Making sense of the Legendre transform

R. K. P. Zia
Department of Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

Edward F. Redish
Department of Physics, University of Maryland, College Park, Maryland 20742

Susan R. McKay
Department of Physics and Astronomy, University of Maine, Orono, Maine 04469

(Received 27 September 2007; accepted 26 March 2009)

The Legendre transform is a powerful tool in theoretical physics and plays an important role in classical mechanics, statistical mechanics, and thermodynamics. In typical undergraduate and graduate courses the motivation and elegance of the method are often missing, unlike the treatments frequently enjoyed by Fourier transforms. We review and modify the presentation of Legendre transforms in a way that explicates the formal mathematics, resulting in manifestly symmetric equations, thereby clarifying the structure of the transform. We then discuss examples to motivate the transform as a way of choosing independent variables that are more easily controlled. We demonstrate how the Legendre transform arises naturally from statistical mechanics and show how the use of dimensionless thermodynamic potentials leads to more natural and symmetric relations. © *2009 American Association of Physics Teachers.*
[DOI: 10.1119/1.3119512]

I. INTRODUCTION

The Legendre transform is commonly used in upper division and graduate physics courses, especially in classical mechanics,[1] statistical mechanics, and thermodynamics.[2,3] Most physics majors are first exposed to the Legendre transform in classical mechanics, where it provides the connection between the Lagrangian $\mathcal{L}(\dot{q})$ and the Hamiltonian $\mathcal{H}(p)$, and then in statistical mechanics, where it yields relations between the internal energy E and the various thermodynamic potentials. Despite its common use, the Legendre transform often appears in an ad hoc fashion, without being presented as a general and powerful mathematical tool as is done for the Fourier transform.

In this paper, we present a pedagogical introduction to the Legendre transform, discuss it as a mathematical process, and display some of its general properties. Because some students prefer algebraic approaches and others prefer geometric ones, we discuss the transform from both points of view and relate them. We then motivate the transform in terms related to physical conditions and constraints. We emphasize some of the symmetries and structures of the transform and present a series of increasingly complex examples beginning with classical mechanics and then in statistical mechanics. We end with some remarks on more general versions of the Legendre transform as well as other areas in which it is widely used.

II. THE LEGENDRE TRANSFORM AS AN ALTERNATIVE WAY TO DISPLAY INFORMATION

Many students can manage the rules for generating a Hamiltonian from a Lagrangian or switching between thermodynamic potentials, but express discomfort when asked about the Legendre transform as a general mathematical tool. One reason is that in introductory physics we often treat a function as a relation between physical rather than mathematical quantities. When we think about physical functions, we tend not to pay attention to the particular functional form the mathematical function uses to encode physical information.[4] For example, if we are describing a position as a function of time, we might write it as $x(t)$. We do not bother to change the symbol x if we decide to give t in milliseconds instead of in seconds. If we write the temperature as a function of position as $T(\vec{r})$, we do not change the symbol if we switch to a different coordinate system or measuring scale. In contrast, the Legendre transform is explicitly about how information is coded in the functional form.

Students are usually first introduced to the Legendre transform as the transformation from the Lagrangian $\mathcal{L}$ to the Hamiltonian $\mathcal{H}$. This transformation involves the switch from the velocity to the momentum variable in the nonrelativistic kinetic energy T. In the context of nonrelativistic particle motion with velocity independent potentials, the transform involves the kinetic energy, the most trivial function to which the Legendre transform can be applied. The result looks like a shift from v to mv as an independent variable, so that it seems pointless. Because the position variable q plays no role in the transform and typically appears only in the potential energy V, the result is often regarded as a mysterious change of the sign of V: $\mathcal{L}=T-V$ versus $\mathcal{H}=T+V$.

In the rest of this section, we motivate the Legendre transform as a general mathematical transformation and describe a method that displays its general properties and symmetries.

For simplicity, we begin with a single variable x. Generally, a function expresses a relation between two parameters: an independent variable or control parameter x and a dependent value F. This information is encoded in the functional form of $F(x)$. For later convenience, we will also denote such a relation or "encoding" as $\{F,x\}$.

In some circumstances, it is useful to encode the information contained in the function $F(x)$ differently. Two common examples are the Fourier transform and the Laplace transform. These transforms express the function F as sums of (complex or real) exponentials and display the information in F in terms of the amount of each component contained in the function rather than in terms of the value of the function. We

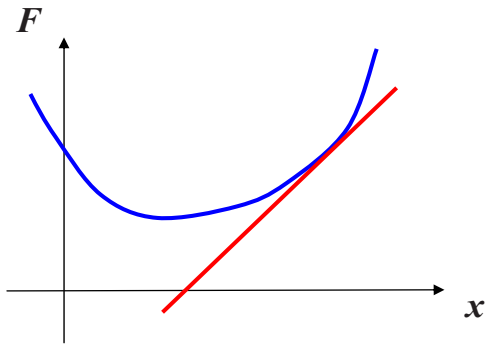

Fig. 1. (Color online) The graph of a convex function $F(x)$. The tangent line at one point is illustrated.

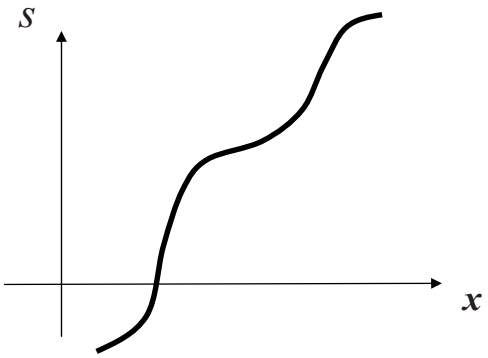

Fig. 2. The graph of $s(x)$, the slope of a convex function.

say that $\{\tilde{F},k\}$ encodes the same information as $\{F,x\}$. For the Fourier transform, $\tilde{F}(k)\equiv\int e^{ikx}F(x)dx$ is an explicit "transformation" between the two encodings.

Given an $F(x)$, the Legendre transform provides a more convenient way of encoding the information in the function when two conditions are satisfied: (1) The function (or its negative) is strictly convex (second derivative always positive) and smooth (existence of "enough" continuous derivatives). (2) It is easier to measure, control, or think about the derivative of F with respect to x than it is to measure or think about x itself.

Because of condition (1), the derivative of $F(x)$ with respect to x can serve as a stand in for x; that is, there is a one-to-one mapping between x and dF/dx. (We remark on relaxing this condition in Sec. VII.) The Legendre transform shows how to define a function that contains the same information as $F(x)$ but as a function of dF/dx.

III. THE MATHEMATICS OF THE LEGENDRE TRANSFORM

We first consider a single, smooth convex function of a single variable. There are many equivalent ways to characterize convex functions. The most convenient one is that the second derivative $d^2F(x)/dx^2$ is always positive. Another characterization of this condition is that the slope function

$$s(x)\equiv\frac{dF(x)}{dx} \tag{1}$$

is a strictly monotonic function of x (because this characterization also permits us to treat functions whose negative is convex).

A graphical way to see how the value of x and the slope of a convex function can stand in for each other can be seen by considering the example in Fig. 1, where the curve drawn to represent F is convex. As we move along the curve to the right (as x increases), the slope of the tangent to the curve continually increases. In other words, if we were to graph the slope as a function of x, it would be a smoothly increasing curve, such as the example in Fig. 2. If the second derivative $d^2F(x)/dx^2$ exists (everywhere within the range of x in which F is defined; part of the condition for a smooth F), there is a unique value of the slope for each value of x, and vice versa. The corresponding mathematical language is that there is a one to one relation between s and x; that is, the function $s(x)$ is single-valued and can be inverted to give a single-valued function $x(s)$.

In this way, we can then start with s as the independent variable, use the inverse function to obtain an unique value of x, and then insert that into $F(x)$ to obtain F as a function of s. The standard notation for such a function is $F(x(s))$. If we insist on a new encoding of the information in F (in terms of s instead of x), this straightforward "function of a function approach" would appear to be the most natural way.

Instead, the Legendre transform of $F(x)$ is defined differently and in a seemingly unnatural way:

$$G(s)\equiv sx(s)-F(x(s)). \tag{2}$$

Typically, this definition is presented with little motivation or explanation, and leaves the students to ponder: Why? Why the extra sx? Why the minus sign? Frequently, the instructor or textbook invokes another magical relation to answer such queries. Only with this peculiar definition can we have the property that "the slope of $G(s)$ is just x":

$$x(s)=\frac{dG}{ds}. \tag{3}$$

This result requires a careful calculation.

A. A graphic-geometric approach

Before providing ways to appreciate this definition of the Legendre transform, as well as how never to forget "which sign goes where," we present a graphical route to the transform. Consider the plot of F versus x in Fig. 3. Choose a value of x, which is represented by the length of the horizontal line labeled by x. Go up to the value on the function curve, $F(x)$. This value corresponds to the length of the vertical line labeled by F. Next, draw the tangent to the curve at that point. The slope here is labeled s, as emphasized by the call out bubble. Extend this tangent until it hits the ordinate (the "F axis"). In this example, the intercept is negative and is labeled as $-G$, with G positive. This value corresponds to the length of the thick vertical line labeled by G. This length is reproduced (thin line) just below the line labeled F. Because the slope of the tangent is s, the length of the dotted vertical line is sx. From this picture, it is clear that $sx=F+G$. In this interpretation, the peculiar definition of the Legendre transform in Eq. (2) appears natural. The minus sign in the definition is seen as a way of retaining the symmetry and

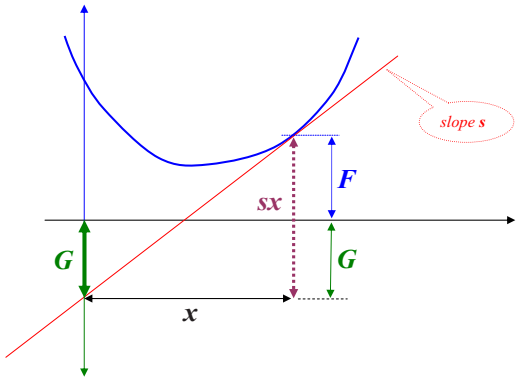

Fig. 3. (Color online) Graphic representation of the Legendre transform, $G(s)$, of $F(x)$ (Ref. 5). See the text for an explanation of the various quantities.

simplicity of the geometrical statement: In the triangle, the slope (tangent) times the adjacent side equals the opposite side, which is the sum of F and G.

B. Symmetric representation of the Legendre transform

This symmetric geometrical construction allows us to display a number of useful and elegant relations that shed light on the workings of the Legendre transform. In particular, we consider the symmetries associated with the inverse Legendre transform, extreme values, and derivative relations.

Ordinarily, the inverse of a transformation is distinct from the transform itself. For example, an inverse Laplace transform is not given by the same formula. The Legendre transform distinguishes itself in that it is its own inverse. In this sense, it resembles (geometric) duality transformations. Symbolically, we may denote this relation as:

$$\{F,x\} \Leftrightarrow \{G,s\}. \tag{4}$$

Specifically, if we perform the Legendre transform a second time, we recover the original function. (If the restriction of convexity is relaxed, this statement must be revised, as remarked in Sec. VII.) In other words, suppose we start with the function $G(s)$ and calculate its Legendre transform. As we will see, $G(s)$ satisfies our conditions: convex and smooth. So, we start with

$$y(s) = \frac{dG}{ds}, \tag{5}$$

and invert the monotonic function $y(s)$ to $s(y)$. Next, we construct

$$H(y) = ys(y) - G(s(y)), \tag{6}$$

which can be rewritten as

$$G = sy - H. \tag{7}$$

If we compare Eqs. (7) and (2), we see that we can identify $\{H,y\}$ with $\{F,x\}$. Thus, the Legendre transform of G is the original function F, leading to the statement: the Legendre transform is its own inverse. This duality of the Legendre transform, shown symbolically in Eq. (4), is best displayed by the symmetric form

$$G(s) + F(x) = sx. \tag{8}$$

This equation should be read carefully. Despite its appearance, there is only one independent variable: either s or x. Referred to as a conjugate pair, these two variables are related to each other, through either $x(s)=dG(s)/ds$ or $s(x)=dF(x)/dx$. A careful writing of Eq. (8) would read either $G(s)+F(x(s))=sx(s)$ or $G(s(x))+F(x)=s(x)x$. To check the consistency with Eqs. (1) and (3), we can start with, say, the first of these equations and differentiate with respect to s. By applying the chain rule for $dF/ds=(dF/dx)(dx/ds)$, we recover $dG/ds=x$.

C. Properties of the extrema

The example in Fig. 3 shows a convex function $F(x)$ with a unique minimum. Let us denote this point by $F_{\min}=F(x_{\min})$. The slope of the tangent vanishes here, that is, $s(x_{\min})=0$. If we substitute this point into Eq. (2), we find that the minimum value of F is

$$F_{\min} = -G(0). \tag{9}$$

It is straightforward to show that a dual relation exists, namely, the minimum value of G is $G_{\min}=-F(0)$. To appreciate the geometric meaning of this relation, we need only to inspect Fig. 3 and see that $-G$ the y-intercept of the tangent to the curve $F(x)$ never reaches beyond $F(0)$.

By exploiting Eq. (8), both this example and the case of general extrema can be cast in an easy-to-remember symmetric form. Suppose F takes on its extremal value at x_{ext}. Then we have a horizontal tangent line and by definition, $s(x_{\text{ext}})=0$. Similarly, if G is at its extremum at s_{ext}, we have $x(s_{\text{ext}})=0$ due to Eq. (3). In either case, the right side of Eq. (8) vanishes and we have

$$G(0) + F(x_{\text{ext}}) = 0 \quad \text{and} \quad G(s_{\text{ext}}) + F(0) = 0. \tag{10}$$

D. Symmetric representation of the higher derivatives

Because the Legendre transform is a dual relation, we can expect manifestly symmetric relations beyond the ones we have seen so far:

$$G(s) + F(x) = sx \tag{11}$$

and

$$\frac{dG}{ds} = x \quad \text{and} \quad \frac{dF}{dx} = s. \tag{12}$$

From Eq. (12) we can obtain an infinite set of relations (if F and G are infinitely differentiable) linking G and F, by taking derivatives of $G+F=sx$ with respect to s or x. Because each function depends on only one variable, the differentials can be easily identified. Thus, differentiating the relations in Eq. (12) with respect to s or x as appropriate, we find

$$\frac{d^2G}{ds^2} = \frac{dx}{ds} \quad \text{and} \quad \frac{d^2F}{dx^2} = \frac{ds}{dx}. \tag{13}$$

Because $(dx/ds)(ds/dx)=1$, we have

$$\left(\frac{d^2G}{ds^2}\right)\left(\frac{d^2F}{dx^2}\right)=1. \tag{14}$$

We again emphasize that the variable s in the first factor and the x in the second are not independent but linked through Eq. (12).

Equation (14) illustrates the importance of (strict) convexity so that neither derivative ever vanishes. An interesting result is that the local curvatures of the Legendre transforms are inverses of each other in a manner reminiscent of the uncertainty relation $\Delta x\Delta k\approx 1$. For simplicity, suppose F is dimensionless but x is not,[6] so that s has the dimension of $1/x$. With this convention, it is easy to check the units of Eqs. (11), (12), and (14). If we differentiate Eq. (14) again, we can write a symmetric relation for the third derivative:

$$\frac{d^3G}{ds^3}\left[\frac{d^2G}{ds^2}\right]^{-3/2}+\frac{d^3F}{dx^3}\left[\frac{d^2F}{dx^2}\right]^{-3/2}=0. \tag{15}$$

Notice that each term is dimensionless, because the units of the various derivatives cancel.

It is possible to derive an infinite set of such relations for higher derivatives by differentiating further. Such an exercise also shows that if F is smooth (with a well defined nth order derivative), then so is G. The relations for higher derivatives do not have forms as simple as Eqs. (11), (12), (14), and (15) but become more and more complex.

IV. EXAMPLES OF THE LEGENDRE TRANSFORM IN SINGLE-PARTICLE MECHANICS

It is useful to provide some physical examples to illustrate these relations. The simplest is a quadratic function $F(x)=\alpha x^2/2$. For this function we easily find that $s=\alpha x$ and $x=s/\alpha$, leading to $G(s)=s^2/2\alpha$. The curvatures in F and G (α and $1/\alpha$, respectively) are inverses of each other as required by Eq. (14). All derivative relations beyond this level are trivial, that is, $0=0$.

This example corresponds to a single nonrelativistic particle with mass m moving in an external potential $V(q)$. The Legendre transform connects the Lagrangian $\mathcal{L}(\dot{q})$ to the Hamiltonian $\mathcal{H}(p)$. Only the kinetic term, which depends on $\dot{q}$ or p, is affected by the transform, because the potential depends on a different variable: q. In this case, $x\to\dot{q}$, $F\to\mathcal{L}$, $\alpha\to m$, $s\to p$, and $G\to\mathcal{H}$, so that $\mathcal{L}=m\dot{q}^2/2\Leftrightarrow\mathcal{H}=p^2/2m$. Because $V(q)$ is just a "spectator" in the Legendre transform, it must appear with opposite signs in F and G (that is, $\mathcal{L}$ and $\mathcal{H}$) to satisfy $F+G=sx$ (that is, $\mathcal{L}+\mathcal{H}=p\dot{q}$, with no q anywhere). Thus, we see the origin of the mysterious sign change in V when we go from the Hamiltonian to the Lagrangian.

Relativistic kinetic energy is a more interesting case. Here, let us go the other way, start with $\{\mathcal{H},p\}$ and generate the velocity as its slope. The relativistic kinetic energy as a function of momentum is $\mathcal{H}(p)=\sqrt{p^2+m^2}$ (with $c=1$), clearly a convex function. Its slope at the point p is

$$v\equiv\frac{d\mathcal{H}}{dp}=\frac{p}{\sqrt{p^2+m^2}}, \tag{16}$$

giving the familiar result

$$p=mv/\sqrt{1-v^2}. \tag{17}$$

Creating the Legendre transform using this pair of variables leads to the Lagrangian[7]

$$\mathcal{L}(v)=pv-\mathcal{H}(p(v))=-m\sqrt{1-v^2}. \tag{18}$$

This example can also be written in terms of the function $F(x)=\cosh\lambda x$. The demonstration is left to the reader (see Ref. 8).

Let us turn to a less familiar example that is so trivial it does not appear in typical textbooks. Yet it sets the stage for examining the role of the Legendre transform in equilibrium statistical mechanics. Consider a particle in a one-dimensional convex potential well, $U(x)$, which has a unique minimum at $x_{\min}$. An example would be a particle attached to a wall by a nonideal spring, with x being the distance from the point where the coils of the spring are fully compressed. The potential is effectively infinite at $x=0$, decreases to a minimum at its natural extension, and then increases for larger x. (We restrict our attention to positive values of x but less than the breaking point of the spring.) Another example of U is the potential that binds two atoms into a molecule (though such U's are rarely convex for all separations).

The particle is stationary only if it is at $x_{\min}$ for all time. If it is subjected to an additional external applied force f, then it will reach a new stationary point x_0, which is the solution to the equation

$$\left.\frac{dU}{dx}\right|_{x_0}=f. \tag{19}$$

To emphasize the dependence of this point on f, we write $x_0(f)$. We can ask the inverse question: If we want the particle to settle at $x_1\neq x_{\min}$, what force do we need to apply? The answer is $f(x_1)$, a force that depends on which x_1 we choose. A little thought leads us to the explicit functional form $f(x_1)=dU/dx|_{x_1}$. There is nothing special about the subscripts here, and we may write

$$f(x)=\frac{dU}{dx}, \tag{20}$$

and $x(f)$ instead of $x_0(f)$.

Although Eq. (20) gives $f(x)$ explicitly, we may ask if there is a counterpart to U which provides the inverse, $x(f)$, explicitly. If so, we can substitute f into the expression and arrive at the new equilibrium position. The answer is the Legendre transform of U, namely, $V(f)=fx-U(x(f))$. We leave it to the reader to show that

$$x(f)=\frac{dV}{df} \tag{21}$$

is the companion to Eq. (20).

All the details can be worked out for the simple example of the mass on a spring with $U(x)=kx^2/2$. This example is the analog of the nonrelativistic kinetic energy Legendre transform. The reader can easily demonstrate that the Legendre transform equation $U+V=fx$ becomes $(f-kx)^2=0$, yielding the relation between f and the new equilibrium point x.

Note that the information about the system (for example, the wall-spring-particle complex) is fully contained in either U or V. The only difference is in the coding: $\{U,x\}$ or $\{V,f\}$.

Although U is the usual potential energy associated with the particle at x, V is a kind of potential associated with the control f. In ordinary classical mechanics, such an approach seems unnecessarily cumbersome for describing the simple problem we posed. For this reason it is rightfully ignored in a course on classical mechanics. We include the example here only as a stepping stone to the Legendre transform in statistical mechanics and thermodynamics. There, multiple potentials are essential.

V. THE LEGENDRE TRANSFORM IN STATISTICAL THERMODYNAMICS

The Legendre transform appears frequently in statistical thermodynamics when different variables are "traded" for their conjugates.[2] Often, one of the variables is easy to think about, and the other is easy to control in physical situations.

The difficulty with making sense of the Legendre transform in thermodynamics arises from two causes: (1) For historical reasons, Legendre transform variables are not always chosen as conjugate pairs. (2) Many variables in thermodynamics are not independent and are constrained by equations of state, for example, $PV=Nk_BT$.

As an example of the first point, the conjugate to the total energy E of a system is the inverse temperature $\beta=1/k_BT$. Yet, our daily experience with the temperature T is so pervasive that T is used in most of the relations. Thus, the familiar equation

$$F = E - TS, \tag{22}$$

which relates the Helmholtz free energy F to the entropy S, obscures the symmetry between β and E, as well as the dimensionless nature of the Legendre transform. If we define the dimensionless quantities

$$\mathcal{S} \equiv S/k_B \text{ and } \mathcal{F} \equiv \beta F, \tag{23}$$

the duality between them can be beautifully expressed as

$$\mathcal{F}(\beta) + \mathcal{S}(E) = \beta E. \tag{24}$$

To elaborate the second point, we typically encounter a bewildering array of thermodynamic functions (for example, entropy, Gibbs and Helmholtz free energies, and enthalpy), a slew of variables (energy, temperature, volume, and pressure), as well as a jumble of thermodynamic relations (with multiple partial derivatives). Because of the multiple constrained variables, none of these examples is as simple as those we have considered, compounding the difficulty of both teaching and learning this material.

Before discussing the generation of the standard thermodynamic potentials, we briefly summarize the basics of statistical mechanics. We will show how the Legendre transform enters thermodynamics through the Laplace transform of partition functions in statistical mechanics.

Equilibrium statistical mechanics is based on the hypothesis[2] that for an isolated system, every allowed microstate is equally probable. The high probability of finding a particular equilibrium macrostate is due to a predominance of the number of microstates corresponding to that macrostate. The classic example is a gas of N identical, free, nonrelativistic structureless particles, confined in a D-dimensional box of volume L^D. For this system, a microstate is specified by the $2DN$ variables corresponding to the positions and momenta of each particle, $\{\vec{r}_i,\vec{p}_i\}$, with $i=1,\ldots,N$. Because the total energy E is a constant for an isolated system, the fundamental hypothesis can be represented as

$$P(\{\vec{r}_i,\vec{p}_i\}) \propto \delta(E - \mathcal{H}(\{\vec{r}_i,\vec{p}_i\})), \tag{25}$$

where $P(\{\vec{r}_i,\vec{p}_i\})$ is the probability of finding the configuration of positions and momenta $\{\vec{r}_i,\vec{p}_i\}$. In this case, the Hamiltonian $\mathcal{H}$ is explicitly given by

$$\mathcal{H} = \sum_i h(\vec{r}_i,\vec{p}_i) = \sum_i \left[\frac{\vec{p}_i^2}{2m} + U(\vec{r}_i) \right], \tag{26}$$

where m is the mass of each particle and U is the confining potential, which is zero for each component of $\vec{r} \in [0,L]$ and infinite otherwise.

The normalization factor for P is

$$\Omega(E) = \int_{r,p} \delta(E - \mathcal{H}(\{\vec{r}_i,\vec{p}_i\})), \tag{27}$$

where the integral is over all $\{\vec{r}_i,\vec{p}_i\}$ from $-\infty$ to ∞. (The infinite values of U restrict the actual position integrations to the volume of the box.) We have also suppressed the other variables that Ω depends on for now: L and m. Note that Ω is just the volume of phase space available to the system and is the microcanonical partition function.

The standard approach evaluates the integral in Eq. (27) as follows. The position integrals can be done explicitly because the only dependence of the Hamiltonian on position is the confinement of the position integrals to the allowed volume. These integrals yield a factor of L^{ND}. The momentum integrals are done by computing the surface area of a sphere in DN dimensions.

The entropy is introduced by the definition $S \equiv k_B \ln \Omega$. We exploit the dimensionless entropy $\mathcal{S}$ and write

$$\mathcal{S}(E) \equiv \ln \Omega(E). \tag{28}$$

To proceed, we have two choices: the route that emphasizes the mathematics or the physics.

A. The route of mathematics

Our task is straightforward: evaluate integrals with a constraint such as Eq. (27). Often, such integrals are not easy to perform. However, exploiting the Laplace transform typically renders the integrand factorizable. For example, the DN integrations in Eq. (27) become products of a single integral. Specifically, we consider the Laplace transform of $\Omega(E)$,

$$Z(\beta) \equiv \int \Omega(E) e^{-\beta E} dE. \tag{29}$$

If we substitute Eq. (27) for $\Omega(E)$, the delta function permits us to do the E integral giving

$$Z(\beta) = \int_{r,p} e^{-\beta \mathcal{H}}. \tag{30}$$

Because $\mathcal{H}$ is a sum over the individual components, the integrand factorizes, and we have

$$\int_{r,p} e^{-\beta\mathcal{H}} = \int_{r,p} \prod_i e^{-\beta h(\vec{r}_i,\vec{p}_i)} = \left[\int d\vec{r}d\vec{p}e^{-\beta h(\vec{r},\vec{p})}\right]^N. \quad (31)$$

The expression in [...] is much easier to handle and is $L^D(2\pi m/\beta)^{D/2}$. An attentive reader will have noticed from Eq. (30) that Z is the canonical partition function and can appreciate the statement: The two partition functions are related to each other by a Laplace transform.

To return to our goal, $\Omega(E)$, we need to perform an inverse Laplace transform, that is,

$$\Omega(E) = \int_{\mathcal{C}} Z(\beta)e^{\beta E}d\beta, \quad (32)$$

where $\mathcal{C}$ is a contour in the complex β plane (running parallel to and to the right of the imaginary axis). We define

$$\mathcal{F}(\beta) \equiv -\ln Z(\beta), \quad (33)$$

and write the integral as

$$e^{\mathcal{S}(E)} = \int_{\mathcal{C}} e^{-\mathcal{F}(\beta)+\beta E}d\beta. \quad (34)$$

To continue, it is necessary to inject some physics. In this case, we expect to be considering many particles, that is, large N. From Eq. (31) we have $\mathcal{F}\propto N$, leading us to expect that the range of E of interest is also $O(N)$. The standard tool for evaluating integrals with large exponentials as integrands is the saddle point (or steepest decent) method. Thus, we seek the saddle point in β, which is defined by setting the first derivative of $\beta E-\mathcal{F}(\beta)$ to zero:

$$\left.\frac{d[\beta E-\mathcal{F}]}{d\beta}\right|_{\beta_0} = 0. \quad (35)$$

In other words, we have

$$\left.\frac{d\mathcal{F}}{d\beta}\right|_{\beta_0} = E. \quad (36)$$

We emphasize that β_0 should be regarded as a function of E here.

In this approach, the integral in Eq. (29) is well approximated by evaluating the integrand at the saddle point, so that

$$\Omega(E) \cong \exp[\beta_0 E-\mathcal{F}(\beta_0)], \quad (37)$$

or using Eq. (28)

$$\mathcal{S}(E)+\mathcal{F}(\beta_0) = \beta_0 E, \quad (38)$$

with the understanding that β_0 and E are related through Eq. (36). There is nothing significant about the subscript on β, and Eq. (38) is identical to Eq. (24). In other words, $\mathcal{S}$ and $\mathcal{F}$ are Legendre transforms of each other. Thus, we see that (for situations involving a large parameter, N in this case) the Laplace and Legendre transforms, Eqs. (29) and (38), respectively, are related to each other as a result of the thermodynamic limit.

B. The route of physics: Interpretation of the equilibrium condition

Under what conditions does the internal energy move from one object to another and under what conditions can it be converted to work? Part of the answer lies in understanding which way the energy will move if we bring two systems into thermal contact. Why does it not go always from the system with more energy to the one with less? Considering this question leads us to the Legendre transform.

When two systems (not necessarily of the same size or energy) are brought together and the combined system isolated, $E_{\text{tot}}\equiv E_1+E_2$ will remain a constant and can be regarded as the control parameter. The individual E_j's are not fixed, and we ask: Starting at some initial values, how do they end up at the final equilibrium partition $\{E_1^*,E_2^*\}$? The answer lies with $\mathcal{S}_{\text{tot}}(E_{\text{tot}}|E_1,E_2)$, the entropy of the combined system subject to the specific partition of E_{tot} into $\{E_1,E_2\}$. The idea is that $e^{\mathcal{S}_{\text{tot}}}$ counts the number of allowed microstates associated with a particular partition and carries information of how probable that partition is. Calculating this quantity is usually not easy. However, if we focus on systems with extensive entropies, then we may write to a good approximation: $\mathcal{S}_{\text{tot}}=\mathcal{S}_1+\mathcal{S}_2$ with $\mathcal{S}_1=\mathcal{S}_1(E_1)$ and $\mathcal{S}_2=\mathcal{S}_2(E_2)$. These statements are not trivial: We are injecting the physics that for the specified conditions, the entropies of each system do not depend on the energy of the other.

Given these assumptions, we ask: For what partition will $\mathcal{S}_{\text{tot}}$ be a maximum, or equivalently, which partition is the most probable? If we write $E_2=E_{\text{tot}}-E_1$ and recall that E_{tot} is fixed, this task is easy. The maximum occurs at E_1^*, which is defined by

$$\left.\frac{d\mathcal{S}_{\text{tot}}}{dE_1}\right|_{E_1^*} = 0. \quad (39)$$

Because $dE_1=-dE_2$, we have

$$\left.\frac{d\mathcal{S}_1}{dE_1}\right|_{E_1^*} = \left.\frac{d\mathcal{S}_2}{dE_2}\right|_{E_2^*}, \quad (40)$$

where $E_2^*=E_{\text{tot}}-E_1^*$. This result is significant, because each side does not depend on the parameters of the other system. Thus, if we associate a quantity with $d\mathcal{S}/dE$, which we define by

$$\beta(E) \equiv \frac{d\mathcal{S}}{dE}, \quad (41)$$

then Eq. (40) becomes

$$\beta_1(E_1^*) = \beta_2(E_2^*). \quad (42)$$

In other words, the most probable partition occurs when the β of one system equals the β of the other. This condition does not depend on the details of the two systems, such as composition, size, or state (gas, liquid, or solid). When the two systems are brought into contact, energy will be transferred between them until they settle at values given by this condition: the equality of $\beta\equiv d\mathcal{S}/dE$, associated with each of them separately. It is natural, therefore, to use this quantity for describing our daily experience; namely, two systems, one hot and one cold, will equilibrate at a common temperature T when brought in contact with each other. Historically, many arbitrary scales were used for T. Their relation to the more natural quantity β was clarified later.

Besides providing a natural scale to describe "hot" and "cold," can the variable β be exploited further? For a given system we can write $\mathcal{S}(E(\beta))$, but is that useful? The answer is connected to the canonical ensemble, the (Helmholtz) free

energy, and the Legendre transform of $\mathcal{S}$. There is no need to reproduce here the standard derivation of this ensemble and the Boltzmann factor $e^{-\beta\mathcal{H}}$. In the previous section, we discussed the transformation between the partition functions $Z(\beta)$ and $\Omega(E)$ and the relation to the Legendre transform between $\mathcal{S}(E)$ and $\mathcal{F}(\beta)$.

C. How does the Legendre transform enter into thermodynamics?

For convenience, we summarize the key relations using dimensionless potentials:

$$\Omega(E) = e^{\mathcal{S}(E)} \quad Z(\beta) = e^{-\mathcal{F}(\beta)}, \tag{43}$$

$$\frac{d\mathcal{S}}{dE} = \beta, \quad \frac{d\mathcal{F}}{d\beta} = E, \tag{44}$$

where $Z(\beta) = \int dE e^{-\beta E}\Omega(E)$ and $\mathcal{F}(\beta) + \mathcal{S}(E) = \beta E$ in the thermodynamic limit. We can now see where the Legendre transform enters and why it is useful. The entropy $\mathcal{S}$ is a function of E, but the internal energy is typically not easy to control. To put more (less) energy into a system, we may heat (cool) it. In other words, we often manipulate E by coupling the system to an appropriate thermal bath, so the temperature or β becomes the control variable. In that case, we can perform a Legendre transform of $\mathcal{S}(E)$ and work with $\mathcal{F}(\beta)$ instead. Because $\{\mathcal{S}, E\}$ and $\{\mathcal{F}, \beta\}$ contain the same information about the system, it makes sense to deal with the more convenient thermodynamic potential when we change the control on a system from one variable to another.

Because the independent variable in a thermodynamic potential is to be regarded as a control (or a constraint) parameter, the slope associated with this function (for example, $d\mathcal{S}/dE$ and $d\mathcal{F}/d\beta$) carries physically significant information, namely, the response of the system to this control. The Legendre transform exchanges the role of the variables associated with control and response. In the example we just discussed, temperature (or β) is taken as the (very familiar) control variable, and the internal energy is regarded as the response. Thus, the free energy $\mathcal{F}(\beta)$ is the more appropriate potential, with $E = d\mathcal{F}/d\beta$ being the response. In the transformed version, which is mathematically and conceptually easier to grasp, E is a constraint (conserved variable for an isolated system) and $\mathcal{S}(E)$ is the more appropriate potential. After we understand the significance of its slope, $d\mathcal{S}/dE$, we can identify the response β as a measure for temperature. There are many other examples of response/control pairs to which the same kind of transformation may be applied, such as particle number and chemical potential, polarizability and electric field, and magnetization and magnetic field.

VI. LEGENDRE TRANSFORM WITH MANY VARIABLES

The thermodynamic potentials depend on many variables other than the total energy E. Each variable that can be independently controlled elicits a distinct response. As we construct Legendre transforms for each of these control/response variable pairs, we generate a new thermodynamic potential. The result is a plethora of thermodynamic functions. We emphasize that all these thermodynamic potentials carry the same information but encoded in different ways. We begin this section by discussing briefly the mathematical structure of the multivariable Legendre transform and then apply it to thermodynamics and statistical mechanics.

Consider the multivariate function $F(\vec{x})$, where $\vec{x}$ stands for M independent variables: $x_1, \ldots, x_M$. For convenience, suppose F is smooth and convex over all of this M-dimensional space. At every point $\vec{x}$ there will be M slopes:

$$s_m = \frac{\partial F}{\partial x_m} \equiv \partial_m F, \tag{45}$$

and $M(M+1)/2$ second derivatives, $\partial_m \partial_\ell F$, which can be regarded as a symmetric matrix. The convexity restriction requires that all of the eigenvalues of this matrix are positive (or negative).[9] In the context of thermodynamics, convexity is the condition for stability in equilibrium systems.[10] A standard corollary is that the relation between $\{x_m\}$ and $\{s_m\}$ is one to one, so that we can replace any one of the x_m's by the corresponding s_m through a Legendre transform.

Because we can transform any number of the x's, we may consider (up to) 2^M functions. For example, if we restrict ourselves to (E, V)—the standard variables for the microcanonical ensemble of the ideal gas—there are four thermodynamic functions: entropy, enthalpy, Gibbs, and the Helmholtz free energy. One way to picture the relation between so many functions is to put them at the corners of an M-dimensional hypercube. Each axis in this space is associated with a particular variable pair (x_m, s_m). Going from one corner to an adjacent corner along a particular edge corresponds to carrying out the Legendre transform for that pair. For the $M=2$ example of $(x_1, x_2) = (E, V)$, the hypercube reduces to a square, which is related, but not identical, to the square that appears in some texts.[2,11] Thanks to the commutativity of partial derivatives, going from any corner to any other corner is a path independent process, so that the function associated with each vertex is unique. For example, if we exchange (x_ℓ, x_m) for (s_ℓ, s_m), the Legendre transform relations would be the simple generalization of Eq. (11):[12]

$$F(x_1, \ldots x_\ell, \ldots x_m, \ldots x_M) + G(x_1, \ldots s_\ell, \ldots s_m, \ldots x_M) = s_\ell x_\ell + s_m x_m, \tag{46}$$

with $\partial_\ell G = x_\ell$, $\partial_m G = x_m$, $\partial_\ell F = s_\ell$, and $\partial_m F = s_m$. We should have given this G some special notation to denote that its variables are all $\{x\}$ except for the two that are $\{s\}$. A possibility is $G^{\ell,m}$, but for simplicity we do not pursue this issue further. One special Legendre transform is noteworthy—the one in which all variables are $\{s\}$. Located at the corner of the hypercube diametrically opposite to F, this function will be denoted by H. In this case, the Legendre transform relation simplifies to

$$H(\vec{s}) + F(\vec{x}) = \vec{s} \cdot \vec{x}. \tag{47}$$

Generalizations for higher derivatives are straightforward. For example, Eq. (14) becomes

$$\sum_m (\partial_\ell \partial_m H)(\partial_m \partial_n F) = \delta_{\ell n}, \tag{48}$$

where δ is the unit matrix. The convexity of F guarantees that the inverse of $\partial_m \partial_n F$ exists.

Let us now apply these considerations to the thermodynamics of a gas. We begin with the microcanonical partition function $\Omega(E,V)$ and consider the mapping

$$F(x_1,x_2) \to \mathcal{S}(E,V) \equiv \ln \Omega, \tag{49}$$

where $x_1 \to E, x_2 \to V, s_1 \to \beta$, and $s_2 \to \eta$. The last of these is related to the pressure P. The Legendre transform with respect to x_1 leads to the Helmholtz free energy. Our symmetric and dimensionless version of $F=E-TS$ is the same as Eq. (24): $\mathcal{F}(\beta,V)+\mathcal{S}(E,V)=\beta E$, with V playing the role of a "spectator." To be precise, we now write Eq. (41) with the partial derivative:

$$\beta \equiv \left.\frac{\partial \mathcal{S}}{\partial E}\right|_V. \tag{50}$$

For the second Legendre transform with respect to $x_2=V$, we define[13]

$$\eta \equiv \left.\frac{\partial \mathcal{S}}{\partial V}\right|_E \tag{51}$$

and arrive at

$$\mathcal{G}(\beta,\eta)+\mathcal{S}(E,V)=\beta E+\eta V, \tag{52}$$

where $\mathcal{G} \equiv \beta G(T,P)$ is the dimensionless Gibbs free energy. The relation between η and the traditional definition of pressure, $P \equiv -\partial E/\partial V|_S$, is $\eta=\beta P$. To derive this relation would take us further into the first law of thermodynamics and the notion of heat transfer. The interested reader should consult a standard text such as Ref. 13.

We return to Eq. (52), move $\mathcal{S}$, and divide both sides by β to arrive at its more common form: $G=E-TS+PV$. The seemingly mysterious signs of the last two terms on the right are, from our perspective, due to the placing of S and the use of T instead of β. In contrast, every term comes with a positive sign in Eq. (52), with all the potentials on the left and all the conjugate variables on the right. Note that there are just two variables in this example, so that $\mathcal{G}$ plays the role of H in Eq. (47), which is an explicit version of Eq. (52).

We next consider the enthalpy, which is laden with extra complications. For various reasons, S (instead of E) is chosen to be the independent variable for arriving at the enthalpy. As a result, instead of β, the natural conjugate variable is T $(=\partial E/\partial S)$. Regarding S as a control variable with which to access E is conceptually difficult. However, it is common to think of transferring heat so that TdS appears as the means of control. If we take the Legendre transform of $E(S)$ in the standard fashion, we would arrive at $TS-E$, which is the Helmholtz free energy except for a sign. The disadvantage is clear, but there are advantages to this approach. In particular, by starting with $E(S,V)$, we naturally arrive at the ordinary pressure, $-P$, as the conjugate to V (instead of η). Note the extra minus sign here. The Legendre transform with respect to V of $E(S,V)$ gives $(-P)V-E$, the (negative of) enthalpy $H=E+PV$. If we allow logic to overcome tradition, we would have defined the last potential as $\mathcal{H}(E,\eta)$ (not to be confused with the Hamiltonian $\mathcal{H}$) through the Legendre transform

$$\mathcal{H}(E,\eta)+\mathcal{S}(E,V)=\eta V, \tag{53}$$

in which the variable E plays the role of a spectator. But, the beauty of pure reason does not always prevail, and we must often abide by the results of our historical paths.

VII. CONCLUDING REMARKS

There are many interesting aspects of the Legendre transform we have not discussed. Covering all aspects would be more appropriate for a textbook. We conclude by touching on just two important generalizations—the Legendre transform of nonconvex functions and functions defined on spaces with nontrivial topology, such as the angle on a circle—and providing references for further reading.

If a function is nonconvex, the Legendre transform becomes multivalued. If we delete all but the principal branch, the Legendre transform develops discontinuous first derivatives. If we perform another transformation, the result would be the convex hull of the original. This topic is intimately related to the Maxwell construction and the coexistence of phases (for example, liquid and vapor). Although most texts on thermodynamics and statistical mechanics discuss the Maxwell construction, few demonstrate its relation to the Legendre transform of nonconvex functions. A good example of a convexified (free energy) function is given in Ref. 14.

A second generalization concerns variables whose domains have a nontrivial topology, the simplest being functions defined on a circle or the surface of a sphere. The angles are the most natural variables for a sphere, but we must be mindful of the periodic nature of $\phi \in (0,2\pi]$ and the co-ordinate singularities at the poles $\theta=0,\pi$. An example is the shape of crystals in equilibrium with its liquid (for example, ^{4}He crystals in coexistence with the superfluid[15]) or vapor (for example, gold crystals[16]). Typical crystal shapes are not spherical and can be described by a nontrivial function $R(\theta,\phi)$, which specifies the distance from the center of mass to a point on the crystal surface labeled by (θ,ϕ). The tangent plane at that point can be associated with the direction of its normal and labeled by $(\tilde{\theta},\tilde{\phi})$. The relation between these and the derivatives $\partial_\theta R$ and $\partial_\phi R$ exists but is not simple. From these derivatives a (generalized) Legendre transform of R can be constructed: $\sigma(\tilde{\theta},\tilde{\phi})$. The function σ is also a significant physical quantity: it is the free energy per unit area (the surface tension) associated with a planar interface, with normal $(\tilde{\theta},\tilde{\phi})$, between the crystalline and the isotropic phases of the material. A bonus is that, unlike typical thermodynamic potentials such as the entropy and free energies, the potential $R(\theta,\phi)$ is not just an abstract concept; it is the shape of a physical object in three dimensions. Further details of this intriguing connection are in Ref. 17.

Finally, we point readers to horizons far beyond those discussed here. Because our purpose is to reach students and instructors in upper undergraduate and core graduate courses, we have limited our considerations to cases with two (or finite M) variables. It is possible to study the Legendre transform with an infinite number of variables. Probably the best known example in physics comes from both quantum field theory[18] and statistical field theory.[19] Associated with each quantum field $\phi(\vec{r},t)$ is a "source field" $J(\vec{r},t)$, in much the same way that a fluctuating local magnetization, $m(\vec{r})$, can be "created" by an inhomogeneous magnetic field $B(\vec{r})$. In the latter system, the fluctuations of m are thermal, rather than quantum, in nature. The source field can be regarded as a control variable for each $\vec{r},t$ (or just $\vec{r}$). Thus, there are an

infinite number of variables, as well as responses, involved. Corresponding to a given $J(\vec{r},t)$ or $B(\vec{r})$, we can calculate, in principle, the "vacuum energy" $\mathcal{U}[J(\vec{r},t)]$ or the free energy $\mathcal{F}[B(\vec{r})]$. These quantities carry information on the quantities of interest: connected Schwinger functions (expectation values of products of ϕ's) or correlations functions (averages of products of m's). More useful than $\mathcal{U}$ is its Legendre transform, Γ, which is known as the effective action. Γ displays the essential information more conveniently in terms of one particle irreducible Schwinger functions or vertex functions. For the effective action of a quantum field, there is a particularly appealing systematic expansion in powers of $\hbar$. The zeroth order term is just the classical action. Similarly, for the Legendre transform of $\mathcal{F}$, there is a systematic expansion in powers of T or β^{-1}. Not surprisingly, the zeroth order term is just the energy associated with $m(\vec{r})$, which is the Hamiltonian $\mathcal{H}[m(\vec{r})]$ that enters the Boltzmann factor $\exp\{-\beta\mathcal{H}\}$. We hope that these comments will help some students who are struggling with field theory or further motivate those who are waiting to delve into the subject.

ACKNOWLEDGMENTS

We thank many colleagues for fruitful discussions as well as Beate Schmittmann for critical readings of the manuscript. This work is supported in part by the U.S. National Science Foundation through Grant Nos. DMR-0705152 and DUE-0524987.

[1] C.-C. Cheng, "Maxwell's equations in dynamics," Am. J. Phys. **34**, 622 (1966); A. L. Fetter and J. D. Walecka, *Theoretical Mechanics of Particles and Continua* (McGraw-Hill, New York, 1980).

[2] K. Huang, *Statistical Mechanics* (Wiley, New York, 1987); H. S. Robertson, *Statistical Thermophysics* (Prentice Hall, New York, 1997).

[3] J. W. Cannon, "Connecting thermodynamics to students' calculus," Am. J. Phys. **72**, 753–757 (2004).

[4] M. Artigue, J. Menigaux, and L. Viennot, "Some aspects of students' conceptions and difficulties about differentials," Eur. J. Physiol. **11**, 262–267 (1990); E. F. Redish, "Problem solving and the use of math in physics courses," to be published in *Proceedings of the Conference, World View on Physics Education in 2005: Focusing on Change*, Delhi, India, August 21–26, 2005; ⟨www.physics.umd.edu/perg/papers/redish/IniaMath.pdf⟩.

[5] In this example, s, x, F, and G are all positive. Thus, the "G axis" points downward, opposite to the "F axis."

[6] This restriction can be lifted, especially if physical quantities with dimensions (for example, the Hamiltonian) are studied. In that case, we must keep more careful track of the units, such as $[s]=[F][x^{-1}]$.

[7] See, for example, Eq. (12.7) in J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, New York, 1999) or Eq. (7.136) in H. Goldstein, *Classical Mechanics*, 2nd ed. (Addison-Wesley, Reading, MA, 1980).

[8] E. Taylor and J. A. Wheeler, *Spacetime Physics* (Freeman, New York, 1966).

[9] In general F may be regarded as a smooth M-dimensional manifold. The eigenvalues of $\partial_m\partial_\ell F(\vec{x})$ are the principal curvatures of this surface at $\vec{x}$.

[10] For systems in nonequilibrium stationary states, negative responses can be easily achieved. See, for example, R. K. P. Zia, E. L. Praestgaard, and O. G. Mouritsen, "Getting more from pushing less: Negative specific heat and conductivity in nonequilibrium steady states," Am. J. Phys. **70**, 384–392 (2002).

[11] This sort of construction is attributed to Born. See, for example, the discussion in W. W. Bowley, "Legendre transforms, Maxwell's relations, and the Born diagram in fluid dynamics," Am. J. Phys. **37**, 1066–1067 (1969).

[12] These partial derivatives are taken with the understanding that all other variables are held fixed. It is common (and reasonable) to consider derivatives with F or G held fixed. In this article we avoid discussing such complications.

[13] We follow the notation in R. K. Pathria, *Statistical Mechanics* (Pergamon, Oxford, 1972).

[14] D. V. Schroeder, *An Introduction to Thermal Physics* (Addison-Wesley, Reading, MA, 2000), Fig. 5.27.

[15] J. Landau, S. G. Lipson, L. M. Määttänen, L. S. Balfour, and D. O. Edwards, "Interface between superfluid and solid ^{4}He," Phys. Rev. Lett. **45**, 31–35 (1980).

[16] J. C. Heyraud and J. J. Métois, "Equilibrium shape of gold crystallites on a graphite cleavage surface: Surface energies and interfacial energy," Acta Metall. **28**, 1789–1797 (1980).

[17] See, for example, M. Wortis, "Equilibrium Crystal Shapes and Interfacial Phase Transitions," in *Chemistry and Physics of Solid Surfaces*, edited by R. Vanselow (Springer, New York, 1988), Vol. VII, pp. 367–406; and R. K. P. Zia, "Anisotropic Surface Tension and Equilibrium Crystal Shapes," in *Progress in Statistical Mechanics*, edited by C. K. Hu (World Scientific, River Edge, NJ, 1988), pp. 303–357. See also R. K. P. Zia and J. E. Avron, "Total surface energy and equilibrium shapes: Exact results for the $d=2$ Ising crystal," Phys. Rev. B **25**, 2042–2045 (1982). The connection between anisotropic surface energy and the minimizing shape was first established over a century ago by G. Wulff, "Zur Frage der Geschwindigkeit des Wachstums und der Auflösung der Krystallflächen," Z. Krystal. Mineral. **34**, 449–530 (1901).

[18] J. Schwinger, "The theory of quantized fields I," Phys. Rev. **82**, 914–927 (1951) and J. Schwinger, "The theory of quantized fields II," Phys. Rev. **91**, 713–728 (1953). For a more recent treatment, see, for example, S. Weinberg, *The Quantum Theory of Fields* (Cambridge U. P., Cambridge, MA, 1996).

[19] A recent text containing chapters on statistical fields is M. Kardar, *Statistical Physics of Fields* (Cambridge U. P., Cambridge, MA, 2007). More complete treatments may be found in C. Itzykson and J. M. Drouffe, *Statistical Field Theory* (Cambridge U. P., Cambridge, MA, 1989) and J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena* (Oxford U. P., New York, 2002).