Heat capacity in bits

P. Fraundorf[a)]
Department of Physics and Astronomy and Center for Molecular Electronics, University of Missouri, St. Louis, St. Louis, Missouri 63121 and Department of Physics, Washington University, St. Louis, Missouri 63130

(Received 15 July 2002; accepted 2 June 2003)

The temperature T may be expressed as the rate of energy increase per unit increase in the state uncertainty under no-work conditions. The consequences of such a choice for heat capacities are explored. I show that the ratio of the total thermal energy E to kT is the multiplicity exponent (log–log derivative of the multiplicity) with respect to energy, as well as the number of base-b units of mutual information that is lost about the state of the system per b-fold increase in the thermal energy. Similarly, the no-work heat capacity C_V is the multiplicity exponent for temperature, making C_V independent of the choice of the intensive parameter associated with energy (for example, kT vs $1/kT$) to within a constant, and explaining why its usefulness may go beyond the detection of thermodynamic phase changes and quadratic modes.
[DOI: 10.1119/1.1593658]

I. INTRODUCTION

The adjective ''deep'' has been applied to simplifying approaches that offer both wider applicability and reduced algorithmic complexity.[1] Minkowski's approach to special relativity via the metric equation is a classic example. Initial impressions of it as ''superfluous erudition''[2] were later eclipsed by its uses in the development of general relativity, in metric interpretation of other forces[3] (for example, string theory), and in simplifying introductions to space–time.[4–6]

Similarly, the insights of Shannon[7] led Jaynes and others[8,9] to clarify the distinction in thermal physics between ''describing the dice'' and ''taking the best guess.'' Conservation of energy describes the dice by specifying that energy lost from one system must appear in another. In contrast, the zeroth law, along with the Boltzmann and Gibbs factors for a given system, are ''dice-independent'' tools of statistical inference,[10] because they generally describe the role of the energy's Lagrange multiplier (reciprocal temperature) when two systems are allowed to equilibrate under various constraints, for example, at constant volume or constant pressure.

This insight for example allows us to infer equipartition and the ideal gas law (as well as their range of validity) from simple assumptions about the systems to which they apply.[12] The approach is becoming increasingly popular in undergraduate texts[11–18] as a way to increase student understanding. It also bolsters physical intuition about matters involving the codes of biology[19] and computer science.[20] The heart of the approach is found in Shannon's and Jayne's observations that the physical entropy S of classical and quantum statistical mechanics is a mathematical measure of statistical uncertainty about the state of a system, most naturally expressed in information units (like bits or bytes) rather than in historical units (for example, joules per kelvin). This interpretation has paved the way to a formal connection between statistical uncertainty and physical entropy, for example, by considering increases in the uncertainty as decreases in the mutual information[21] between a system and its environment. It then follows that temperature (the reciprocal of the no-work uncertainty slope dS/dE) is quantifiable in units of the energy per unit increase in the state uncertainty, and that heat capacities, that is, dE/dT normally in joules per kelvin, may be assigned information units as well.

What is the physical meaning of a heat capacity without reference to traditional temperature units? Bits of what? An answer to this question does not appear to be common knowledge, so we outline an answer here. The insight for students into the mechanism that underlies the utility of heat capacity in calculations involving thermal equilibrium might better prepare them for applying the principles of statistical inference to more complex systems, that is, to those not necessarily at equilibrium.

II. DESCRIBING THE DICE

The dice of thermal physics are usually physical systems capable of accommodating thermal energy (as well as other quantities that may be conserved, like volume and particles) in a multiplicity of ways. This multiplicity is the key to understanding, particularly if systems are seen from the (microcanonical) vantage point of the conserved quantities. For example, many gases, liquids, and solids behave over part of their temperature range as though the multiplicity Ω is proportional to $E^{\nu N/2}$, where E is the thermal energy, N is the number of molecules, and ν is the number of ways of storing the thermal energy per molecule. Such systems are called quadratic, because the proportionality results from a sum of squares relation between E and the state coordinates involved in storing the energy.

III. TAKING THE BEST GUESS

The application of gambling theory to physical systems begins with the question: Where might we expect to find a conserved quantity X that has been randomly shared between systems for so long that prior information about the whereabouts of X is irrelevant? In the jargon of the field, this question is the same as asking: How is X likely to be shared after equilibration?

A. Why temperatures help

The science of decision making in the presence of uncertainty (that is, statistical inference or gambling theory) suggests that the best bet is the distribution of X that can happen in the most ways, provided there is no reason to prefer one way over another. In the jargon of probability theory, we assume equal *a priori* probabilities, when evidence to the contrary is not available.

For example, if systems A and B have a total energy E to share between them, then the best bet after equilibration will be that value of $E_A = E - E_B$ that has the largest total multiplicity $\Omega = \Omega_A \Omega_B$. If we set to zero the derivative of Ω with respect to E_A, we find that this maximum requires that $(1/\Omega_A)(d\Omega_A/dE_A) = (1/\Omega_B)(d\Omega_B/dE_B)$, and hence $d(\ln \Omega)/dE$ is the same for both systems. These derivatives (and most others in this paper) are taken under microcanonical constraints, that is, they are partial derivatives with other extensive quantities (like volume or number of particles) held constant.

Information theorists define the logarithm of multiplicity as the uncertainty via the relation $S = k \ln \Omega$, and measure it in units of nats, bits, bytes, or J/K if k is 1, $1/\ln 2$, $1/\ln 256$, or 1.38×10^{-23}, respectively. Thus the best bet for two systems sharing a conserved quantity X in the absence of information to the contrary is that X will rearrange itself between the two systems until each system's uncertainty slope dS/dX has reached a common value. This prediction of a property (corresponding to temperature or its reciprocal) that approaches a common value between equilibrated systems is a quantitative version of the zeroth law of thermodynamics, based purely on the science of statistical inference. It applies to any system randomly sharing conserved quantities.

When the energy E is the quantity that is shared, the uncertainty slope dS/dE is the reciprocal temperature or the coldness[15] $1/T$. Hence, the temperature is a property that is a measure of the propensity of a system to share energy thermally. For example, the uncertainty slope for quadratic systems calculated from the multiplicity given in Sec. II yields the equipartition relation: $E/N = \nu kT/2$.

When V is the quantity shared randomly between systems, the uncertainty slope dS/dV is the free-expansion coefficient (see, for example, the derivations in Secs. 3.22 and 5.7 of Ref. 15) equal to $(dS/dE \times \text{force} \times dx)/(\text{area} \times dx) = P/T$ at equilibrium. For an ideal gas, $\Omega \propto V^N$ and in this case dS/dV yields the ideal gas equation of state $PV = NkT$. If N is the quantity shared randomly, the uncertainty slope dS/dN is the chemical affinity, equal to $-\mu/T$ at equilibrium. From this consideration, for example, reaction equilibrium constants may be calculated.

Before leaving this review of simple applications of the predictive power of "equilibrating multipliers," a graphical way for plotting relations between extensive work parameters (like volume) and their corresponding multipliers (for example, the free-expansion coefficient) in the context of $S = k \ln \Omega$ and its derivatives is illustrated for an ideal gas in Fig. 1. Plots of this form will be used to examine the natural units for heat capacity in what follows. Here Boyle's law (plotted as a relation between V and its multiplier P/kT) emerges in the equation of state plot [Fig. 1(c)] on the lower left, while the fact that the number of particles N is the exponent of both V and P/kT, in the expression for multiplicity Ω, emerges in the exponent map [Fig. 1(b)] on the upper right. Note also from the log–log plot of multiplicity versus work parameter on the lower right [Fig. 1(d)] that the entropy is a smooth and monotone increasing function of V, simplifying entropy maximizations and ensuring that the multiplier (P/kT in this case) remains positive.

B. Range of application

The quantitative version of the zeroth law given previously applies to all equilibrium thermal systems, including spin systems capable of population inversions and hence negative absolute temperatures. Moreover, as a theorem of statistical inference not involving energy at all, it applies to thermally unequilibrated systems sharing other conserved quantities (including money, for example), provided that the only prior information is how the multiplicity of ways a quantity can be distributed depends on the amount of that conserved quantity. If we have other kinds of information, such as knowledge of a system's temperature but not its total energy, then the broader class of maximum entropy strategies in statistical inference (for example, Gibbs' canonical and grand ensembles) predict the distribution of outcomes.

The quantitative role in statistical inference for the uncertainty S and its derivatives also extends naturally to systems with internal correlations or mutual information.[21] Whether correlated subsystems are easily replicable (for example, xerox copies of a valid street map) or difficult to clone (for example, qubits[22]), the consideration of the knowledge that otherwise unknown subsystems are correlated allows one to extend the tools of statistical inference (and hence the second law of thermodynamics) to complex systems whose entropy is less than the sum of its subsystem entropies. A simple classic example of such a system consists of Szilard's bisymmetric single-atom vacuum-pump memory and an external observer.[25] This memory can be placed into either of its two possible states, decreasing the state uncertainty by 1 bit, by either an isothermal compression requiring $kT \ln 2$ of work, or an arbitrarily slow re-orientation guided by an observer who knows (has mutual information that specifies) on which side of a removable central partition the atom can be found.

In addition to illustrating the thermodynamic value of mutual information, the vacuum-pump memory example is also an example of an (observer plus test system) world, where information that an observer has about the state of a test system may be seen as mutual information in the world of which both are a part. From this perspective, a second law that puts constraints on the entropy of isolated systems may be applied rigorously not only to subsystems that share energy, but to the information that observers external to a system of interest have about its state. This relation between mutual information and observer knowledge is precisely why statistical inference (that is, the formal use of gambling theory) in thermal physics may be predicated on uncertainty about a system as seen by an external observer.

Pedagogical approaches that leverage this approach include thermodynamic-engine "possibility problems" which use only the first and second laws. Examples include the usual Carnot heat engine and refrigerator efficiency limits on reversible energy flow between two different thermal reservoirs, the at first glance nonintuitive "ice-water invention" problem of converting boiling water to ice water reversibly given no available work and only a room temperature reservoir as leverage, tantalizing results pointed out by Jaynes[23] on "reversible home heating with flames" and "zero energy ovens for eskimos," constraints on the channel capacity and

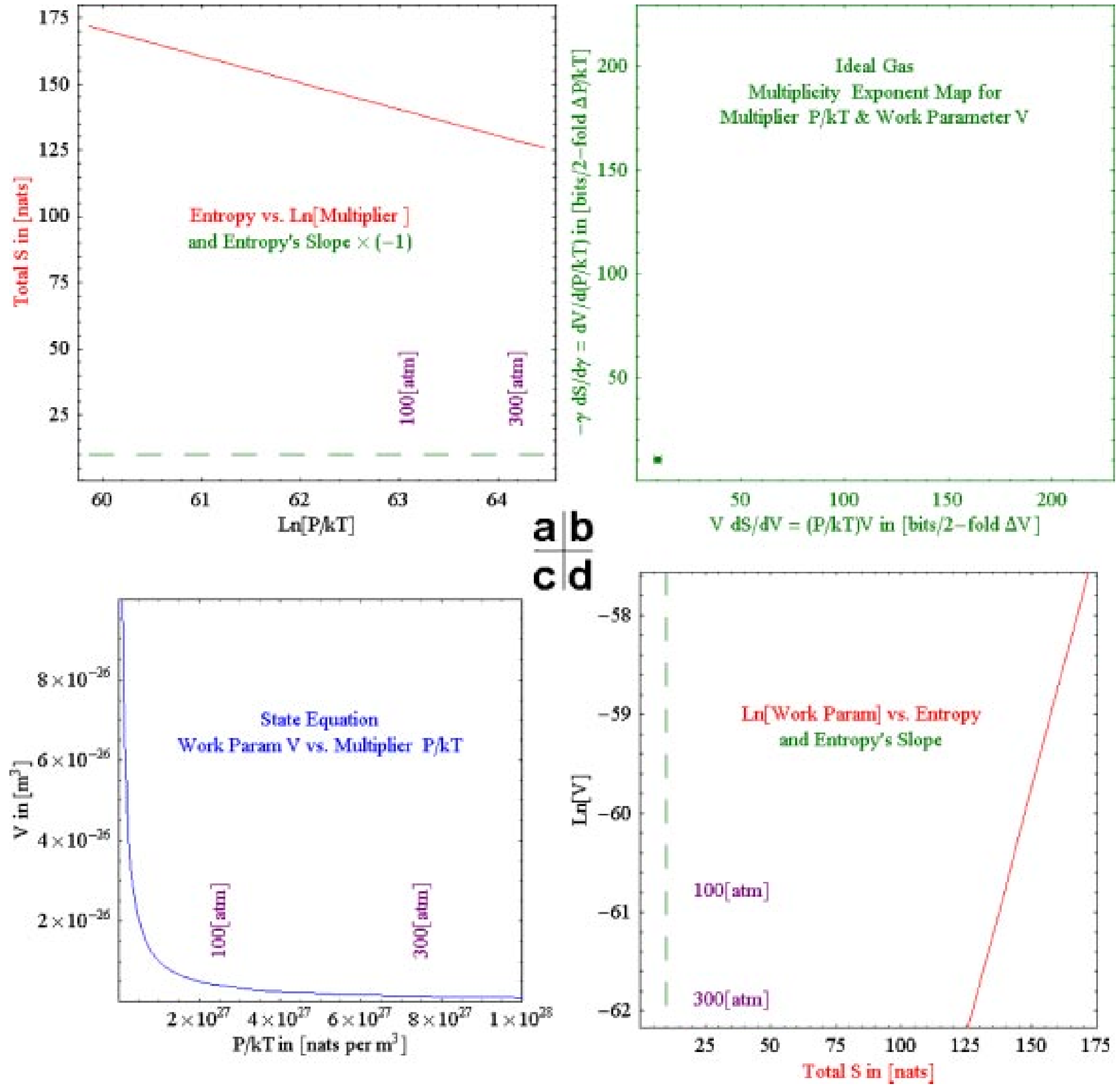

Fig. 1. Plots of the dependence of the entropy on the volume, and the latter's Lagrange multiplier $\gamma = P/kT$, for an ideal (Sakur–Tetrode) 10-atom argon gas at 290 K. (a) S (solid line) and its slope $\gamma\, dS/d\gamma$ (dashed line) vs $\ln \gamma$, (b) the multiplicity exponent map $\gamma\, dS/d\gamma$ vs $V\, dS/dV$, (c) state equation plot of V vs $dS/dV = \gamma = P/kT$, (d) $\ln V$ vs S and its slope $V\, dS/dV$ (dashed line). The entropy expression used for this is $S/k = N[\ln(V/N) + \frac{3}{2}\ln(4\pi mE/3Nh^2) + \frac{5}{2}]$, where V is the total volume, E is the total energy, N is the number of atoms, m is the atomic mass, and h is Planck's constant. Note that Boyle's law appears in the ideal gas equation of state plot in the lower left, that the adjacent plots show the entropy to be a monotonic increasing function of both V and P/kT, and that the exponent map on the upper right shows the log–log derivative of the multiplicity for both V and P/kT to be fixed at the number of particles $N = 10$.

energy dissipation of molecular machines,[24] and the $W_{\text{in}}/kT_{\text{out}}$ lower limit on work thermalized for each nat of computer memory erased in preparation for a subsequent write.[25] Such applications of the first and second laws constrain what is possible while involving no specifics about the actual implementation. For introductory classes, the sustainable flow of available work through biological systems (regardless of their distance from equilibrium), long the province of ecologists,[26] becomes fair game as a physics subject of broad interest. Finally, in courses with few mathematical prerequisites, we can avoid logarithms (and reinforce student insight into the meaning of computer science units) by using only powers of 2. For example, $N_{\text{choices}} = 2^{n_{\text{bits}}} = 2^{1.4S/k}$ quantifies the relation between multiplicity, measures of uncertainty in bits, and traditional units for entropy.

IV. THE RATIO OF THERMAL ENERGY TO *kT*

The statistical definition of temperature can be rewritten as:

$$\frac{1}{kT} \equiv \frac{\partial(\ln \Omega)}{\partial E} = \frac{\partial(\ln \Omega)}{\partial(\ln E)}\,\frac{1}{E}. \tag{1}$$

We can think of the quantity $\partial(\ln \Omega)/\partial(\ln E)$ as the instantaneous exponent of energy or as the slope of the multiplicity versus energy curve on a log–log plot. We can rearrange Eq. (1) so that it looks very much like the familiar equipartition theorem often given to introductory students without qualification, except that Eq. (1) applies to all thermal systems under conditions of maximum ignorance (that is, at equilib-

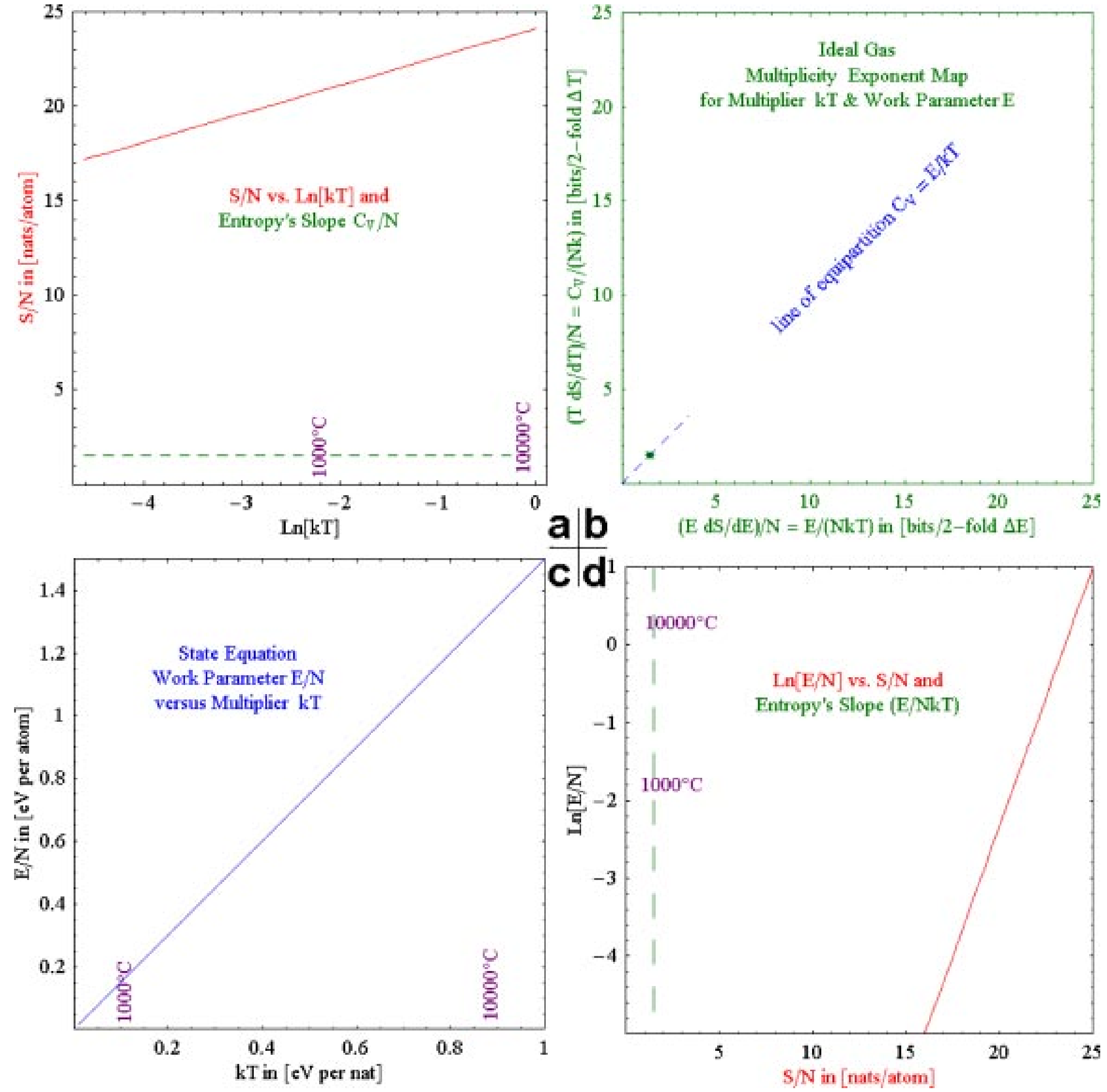

Fig. 2. Plots of the dependence of the entropy on the energy E and energy multiplier kT for the ideal (Sakur–Tetrode) 10-atom argon gas of Fig. 1 in a fixed volume of 0.4 zeptoliters. (a) S and its slope C_V/k (dashed line) vs ln(kT), (b) multiplicity exponent map C_V/k vs E/kT, (c) equation of state plot E/N vs $dE/dS = kT$, and (d) ln(E/N) vs S/N and its slope E/kT (dashed line). Note how the lower left corner panel axes lead into both adjacent panels, and how the derivative plots in turn lead to the exponent map in the upper right, which for the case of this simple quadratic system yields a single point at $C_V/N = E/NkT = 3/2$ on the "line of equipartition."

rium). The analog to degrees of freedom over 2 then becomes:

$$\xi \equiv \frac{E}{kT} = E\frac{\partial(S/k)}{\partial E} = E\frac{\partial \ln \Omega}{\partial E} \overset{E>0}{=} \frac{\partial \ln \Omega}{\partial \ln E} = \frac{\partial(\log_b \Omega)}{\partial(\log_b E)}, \tag{2}$$

where b is any positive real number. For quadratic systems, it is easy to see that the log–log derivative in Eq. (2) is equal to $\nu N/2$. For any system, ξ defined in Eq. (2) measures the instantaneous energy exponent, as well as the number of nats of information lost about the state of the system per e-fold increase in the thermal energy of the system. The last term in Eq. (2) notes that the value of ξ is independent of the base b of the logarithms used. Thus we can also think of ξ as the number of bits of information lost per twofold increase in the thermal energy, or more generally the number of base-b units of information lost per b-fold increase in thermal energy. In our search for the meaning of the heat capacity in natural units, this concept of multiplicity exponent (in this case for energy E) is our first big clue.

The relationships in Eq. (2) are illustrated in the bottom half of Fig. 2 for a simple quadratic system. On the left is plotted E/N per atom vs kT for a low density argon gas, and on the right is plotted ln(E/N) vs S/k using the quadratic system multiplicity with $\nu = 3$. The dashed line on the right is the derivative of the S/k vs ln(E/N) curve, or in other words the degrees of freedom over 2, that is, $\xi = 1.5$.

Before we move on, we also point out a relation that applies if the energy origin has been chosen so that $E \rightarrow 0$ as $T \rightarrow 0$, something we might expect for a measure of thermal energy. In terms of the heat capacity C_V, we can write:

$$\xi = \frac{E}{kT} = \frac{\int_0^T C_V dT}{kT} \overset{T>0}{=} \frac{1}{\Delta T}\int_0^T C_V dT k \Delta T \equiv \left\langle \frac{C_V}{k} \right\rangle, \tag{3}$$

where the middle equality applies only for systems not in a

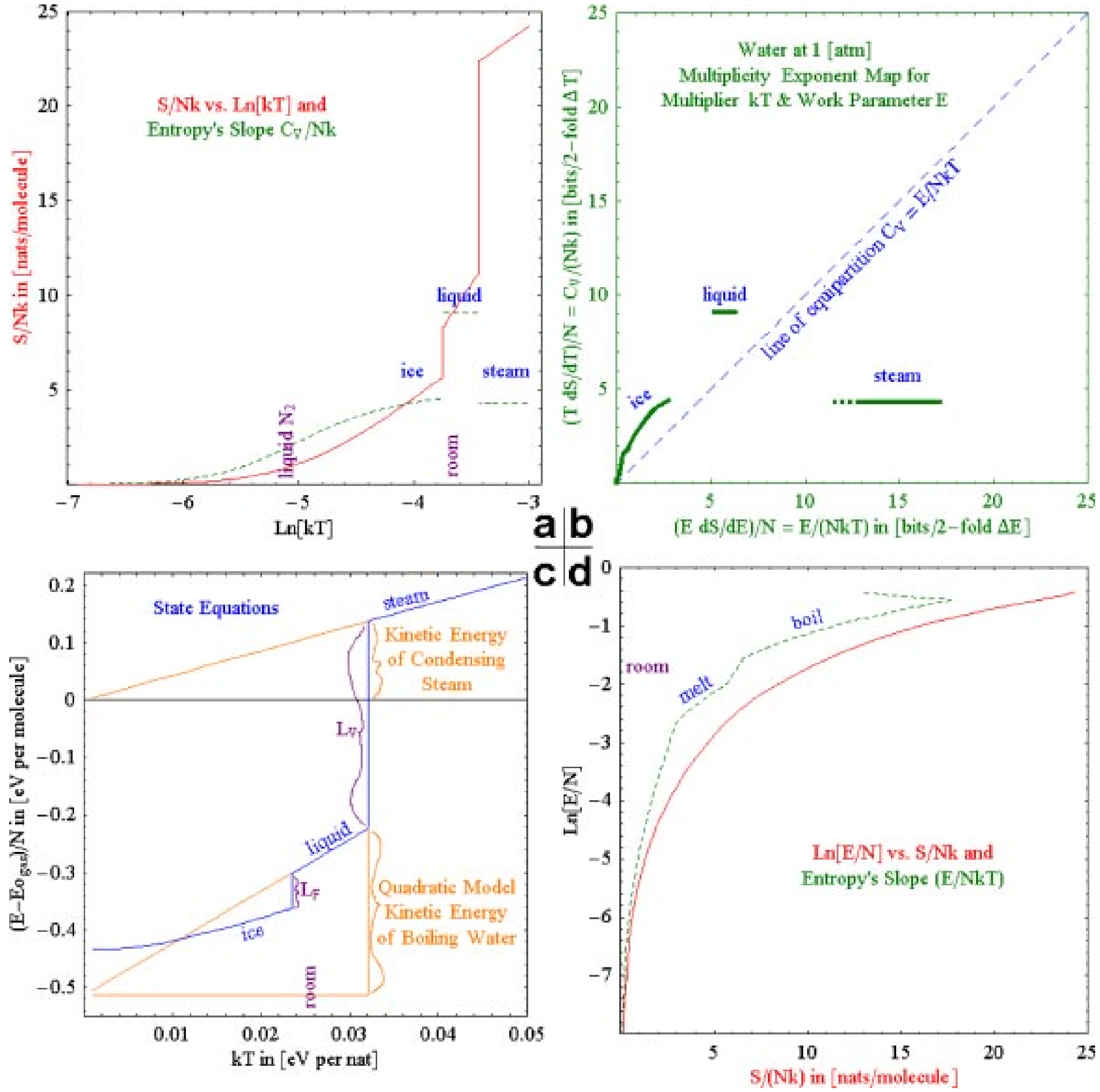

Fig. 3. Plots of the dependence of entropy on the energy per molecule for a steam, water, and Debye model ice system ($\theta = 333$ K assuming a sound speed near 3500 m/s). For simplicity, we have ignored the effect of volume changes on energy, and thus for example, steam has a lower heat capacity than it would in an experiment done at 1 atm. (a) S/N and its slope C_V/Nk (dashed line) vs $\ln(kT)$, (b) multiplicity exponent map C_V/Nk vs E/NkT, (c) equation of state plot E/N vs $dE/dS = kT$, and (d) $\ln(E/N)$ vs S/N and its slope E/NkT (dashed line). The layout of the panels is similar to that in Fig. 2, but curve shapes are now complicated by two phase changes, and the Debye model "freezing out" of thermal energy storage modes in the model used for ice. As discussed in the text, locally defined energy zeros are useful in understanding features seen in the lower left panel, phase changes are more clearly distinguishable in the upper left than in the lower right entropy plots, and the plots on the right side vary drastically with the choice of the energy zero.

population inversion, so that the absolute temperature $T > 0$ and $\Delta T = T$. Thus when $T > 0$, ξ is the average of the heat capacity over temperatures between absolute zero and T.

We have shown that the log–log derivative of the multiplicity with respect to the energy has a simple information theoretic interpretation and is elegantly given (regardless of the units for T) by E/kT as well. From the perspective of an experimentalist, however, E/kT has one glaring disadvantage: its numerical value depends on our choice for the zero of the thermal energy.

To illustrate the problem, consider the cooling of water until it becomes ice (see, for example, the lower left panel in Fig. 3). As water initially cools, the temperature drop per unit energy removed is roughly constant. One might easily say: "Given the uniform rate of decrease of energy with temperature, this system looks like a quadratic system with about 18 degrees of freedom per molecule," so ξ must be about 9 bits per twofold increase in thermal energy. At the freezing point, the temperature stops dropping as the energy continues to be removed. If this system were quadratic, the degrees of freedom would have gone to infinity! Once all is frozen, the temperature continues its drop, this time its rate of decrease suggests a quadratic system with about 8 degrees of freedom per molecule, or ξ closer to 4 bits per twofold increase in thermal energy. Because kT may change little during this experiment, how can E/kT be changing so much? The answer of course is that our inferences about the degrees of freedom have followed not by measuring the total energy E, but only from the rates of change in the energy. Moreover, in the process our inferred "quadratic zero" of thermal energy has been shifting.

The effects of these shifts are seen explicitly in the energy versus temperature plot (lower left) of Fig. 3. In particular, the quadratic (or straight line extrapolated) zero point energy of thermal motion for water is much lower than that for steam, and lower than that for ice. Thus equipartition makes sense for water only if a fictituous zero for the thermal energy is chosen, and then only far from phase changes (where energy is released when particles give up their freedom to move, at the same time enabling new ways of storing thermal energy) and ‘‘freeze out’’ zones (where quantization makes some modes of energy storage less and less accessible). Also note that as in Fig. 2 the lower right column of Fig. 3 shows S vs $\ln E$ and its slope, which from Eq. (2) is nothing other than E/kT using as the origin of energy the actual system energy near $T=0$. As we can see, the value of E/kT using this energy origin shows no regions of constancy at all.

The question then is can we modify our estimate for the log–log derivative of thermal energy so as to reflect only data on temperature changes over a limited energy range? Such a quantity might allow us to probe the ways that thermal energy is being accommodated, one energy range at a time.

V. HEAT CAPACITIES

The no-work (for example, constant volume or constant magnetization) heat capacity, in natural units, can be written as:

$$\frac{C_V}{k}=\frac{\partial}{\partial T}[\xi T]=\left[1+T\frac{\partial}{\partial T}\right]\xi=T\frac{\partial(S/k)}{\partial T}. \tag{4}$$

Here T is absolute temperature in any units you like, and the partials are taken with work parameters (like the volume) held constant. This equation is illustrated in natural units in the top left of Figs. 2 and 3, where one can find both S/k and its derivative C_V/k plotted as a function of $\ln T$. Although the upper left plots of S/k vs $\ln T$ are arrayed symmetrically to the lower right plots of S/k vs $\ln E$, it is clear from Fig. 3 that for water the position of the phase changes and regions between phase changes are much easier to see in the plots of C_V than in the plots of E/kT. The reason in part is that because C_V is an energy derivative, it does not depend on the choice of energy origin, and instead depends only on changes in energy and temperature local to the current state. Thus although ξ depends strongly on the choice of U_0, C_V does not.

To see what C_V actually measures, let us suppose that we have a system whose multiplicity for energies near a ‘‘quadratic zero’’ U_1 obeys

$$\Omega=\left(\frac{U-U_1}{\varepsilon_0}\right)^{\nu N/2},$$

where U_1 is a constant for some range of U values. This physical situation arises often between phase changes, when the thermal energy is distributed over a fixed number of coordinates with ground state energy ε_0. It then follows that

$$S/k=\frac{\nu N}{2}\ln\left(\frac{U-U_1}{\varepsilon_0}\right),\quad \partial S/\partial E=\frac{\nu N}{2(U-U_1)},$$

$$\frac{U-U_1}{kT}=\frac{\nu N}{2}=\frac{C_V}{k}.$$

Thus C_V/k estimates not U/kT, but $(U-U_1)/kT$, where U_1 is a quadratic energy zero inferred by linearly extrapolating a local segment of the E vs kT plot to $T=0$.

This linear extrapolation is trivial for the quadratic system in Fig. 2, and thus $C_V=E/kT$ for all values of E. If the system is not simply quadratic (for example, if it has phase changes or modes of energy storage that freeze out), then C_V/k is simply a local estimate of the thermal energy over kT for a quadratic system. Thus C_V/k modifies the log–log derivative of multiplicity with respect to energy, combining it with its rate of increase per e-fold change in temperature to yield an estimate of $(U-U_1)/kT$.

This observation also provides a different perspective on the mechanism by which the heat capacity diverges during a phase change. The ratio of thermal energy to kT (or the log–log derivative of multiplicity) should have no singularities, because both energy and $1/kT$ are expected to be finite for finite systems. It therefore must be the second term in the two-term expression for heat capacity in Eq. (4), the temperature derivative of ξ, which blows up.

We now return to the lower left panel of Fig. 3 and consider the interesting question: Does the thermal energy of steam increase, decrease, or go negative when it condenses to water? From the above considerations, we can see that the total energy decreases (perhaps even goes negative) if the choice of energy zero is held constant, because steam loses the latent heat of vaporization when it condenses. However, because the specific heat of water at boiling is higher than that for steam, the energy of random motion measured with respect to the locally inferred quadratic zero of thermal energy (for example, at 100 °C) actually goes up. In other words, a small part of the binding energy, liberated when water molecules fall into the potential well of their neighbors, goes to increase the energy of random motion in the condensed phase, relative to that available to particles in the uncondensed gas.

VI. BEYOND EQUIPARTITION

The above observations suggest that introductory texts might consider giving C_V in nats per molecule for common substances (along with the usual values in J/K per unit mass). Some texts already focus their discussion on molar heat capacity divided by the gas constant R, which is the same thing. This dimensionless heat capacity provides insight into the way thermal energy is partitioned among the molecules, even when it does not exactly take on a half-integer value. Well away from phase changes, C_V in natural units also offers clues to the number of degrees of freedom that new thermal energy at a given temperature can access. For example, the high value of this quantity for water (corresponding to 18 degrees of freedom per molecule, divided by 2) tells us something about the collective excitations involving more than one water molecule.

In shifting the focus from historical temperature units to the multiplicities that underlie our inferences, we can say that both E/kT and C_V/k measure the number of bits of increase in state uncertainty per twofold increase in energy with respect to their respective choices for the energy origin. In this sense, they represent physical quantities like mechanical degrees of freedom, but with wider applicability. After all, the concept of degrees of freedom competing on equal grounds for randomly distributed thermal energy presumes not only multiplicities linear with energy on a log–log plot (like a high-temperature Einstein solid), but also quadratic energies (that is, energies proportional to a sum of the squares of some randomly occupied coordinates of state). Even though it is easy to come away from introductory physics with the impression that absolute temperature is a measure of average kinetic energy (see, for example, page 349 of Ref. 17), the idea that every active coordinate gets $kT/2$ even more strongly resists extension to systems with one or more entropy maxima. We begin, however, with a nonquadratic example for which entropy is still a monotonic increasing function of energy.

A. Debye solids

The Debye heat capacity of a solid is an example in which C_V/k depends strongly on temperature.[11] In the low temperature limit, $E=(6N/2)(\pi^4T^3/5\theta^3)\,kT$, so that $\xi=(6N/2)(\pi^4T^3/5\theta^3)$ and $C_V/k=(6N/2)(\pi^4T^3/5\theta^3+T\,3\pi^4T^2/5\theta^3)=4(6N/2)(\pi^4T^3/5\theta^3)$. Here θ is the Debye temperature, which is related to the density and speed of sound in the solid. In the limit $T\rightarrow 0$, only a quarter of C_V/k comes from equipartition (E/kT), the remaining three quarters from the $\ln T$ derivative of E/kT in Eq. (4) (in effect, from the unfreezing of new modes of thermal energy storage). As can be seen from the lower left panel of Fig. 3, such an unfreezing is associated with a lowering of the thermal energy zero locally inferred from the heat capacity, that is, by an increase in the slope as one moves from the solid at absolute zero toward the melting point. Thus attempts to infer degrees of freedom over 2 (that is, ξ) from the heat capacity by assuming that $\xi\simeq C_V/k$ yield a fourfold overestimate of E/kT and hence the effective number of modes for thermal energy storage at low T. This overestimate decreases as the temperature is increased to beyond the Debye temperature θ. In the high temperature limit, $\xi\simeq C_V/k\simeq 3N$, as one would expect from a classical lattice model with $\Omega\propto E^{3N}$.

In summary, the multiplicity exponents examined in Fig. 3 deviate from the quadratic cases (for example, Figs. 1 and 2) by the creation of new coordinates for thermal energy storage in two ways. For the phase changes from ice to water and water to steam, a new fixed set of coordinates is created as the system's old modes of binding are abruptly undone. The heat capacity $C_V/k=T\,dS/dT$ effectively estimates the number of competing coordinates, by locating the new energy zero from which they operate, while $E/kT=E\,dS/dE$ relative to the $T=0$ energy simply measures the total energy in units of the current value of kT. On the other hand, during the heating of ice from $T=0$, the unfreezing of new coordinates adds a second nonzero term to the heat capacity, so that C_V/k overestimates the number of competing coordinates with respect to zero energy for the active set, in this case of the E/kT coordinates competing at any given temperature.

B. Two-state paramagnets

Systems whose thermal energy-storing coordinates have an upper limit on the amount of energy they can accommodate, as for example spins in an external magnetic field, are even more challenging to the idea that thermal energy is distributed in packets of size $kT/2$ over a set of competing modes of storage. As long as the energy is low enough that no single coordinate approaches its maximum energy value, one will find behavior similar to that we have already seen, that is, in solids, liquids, and gases for which the number of ways to accommodate the thermal energy increases as the energy goes up. However, when individual coordinate energies begin to approach their maximum value, the degrees of freedom (as one might imagine) can reverse their enthusiasm to compete.

In particular, there will be but one way (neglecting degeneracies) for the system to store the maximum amount of energy. With only one way to accommodate either the minimum energy or maximum energy, the multiplicities will approach unity at both end points of the continuum. Because large systems may have many ways to store intermediate amounts of energy, the multiplicity (and entropy) as a function of system energy will have a maximum somewhere in between. At such maxima, $dS/dE=1/kT$ will be zero, while on the high energy side dS/dE will be negative, signaling population-inverted states not accessible by thermal contact with reservoirs at positive absolute temperature. Temperature and reciprocal temperature are simply different forms for the multiplier used to characterize a system's propensity to share thermal energy.[8,10] The consideration of spin systems that are capable of taking on (and sharing energy from) negative absolute temperature states suggests that reciprocal temperature has more fundamental significance than temperature, and that the absolute zeros of temperature (approached from negative or positive directions) are at opposite ends of a continuum.[27] But if reciprocal temperature is more fundamental, the heat capacity should be no less simply connected to the log–log derivative via the reciprocal temperature. This rearrangement of Eq. (4) confirms that the heat capacity is simply related to the multiplicity exponent for $1/kT$ as well:

$$\begin{aligned}\frac{C_V}{k}&=-\beta^2\frac{\partial E}{\partial\beta}=-\beta^2\frac{\partial}{\partial\beta}\left[\frac{\partial\ln\Omega}{\partial\ln E}\,\frac{1}{\beta}\right]\\&=\left[1-\beta\frac{\partial}{\partial\beta}\right]\frac{\partial(\ln\Omega)}{\partial(\ln E)}\\&=-\beta\frac{\partial\ln\Omega}{\partial\beta}=-\beta\frac{\partial(S/k)}{\partial\beta}.\end{aligned}\tag{5}$$

Had we historically adopted as our measure of willingness to share energy a power of the uncertainty slope other than $T=\beta^{-1}$ or β^1, say $\gamma=\beta^\alpha$, the heat capacity would more generally become $-\alpha$ times the log–log derivative of multiplicity with respect to that measure. Hence log–log derivatives of multiplicity with respect to the Lagrange multiplier of a conserved quantity (in effect, heat capacities when the conserved quantity is energy) are to within a constant, independent of the specific multiplier choice (for example, temperature or its reciprocal). Thus for the spin system plots (Fig. 4) the multiplier used will be $1/kT$ instead of kT. However, C_V/k from Eq. (5) is still the multiplicity exponent of choice both on historical grounds, and because it remains positive for spin systems.

A note about the partial derivatives in this section is in order. The derivatives are not work-free because they will be taken at constant magnetic field (an intensive parameter) rather than at constant magnetization (the corresponding extensive or work parameter). We have argued that Eqs. (2) and (4) apply generally for partials with all work-parameters held constant. They also are valid for the ideal gas under constant pressure, if energy is replaced by enthalpy. This example will

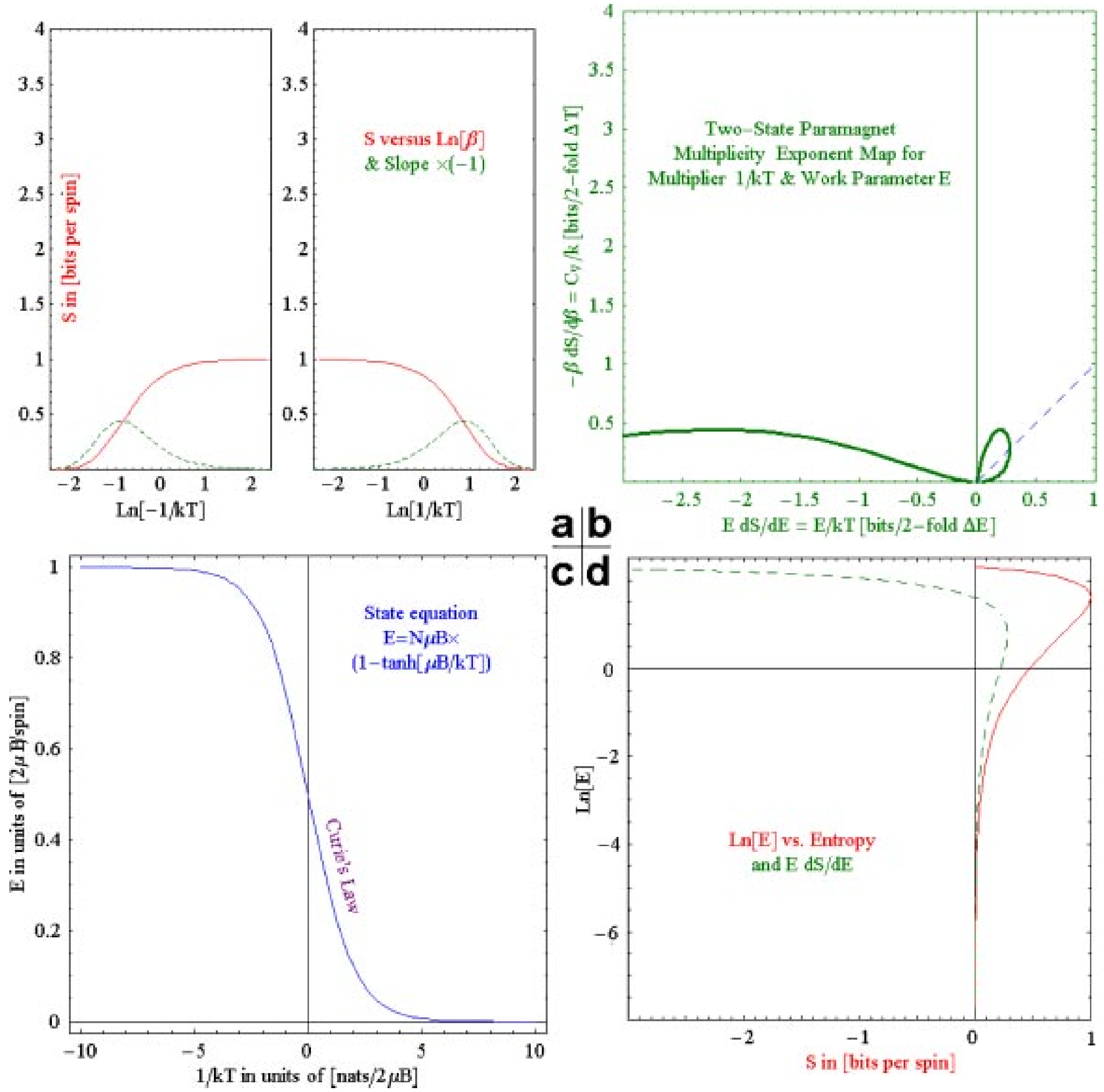

Fig. 4. Plots of the dependence of entropy on energy per spin for a two-state paramagnet consisting of 10 noninteracting spins. (a) S/N and its slope C_V/Nk (dashed line) vs $\ln(1/kT)$, (b) multiplicity exponent map C_V/Nk vs E/NkT, (c) equation of state plot E/N vs $dE/dS=1/kT$, and (d) $\ln(E/N)$ vs S/N and its slope E/NkT (dashed line). The layout is similar to that in Figs. 2 and 3 except that $1/kT$ replaces kT on the left-hand side. As discussed in the text, $1/kT$ runs from ∞ to $-\infty$ as T goes from 0^+ to 0^-, the upper left and lower right entropy plots show a single maximum, and in the exponent map, the heat capacity remains positive while E/kT takes on negative values that correspond to the negative reciprocal temperature side of the entropy peak. If the energy zero is reset halfway between minimum and maximum energy values, the lower right entropy peak becomes symmetric, and the loop in the positive quadrant of the exponent map disappears.

show that they are valid for noninteracting spins. The maximum entropy formalism (with a suitably generalized definition of enthalpy) likely holds clues to the usefulness of multiplicity exponents under a wider range of (for example, ensemble rather than only no-work) constraints,[28] although this is not explored further here.

We begin with a system of half-integral spins with a single entropy maximum. If we expand $S(E)$ to second order about the energy of its maximum, $E_{S\,\max}$, we obtain $S(E)$ as $S(E_{S\,\max})+\frac{1}{2}A(E-E_{S\,\max})^2$, where

$$A=\left(\frac{d^2S}{dE^2}\right)_{E=E_{S\,\max}}<0.$$

Hence for energies near $E_{S\,\max}$, $dS/dE\simeq A(E-E_{S\,\max})$. Thus deviations from $E_{S\,\max}$ are negative for $T>0$, and (at least for small deviations and hence high temperatures) are proportional to the reciprocal temperature $\beta=1/kT$. This proportionality was discovered by Pierre Curie, and bears his name (see, for example, page 105 of Ref. 18). The region of proportionality can be seen in the center of the state equation plot of a two-state paramagnet with 10 noninteracting spins, in the lower left of Fig. 4.

For a system of N half-integer spins, Stirling's approximation allows us to approximate the number of accessible states[12,18] by $\Omega\simeq N^N/(N_\uparrow^{N_\uparrow}N_\downarrow^{N_\downarrow})$. Here $N_\uparrow+N_\downarrow=N$ because all of the N spins are either up or down. If we define the energy relative to the all down state as $E=2\mu BN_\uparrow$, where μ is the magnetic moment per spin, and B is the superposed magnetic field, then we can write Ω as a function of E. Equation (1) then yields $E\simeq 2\mu BN(1-\tanh[x])$, where $x=\mu B/kT$, so that $E/kT=Nx(1-\tanh[x])$ and $C_V/k=N(x/\cosh[x])^2$. Here x takes on positive and negative values, ranging from $+\ln(N)$ to $-\ln(N)$, respectively, for orientation energies E with allowed values from 0 to $2\mu BN$.

Thus for E near to but less than $E_{S\,\max}=\mu BN$, $C_V/k \simeq Nx^2>0$, while $(E-E_{S\,\max})/kT\simeq -Nx^2<0$. Because β for these systems may be positive or negative, E/kT will be negative for some β values, regardless of our choice of the thermal energy zero. The interpretation of E/kT as half of some number of degrees of freedom competing successfully for $kT/2$ units of thermal energy is now complicated by the fact that the degrees of freedom (in this case spins) with maximum energy will be competing to lose the energy they have. When more than half of the spins are at maximum energy, E/kT becomes negative. Although negative degrees of freedom may cause discomfort for some, a negative value for E/kT as the multiplicity exponent should disturb no one because, to paraphrase a related comment by Schroeder,[18] there is no law of physics guaranteeing that there will not be fewer ways to distribute energy as more energy is added.

As shown in the exponent map (upper right) of Fig. 4, C_V/k overestimates E/kT at low temperatures (new degrees of freedom come into play in the upper portion of the loop), although the estimate becomes exact when x decreases to $x \cong 1.279$ (the solution of $e^{-x}=x-1$ for x). At higher temperatures, C_V/k underestimates E/kT, which begins to decrease as x decreases and T increases from this point. The relationship between E/kT and C_V switches again when x passes through zero, because $C_V/k \geqslant 0$ while the energy's multiplicity exponent becomes negative because the uncertainty about the system's state decreases with added energy past this point. Also the average heat capacity goes to zero and no longer equals E/kT, because the average must be obtained piecewise when temperature (unlike reciprocal temperature) breaches the discontinuity from plus infinity to minus infinity.

Thus in addition to information units for heat capacity, we gain from this approach a way to minimize consternation over negative degrees of freedom and to visualize the limits of equipartition as well. The concept of equipartition offers a way to think about the size of the energy fluctuations (as well as the effects of heat flow on temperature) for degrees of freedom (or thermal coordinates) that have more than their ground state energy, provided they have much less energy than any upper limits on energy capacity that might apply. However, as we might infer from the exponent maps (upper right) in Figs. 3 and 4, equipartition's usefulness taken literally is limited only to the simplest of models. Thinking about degrees of freedom over two as energy's multiplicity exponent, and heat capacity as temperature's multiplicity exponent (that is, a multiplicity exponent for energy that uses a locally inferred zero), works well in any case.

VII. SUMMARY

In brief, we have looked at natural (as distinct from historical) units for the common thermodynamic quantities, so that we might explore the possibility that common uses of T, as a measure of the willingness to share thermal energy, have inherited their present emphasis partly because they predate our present understanding of multiplicity. The recognition that a major part of thermodynamics is pure statistical inference, both clarifies the origin and limits of various postulated relationships (like equipartition), and opens the door to second law treatments of mutual information in highly nonequilibrium systems. Applications range from data compression and error correction in communication lines,[29] through present studies of evolving nucleic acid codes[19] and computing,[30,31] to future studies of neural nets and memetic codes.[32] If we look back at how these insights might impact pedagogy, we see that heat capacity is a particularly knotty concept because for most of us, it has always been a change in energy per degree kelvin. In fundamental units, if heat capacity has any dimensions at all, they are information units, because for example "change in energy per unit change in energy per bit" leaves us with nothing but bits in the bargain.

We point out a simple interpretation here. To begin with, the thermal energy E divided by temperature $T=\partial E/\partial S=\beta^{-1}$ is fundamentally $E/kT=E(\partial S/\partial E)$, that is, the log–log derivative of multiplicity with respect to energy (for example, a measure of the bits of uncertainty increase per two-fold increase in energy). The no-work heat capacity then becomes $C_V/k=T(\partial S/\partial T)=-\beta(\partial S/\partial \beta)$, an estimate of E/kT with an energy zero inferred from the local slope of the E vs kT plot. Thus these two quantities are related in concept, regardless of the variable (for example, T or $1/kT$) chosen to keep track of a system's willingness to share energy thermally.

The quantity E/kT plays the role of degrees of freedom over two in quadratic systems, but depends on the energy zero and, regardless, can take on negative values in systems with entropy maxima. Its limitations are those of the concept of equipartition itself. The quantity C_V/k provides deep insight into the ways a system accommodates new thermal energy. We further show here that the correspondence between work-parameter multiplicity-exponents (like E/kT) and the corresponding multiplier multiplicity-exponents (like C_V/k) may look more like a patchwork quilt (in the water case) or spaghetti (in the paramagnet case) than the simple functional relation (or even equality) that one might expect from the usual talk of temperature (dE/dS) as energy per molecule. Just as velocity can be quite independent of acceleration, requiring a map to sort out the relationship between it and its derivative in a particular process, so maps of energy and temperature (or their multiplicity exponents) can be quite diverse once we see temperature as an energy derivative rather than as simply energy per molecule or degree freedom.

Finally, because E/kT and C_V/k are defined here only in terms of state multiplicity and the conserved variable being shared (energy), and are relatively independent of the form chosen for the multiplier (for example, temperature), their analogs in problems that involve the sharing of other conserved quantities (for example, volume, particles, or even dollars) may be easier for students to recognize and put to use.

ACKNOWLEDGMENTS

This work has benefited indirectly from support by the U.S. Department of Energy and the Missouri Research Board, as well as by Monsanto and MEMC Electronic Materials Companies. It has benefited directly from editorial feedback and clarifying reviews (including one by Keith Stine at University of Missouri, St. Louis), interest by students, and interactions with colleages (including David Hestenes, Art Hobson, Ed Jaynes, John Mallinckrodt, Tom

Moore, Dan Schroeder, Bruce Sherwood, and Edwin Taylor) committed to never-ending work on the details of content modernization.

[a] Electronic mail: pfraundorf@umsl.edu

[1] E. F. Taylor, ''The boundaries of nature, special and general relativity and quantum mechanics: A second course in physics,'' Am. J. Phys. **66**, 369–376 (1998).

[2] A. Einstein, *Einstein's 1912 Manuscript on the Special Theory of Relativity—a Facsimile* (George Braziller, New York, 1996), p. 34.

[3] L. Randall, ''Extra dimensions and warped geometries,'' Science **296**, 1422–1427 (2002).

[4] E. Taylor and J. A. Wheeler, *Spacetime Physics* (Freeman, San Francisco, 1963). The first edition contains some material not found in the second edition.

[5] J. S. Bell, ''How to teach special relativity,'' in *The Speakable and Unspeakable in Quantum Mechanics* (Cambridge U.P., Cambridge, 1987), pp. 67–80.

[6] E. Taylor and J. A. Wheeler, *Exploring Black Holes* (Addison–Wesley–Longman, Reading, MA, 2001).

[7] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, IL, 1949).

[8] E. T. Jaynes, ''Information theory and statistical mechanics,'' Phys. Rev. **106**, 620–630 (1957).

[9] E. T. Jaynes, ''Information theory and statistical mechanics. II,'' Phys. Rev. **108**, 171–190 (1957).

[10] M. Plischke and B. Bergersen, *Equilibrium Statistical Physics* (Prentice–Hall, New York, 1989), Chap. 2.

[11] L. A. Girifalco, *Statistical Physics of Materials* (Wiley, New York, 1973), p. 72.

[12] C. Kittel and H. Kroemer, *Thermal Physics* (Freeman, New York, 1980), p. 39.

[13] K. Stowe, *Introduction to Statistical Mechanics and Thermodynamics* (Wiley, New York, 1984), p. 116.

[14] D. S. Betts and R. E. Turner, *Introductory Statistical Mechanics* (Addison–Wesley, Wokingham, England, 1992).

[15] C. Garrod, *Statistical Mechanics and Thermodynamics* (Oxford U.P., New York, 1995).

[16] T. A. Moore, *Six Ideas that Shaped Physics: Unit T* (McGraw–Hill, New York, 1998).

[17] R. Baierlein, *Thermal Physics* (Cambridge U.P., New York, 1999).

[18] D. V. Schroeder, *An Introduction to Thermal Physics* (Addison–Wesley, San Francisco, 2000).

[19] T. D. Schneider, ''Evolution of biological information,'' Nucleic Acids Res. **28** (14), 2794–2799 (2000) or ⟨www-lmmb.ncifcrf.gov/ toms/⟩.

[20] S. Lloyd and J. Paul Penfield, ''Hot bits: Seeing the world in terms of information and entropy,'' National Electrical Engineering Department Heads Association Annual Meeting, Destin, FL, 1999; ⟨www-mtl.mit.edu/Courses/6.095/⟩.

[21] S. Lloyd, ''Use of mutual information to decrease entropy: Implications for the second law of thermodynamics,'' Phys. Rev. A **39** (10), 5378–5386 (1989).

[22] W. K. Wootters and W. H. Zurek, ''A single quantum cannot be cloned,'' Nature (London) **299**, 802–803 (1982).

[23] E. T. Jaynes, ''A note on thermal heating efficiency'' (1991), now in Am. J. Phys. **71** (2), 180–182 (2003).

[24] T. D. Schneider, ''Theory of molecular machines. II. Energy dissipation from molecular machines,'' J. Theor. Biol. **148** (1), 125–137 (1991).

[25] C. H. Bennett, ''Demons, engines and the second law,'' Sci. Am. **257** (11), 108–116 (1987).

[26] E. P. Odum, *Fundamentals of Ecology* (Saunders, Philadelphia, 1971), 3rd. ed.

[27] J. Castle, W. Emmenish, R. Henkes, R. Miller, and J. Rayne, *Science by Degrees: Temperature from Zero to Zero*, Westinghouse Search Book Series (Walker and Company, New York, 1965).

[28] P. Fraundorf, ''Information physics: from energy to codes,'' arXiv:physics/9611022v3.

[29] R. W. Yeung, *A First Course in Information Theory, Information Technology: Transmission, Processing, and Storage* (Kluwer Academic/Plenum, New York, 2002).

[30] N. Gershenfeld, ''Signal entropy and the thermodynamics of computation,'' IBM Syst. J. **35**, 577–585 (1995).

[31] D. Gottesman and H.-K. Lo, ''From quantum cheating to quantum security,'' Phys. Today **53** (11), 22–27 (2000), quant-ph/0111100.

[32] R. Dawkins, *The Selfish Gene* (Oxford U.P., New York, 1989), 2nd ed.

A NEW KIND OF LOVE

Personally, I find Wolfram's enthusiasm for his own ideas refreshing. I am reminded of a comment made by the Buddhist teacher Guru Amrit Desai, when he looked out of his car window and saw that he was in the midst of a gang of Hell's Angels. After studying them in great detail for a long while, he finally exclaimed, ''They really love their motorcycles.'' There was no disdain in this observation. Guru Desai was truly moved by the purity of their love for the beauty and power of something that was outside themselves.

Well, Wolfram really loves his cellular automata . . .

Ray Kurzweil, on Stephen Wolfram's *A New Kind of Science*
(http://www.kurzweilai.net/articles/art0464.html?printable=1,2002).