhu, P. Somerharju, and J. Virtanen. Lipid headgroup superlattice modulates the activity of surface-acting cholesterol oxidase in ternary phospholipid/cholesterol bilayers. *Biochemistry*, 45(36):10855–10864, 2006.

I. R. Cooke and M. Deserno. Coupling between lipid shape and membrane curvature. *Biophys. J.*, 91(2):487–495, 2006.

I. R. Cooke, K. Kremer, and M. Deserno. Tunable Generic Model for Fluid Bilayer Membranes. *Phys. Rev. E*, 72(1):011506, 2005.

J. F. Danielli and H. Davson. A contribution to the theory of permeability of thin films. *J. Cell. Comp. Physiol.*, 5(4):495–508, 1935.

R. W. R. Darling. *Differential Forms and Connections*. Cambridge University Press, Cambridge, 1 edition, 1994.

M. Deserno. Fluid lipid membranes: From differential geometry to curvature stresses. *Chem. Phys. Lipids.*, 185:11–45, 2015. doi: 10.1016/j.chemphyslip.2014.05.001.

M. Deserno. Biomembranes balance many types of leaflet asymmetries. *Curr. Opin. Struct. Biol.*, 87:102832, 2024.

M. Deserno, M. M. Müller, and J. Guven. Contact lines for fluid surface adhesion. *Phys. Rev. E*, 76(1):011605, 2007.

K. A. Dill and P. J. Flory. Molecular organization in micelles and vesicles. *Proc. Nat. Acad. Sci. (USA)*, 78(2):676–680, 1981.

M. do Carmo. *Differential Geometry of Curves and Surfaces*. Prentice Hall, Englewood Cliffs, NJ, 1976.

M. do Carmo. *Riemannian Geometry*. Birkhäuser, Boston, MA, 1992.

B. Eicher, D. Marquardt, F. A. Heberle, I. Letofsky-Papst, G. N. Rechberger, M.-S. Appavou, J. Katsaras, and G. Pabst. Intrinsic curvature-mediated transbilayer coupling in asymmetric lipid vesicles. *Biophys. J.*, 114(1):146–157, 2018.

Y. Elani, S. Purushothaman, P. J. Booth, J. M. Seddon, N. J. Brooks, R. V. Law, and O. Ces. Measurements of the effect of membrane asymmetry on the mechanical properties of lipid bilayers. *Chem. Comm.*, 51(32):6976–6979, 2015.

E. Evans and V. Heinrich. Dynamic strength of fluid membranes. *Comptes Rendus Physique*, 4(2):265–274, 2003.

E. A. Evans. Bending resistance and chemically induced moments in membrane bilayers. *Biophys. J.*, 14(12):923–931, 1974.

E. A. Evans. Minimum energy analysis of membrane deformation applied to pipet aspiration and surface adhesion of red blood cells. *Biophys. J.*, 30(2):265–284, 1980.

H. Flanders. *Differential Forms with Applications to the Physical Sciences*. Dover, New York, 1989.

S. L. Foley and M. Deserno. Quantifying uncertainty in transmembrane stresses and moments in simulation. In M. Deserno and T. Baumgart, editors, *Biophysical Approaches for the Study of Membrane Structure—Part B: Theory and Simulations*, volume 701 of *Methods in Enzymology*, chapter 3, pages 83–122. Elsevier, 2024.

S. L. Foley, M. Varma, A. Hossein, and M. Deserno. Elastic and thermodynamic consequences of lipid membrane asymmetry. *Emerging Topics in Life Sciences*, 7(1):95–110, 2023.

T. Frankel. *The Geometry of Physics: An Introduction*. Cambridge University Press, New York, 3 edition, 2004.

S. Garg, L. Porcar, A. C. Woodka, P. D. Butler, and U. Perez-Salas. Noninvasive neutron scattering measurements reveal slower cholesterol transport in model lipid membranes. *Biophys. J.*, 101(2):370–377, 2011.

J. A. Hamilton. Fast flip-flop of cholesterol and fatty acids in membranes: implications for membrane transport proteins. *Curr. Opin. Lipidol.*, 14(3):263–271, 2003. doi: 10.1097/00041433-200306000-00006.

W. Helfrich. Elastic properties of lipid bilayers: theory and possible experiments. *Z. Naturforsch. C*, 28(11-12):693–703, 1973.

W. Helfrich. The size of bilayer vesicles generated by sonication. *Phys. Lett. A*, 50 (2):115–116, 1974a.

W. Helfrich. Blocked lipid exchange in bilayers and its possible influence on the shape of vesicles. *Z. Naturforsch. C*, 29(9-10):510–515, 1974b.

A. Hossein and M. Deserno. Spontaneous curvature, differential stress, and bending modulus of asymmetric lipid membranes. *Biophys. J.*, 118(3):624–642, 2020.

J. N. Israelachvili, D. J. Mitchell, and B. W. Ninham. Theory of self-assembly of hydrocarbon amphiphiles into micelles and bilayers. *J. Chem. Soc., Farad. Transact. 2: Molecular and Chemical Physics*, 72:1525–1568, 1976.

H. M. James and A. S. Coolidge. The ground state of the hydrogen molecule. *J. Chem. Phys.*, 1(12):825–835, 1933.

R. A. L. Jones. *Soft Condensed Matter*. Oxford University Press, Oxford, 2002.

E. Kreyszig. *Differential Geometry*. Dover, New York, 1991.

M. Krompers and H. Heerklotz. A guide to your desired lipid-asymmetric vesicles. *Membranes*, 13(3):267, 2023.

N. Kučerka, S. Tristram-Nagle, and J. F. Nagle. Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains. *J. Mem. Biol.*, 208(3):193–202, 2006.

L. D. Landau and E. M. Lifshitz. Theory of elasticity, vol. 7. *Course of Theoretical Physics*, 3, 1986.

F. Leeb and L. Maibaum. Spatially resolving the condensing effect of cholesterol in lipid bilayers. *Biophys. J.*, 115(11):2179–2188, 2018. doi: 10.1016/j.bpj.2018.10.024.

R. Leventis and J. R. Silvius. Use of cyclodextrins to monitor transbilayer movement and differential lipid affinities of cholesterol. *Biophys. J.*, 81(4):2257–2267, 2001.

R. Lipowsky. Multiscale remodeling of biomembranes and vesicles. In M. Deserno and T. Baumgart, editors, *Biophysical Approaches for the Study of Membrane Structure—Part B: Theory and Simulations*, volume 701 of *Methods in Enzymology*, chapter 6, pages 175–236. Elsevier, 2024.

R. Lipowsky and H.-G. Döbereiner. Vesicles in contact with nanoparticles and colloids. *Europhys. Lett.*, 43(2):219, 1998.

J. H. Lorent, K. R. Levental, L. Ganesan, G. Rivera-Longsworth, E. Sezgin, M. Doktorova, E. Lyman, and I. Levental. Plasma membranes are asymmetric in lipid unsaturation, packing and protein shape. *Nat. Chem. Biol.*, 16(6):644–652, 2020.

D. Lovelock and H. Rund. *Tensors, Differential Forms, and Variational Principles.* Dover, New York, 1989.

C. G. Malmberg and A. A. Maryott. Dielectric constant of water from 00 to 1000 c. *J. Res. Natl. Bur. Stand.*, 56(1):1–8, 1956.

S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. De Vries. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B*, 111(27):7812–7824, 2007.

L. Miao, U. Seifert, M. Wortis, and H.-G. Döbereiner. Budding transitions of fluid-bilayer vesicles: The effect of area-difference elasticity. *Phys. Rev. E*, 49(6):5389, 1994.

M. S. Miettinen and R. Lipowsky. Bilayer membranes with frequent flip-flops have tensionless leaflets. *Nano Letters*, 19(8):5011–5016, 2019.

C. E. Morris and U. Homann. Cell surface area regulation and membrane tension. *J. Membr. Biol.*, 179:79–102, 2001.

O. G. Mouritsen and M. J. Zuckermann. What's so special about cholesterol? *Lipids*, 39(11):1101–1113, 2004.

W. D. Nes. Biosynthesis of cholesterol and other sterols. *Chem. Rev.*, 111(10): 6423–6451, 2011. doi: 10.1021/cr200021m.

T. Portet and R. Dimova. A new method for measuring edge tensions and stability of lipid bilayers: effect of membrane composition. *Biophys. J.*, 99(10):3264–3273, 2010.

W. Rawicz, K. C. Olbrich, T. McIntosh, D. Needham, and E. Evans. Effect of chain length and unsaturation on elasticity of lipid bilayers. *Biophys. J.*, 79(1): 328–339, 2000.

J. D. Robertson. The ultrastructure of cell membranes and their derivatives. In *Biochemical society symposium*, volume 16, pages 3–43, 1959.

B. Różycki and R. Lipowsky. Spontaneous curvature of bilayer membranes from molecular simulations: Asymmetric lipid densities and asymmetric adsorption. *J. Chem. Phys.*, 142(5), 2015.

A. Sandra and R. E. Pagano. Phospholipid asymmetry in lm cell plasma membrane derivatives: polar head group and acyl chain distributions. *Biochemistry*, 17(2): 332–338, 1978.

P. K. Schick, K. B. Kurica, and G. K. Chacko. Location of phosphatidylethanolamine and phosphatidylserine in the human platelet plasma membrane. *J. Clin. Investig.*, 57(5):1221–1226, 1976.

B. Schutz. *Geometric methods in mathematical physics.* Cambridge University Press, 1980.

U. Seifert, L. Miao, H.-G. Döbereiner, and M. Wortis. Budding transition for bilayer fluid vesicles with area-difference elasticity. In R. Lipowsky, D. Richter, and K. Kremer, editors, *The Structure and Conformation of Amphiphilic Membranes: Proceedings of the International Workshop on Amphiphilic Membranes, Jülich, Germany, September 16–18, 1991*, pages 93–96. Springer, 1992.

S. Shukla and T. Baumgart. Enzymatic trans-bilayer lipid transport: Mechanisms, efficiencies, slippage, and membrane curvature. *Biochimica et Biophysica Acta (BBA)—Biomembranes*, 1863(3):183534, 2021.

J. R. Silvius. Role of cholesterol in lipid raft formation: lessons from lipid model systems. *Biochim. Biophys. Acta (BBA)—Biomembranes*, 1610(2):174–183, 2003.

S. J. Singer and G. L. Nicolson. The fluid mosaic model of the structure of cell membranes: Cell membranes are viewed as two-dimensional solutions of oriented globular proteins and lipids. *Science*, 175(4023):720–731, 1972.

R. Smith and C. Tanford. The critical micelle concentration of l-α-dipalmitoylphosphatidylcholine in water and water/methanol solutions. *J. Mol. Biol.*, 67(1):75–83, 1972.

A. J. Sodt, R. M. Venable, E. Lyman, and R. W. Pastor. Nonadditive compositional curvature energetics of lipid bilayers. *Phys. Rev. Lett.*, 117(13):138104, 2016.

M. M. Sperotto and A. Ferrarini. Spontaneous lipid flip-flop in membranes: a still unsettled picture from experiments and simulations. In R. M. Epand and J.-M. Ruysschaert, editors, *The Biophysics of Cell Membranes; Biological Consequences*, pages 29–60. Springer, New York, NY, 2017.

M. Spivak. *A Comprehensive Introduction to Differential Geometry*, volume 2. Publish or Perish, Boston, MA, 1970.

M. Spivak. *A Comprehensive Introduction to Differential Geometry*, volume 3. Publish or Perish, Boston, MA, 1975a.

M. Spivak. *A Comprehensive Introduction to Differential Geometry*, volume 4. Publish or Perish, Boston, MA, 1975b.

T. L. Steck and Y. Lange. How slow is the transbilayer diffusion (flip-flop) of cholesterol? *Biophys. J.*, 102(4):945–946, 2012.

T. L. Steck and Y. Lange. Transverse distribution of plasma membrane bilayer cholesterol: Picking sides. *Traffic*, 19(10):750–760, 2018.

T. L. Steck, J. Ye, and Y. Lange. Probing red cell membrane cholesterol movement with cyclodextrin. *Biophys. J.*, 83(4):2118–2125, 2002.

S. Svetina, M. Brumen, and B. Žekš. Lipid bilayer elasticity and the bilayer couple interpretation of red cell shape transformations and lysis. *Stud. Biophys.*, 110 (1-3):177–184, 1985.

M. M. Terzi and M. Deserno. Lipid membranes: From self-assembly to elasticity. In D. J. Steigmann, editor, *The Role of Mechanics in the Study of Lipid Bilayers*, pages 105–166. Springer, Cham, Switzerland, 2018.

N. J. Turro and A. Yekta. Luminescent probes for detergent solutions. a simple procedure for determination of the mean aggregation number of micelles. *J. Am. Chem. Soc.*, 100(18):5951–5952, 1978.

M. Varma and M. Deserno. Distribution of cholesterol in asymmetric membranes driven by composition and differential stress. *Biophys. J.*, 121(20):4001–4018, 2022.

R. M. Venable, F. L. H. Brown, and R. W. Pastor. Mechanical properties of lipid bilayers from molecular dynamics simulation. *Chem. Phys. Lipids*, 192:60–74, 2015.

A. J. Verkleij, R. F. A. Zwaal, B. Roelofsen, P. Comfurius, D. Kastelijn, and L. L. M. Van Deenen. The asymmetric distribution of phospholipids in the human red cell membrane. a combined study using phospholipases and freeze-etch electron microscopy. *Biochimica et Biophysica Acta (BBA)—Biomembranes*, 323(2):178–193, 1973.

X. Wang and M. Deserno. Determining the lipid tilt modulus by simulating membrane buckles. *J. Phys. Chem. B*, 0(0):null, 2016. doi: 10.1021/acs.jpcb.6b02016.

T. J. Willmore. *An Introduction to Differential Geometry.* Dover, New York, 2012.

T. A. Witten and P. A. Pincus. *Structured fluids: polymers, colloids, surfactants.* Oxford University Press, Oxford, 2004.