Simple and synergistic ways to understand the Boltzmann distribution function

Larry Engelhardt, Marie Lopez del Puerto and Norman Chonacky

Citation: American Journal of Physics **83**, 787 (2015); doi: 10.1119/1.4923204
View online: https://doi.org/10.1119/1.4923204
View Table of Contents: https://aapt.scitation.org/toc/ajp/83/9
Published by the American Association of Physics Teachers

Simple and synergistic ways to understand the Boltzmann distribution function

Larry Engelhardt[a)]
Department of Physics and Astronomy, Francis Marion University, Florence, South Carolina 29501

Marie Lopez del Puerto[b)]
Department of Physics, University of St. Thomas, St. Paul, Minnesota 55105

Norman Chonacky[c)]
Department of Applied Physics, Yale University, New Haven, Connecticut 06520-8284

(Received 11 August 2014; accepted 16 June 2015)

This paper presents three approaches to providing undergraduate students with a conceptual understanding of the Boltzmann distribution: (1) a simple logical argument for why it is described by an exponential function, (2) a computational experiment to demonstrate and validate this result, and (3) a computer simulation of a laboratory experiment that allows this result to be observed. Together, these three perspectives complement one another to broaden students' understanding and prepare them for more formal, complete treatments. These examples illustrate how the convergence of theoretical, computational, and experimental approaches, applied to a single physical problem, contribute to a deeper and more unified understanding of statistical systems than could otherwise be had using any one of the methods alone. © *2015 American Association of Physics Teachers.*
[http://dx.doi.org/10.1119/1.4923204]

I. INTRODUCTION

The Boltzmann distribution function is of central importance to statistical mechanics. For example, page 1 of Richard Feynman's book on statistical mechanics begins,[1] "The key principle of statistical mechanics is as follows: If a system in equilibrium can be in one of N states, then the probability of the system having energy E_n is[2]

$$p(E_n) = (1/Z)e^{-E_n/kT}, \tag{1}$$

where

$$Z = \sum_{n=1}^{N} e^{-E_n/kT}, \tag{2}$$

$k =$ Boltzmann's constant, and $T =$ temperature."[3] Here, Eq. (1) defines the Boltzmann distribution function $p(E_n)$ and Eq. (2) defines the partition function Z. Given the importance of the Boltzmann distribution function (BDF), every statistical mechanics book[4–7] includes a derivation of Eq. (1). A variety of alternative derivations have also been presented in the literature,[8–18] as have alternative derivations of other closely related quantities.[19,20]

In addition to being important in statistical mechanics, the BDF is also an important tool for students studying modern physics, computational physics, and condensed matter (or solid state) physics. For example, simulations of the two-dimensional Ising model are a standard problem in computational physics texts,[21–23] and temperature dependence is important for many condensed matter systems.[24–26] Unfortunately for beginning students, derivations of Eq. (1) are quite mathematically sophisticated, involving partial derivatives, series expansions, factorials, Stirling's approximation, and Lagrange multipliers.

The goal of the present article is to provide three much simpler ways to introduce and understand the BDF. Each way involves a different approach and provides a different perspective on the systems and properties to which the BDF applies. In Sec. II, we describe a simple *logical* argument for why the BDF is described by an exponential function. In Sec. III, we describe a *computational* experiment to demonstrate this result. In Sec. IV, we report on a computational simulation inspired by a *laboratory* experiment, which verifies this result.

The Boltzmann distribution function as we apply it will refer to a system S that meets the following conditions: (1) any particles comprising S must be distinguishable—i.e., we will treat them classically; (2) S must be in thermal equilibrium at a temperature T that remains fixed over relevant time scales; (3) particles in S must behave stochastically, i.e., all states of the system must be accessible to all particles, which move among these states by random processes.

As with many of the quotable statements by Feynman, the one above is simple and general, but laden with deep meanings. Before continuing to Secs. II–IV, it is well worth unpacking meanings for *system*, *state*, and *temperature*, as well as clarifying the statement's application in the context of this paper. By "system" we mean any defined set (or subset) of particles. By "state" (or "microstate") we mean a specific configuration of a system with energy E_n. For a given system, there can be multiple different configurations—different microstates—all of which have the same energy E_n. Then the BDF provides the probability, $p(E_n)$, that a system will be found to be in one of the particular microstates with energy E_n. The concept of microstates can be introduced to students using an analogy with dice—there are several ways to get a pair of dice to add up to six, each of which would correspond to a different microstate of the system of dice. There are also aspects of "temperature" that are familiar to all students. For example, lower temperatures correspond to less energy, and heat will flow from a system with high temperature to a system of low temperature. However, temperature does not have a simple, familiar mathematical definition. Rather, its mathematical definition (in our approach) is related to the shape of the BDF, as we discuss in Sec. II.

To understand—and use—the concepts of both "system" and "temperature," it is important to note that there are multiple ways to define a system. For example, when two gases are mixing, the molecules of each type can be considered as a sub-system that together comprise a larger super-system. Only when a system is in thermal equilibrium, and is hence described by a well-defined energy distribution, can it have a well-defined temperature. The temperature of that system is then maintained by the environment that surrounds the system, which might be as simple as a gas-filled balloon (the system) surrounded by a room filled with air (the environment). This external environment is referred to as a "heat bath," as we discuss in the following sections.

II. SIMPLE THEORETICAL ARGUMENT

To understand why Eq. (1) correctly describes the probability distribution of a system in equilibrium, consider a system S that is composed of subsystems X and Y (see Fig. 1). We assume that these subsystems are in equilibrium—both with each other and with a heat bath B—so a single temperature T describes all three systems (X, Y, and S).

Let the energies of the subsystems be E_X and E_Y, and let the energy of the combined system be E_S. The probability of each of these systems occupying a particular microstate will depend on that microstate's energy and will be represented $p_X(E_X)$ for system X, $p_Y(E_Y)$ for system Y, and $p_S(E_S)$ for the combined system. Using this notation, we will now argue why the probability distribution must have the form shown in Eq. (1).

First, we note that energies add, so the energy of the combined system must equal the sum of the energies of the subsystems,

$$E_S = E_X + E_Y. \tag{3}$$

This is so simple that it might seem trivial, but it is important to contrast the mathematics of energies and probabilities: energies *add* whereas probabilities *multiply*. Hence,

$$p_S(E_S) = p_X(E_X) \times p_Y(E_Y). \tag{4}$$

More precisely, "conjunctive" probabilities multiply. This term refers to the probability of two events (e.g., particular microstates with energies E_X and E_Y) *both* occurring. For example, the probability of rolling "snake eyes" (a pair of ones) with two dice is $1/6 \times 1/6 = 1/36$. However, it is true that probabilities multiply only if the two probabilities do not influence one another. This is true for $p_X(E_X)$ and $p_Y(E_Y)$ thanks to the large bath B. If B were not present, then a large value of E_X would necessitate a small value of E_Y(since there would not be much energy left for system Y), so Eq. (4) would not be valid. (See Ch. 1 of Ref. 27 for more about conditional probabilities.)

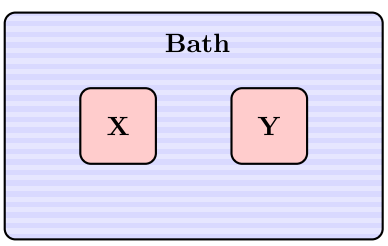

Fig. 1. Visual representation of a system (S) composed of two subsystems (X and Y) in contact with a bath.

We now assume that each of these probabilities $p_i(E)$ is described by a single, universal function of E, which we call $f(E)$, times a system-dependent normalization factor A_i, such that

$$p_i(E) = A_i f(E), \tag{5}$$

where i denotes the system of interest ($i = $ X, Y, or S). To find out the specific form of $f(E)$, we simply substitute Eq. (3) into Eq. (4) to obtain[28]

$$p_S(E_X + E_Y) = p_X(E_X)\, p_Y(E_Y), \tag{6}$$

or

$$f(E_X + E_Y) = a f(E_X) f(E_Y), \tag{7}$$

where $a = A_X A_Y / A_S$. There is one—and *only* one—continuous, real function that has the property that $f(x+y) = af(x) \times f(y)$: the exponential function,[29]

$$f(E) = e^{cE}, \tag{8}$$

with a proportionality constant of $a = 1$. Hence, the Boltzmann distribution function *must* be exponential in order to simultaneously satisfy the additive property of energy and the multiplicative property of probability.

In Eq. (8), there is still an undetermined parameter c. Substituting Eq. (8) back into Eq. (7), it is clear that c must have the same value for all three systems (X, Y, and S) such that $e^{c(E_X+E_Y)} = e^{cE_X} e^{cE_Y}$; and since $a = 1$, we also see that $A_S = A_X A_Y$. The details of both parameters, A_i and c, are discussed below.

In order for Eq. (5) to describe a probability, it must be normalized such that the total probability of all microstates is equal to unity. Summing over all probabilities from Eq. (5) gives[30]

$$\sum_n p_i(E_n) = A_i \sum_n e^{cE_n}, \tag{9}$$

where it should be noted that this is a sum over all microstates n and there can be multiple microstates with the same energy E_n. Setting the right-hand side of Eq. (9) equal to 1 and solving for A_i immediately yields

$$A_i = \frac{1}{\sum_n \exp(cE_n)}. \tag{10}$$

Hence, A_i is simply the normalization factor and is given by Eq. (10). The denominator in Eq. (10) is called the partition function $Z_i = \sum_n \exp(cE_n)$, which has the remarkable property that $Z_S = Z_X Z_Y$. From this, it also follows that $A_S = A_X A_Y$, satisfying that requirement from Eqs. (7) and (8).

We expect the probability distribution to depend on temperature, so the remaining parameter c must be a function of T. To determine exactly how c depends on T, we now present some simple physical arguments. At a low temperature, we know that a system will have a low energy—i.e., low energies have a high probability, and high energies have a low probability. Thus, $p_i(E)$ must be a decreasing function of E,

so the parameter c must be negative. To avoid using a negative parameter, we introduce a new parameter, $\beta \equiv -c$, so Eq. (5) becomes

$$p_i(E) = A_i \exp(-\beta E), \tag{11}$$

where $\beta > 0$. Note that the negative sign also makes it clear that $p_i(E) \to 0$ for large E, which is necessary for normalization.

Figure 2 shows the behavior of Eq. (11) for different values of β. Physically, we know that as temperature is increased, higher energies become more probable; but from Fig. 2, β must be decreased in order to increase the probability of high energies. Therefore, β must be inversely related to temperature. In particular, it turns out that β is inversely proportional to T, the proof of which is beyond the scope of the present argument. If we were to assume a proportionality constant of 1, this would yield $p_i(E) = A_i \exp(-E/T)$; however, $\exp(-x)$ is only defined for dimensionless values of x. This can be corrected by replacing $T \to kT$, where k is a constant with dimensions of energy/temperature (called the Boltzmann constant), such that $E/(kT)$ becomes a dimensionless ratio. This replacement leaves us with the result

$$\beta = \frac{1}{kT}, \tag{12}$$

and upon substituting Eq. (12) into Eqs. (11) and (10), we obtain Eqs. (1) and (2).

It should be stressed that the argument presented above applies to any type of system. We have not specified how the interactions within a system give rise to the microstate energies, so systems X and Y could describe gases, or lattices of magnetic moments, or any other type of system, including systems with negative energies such as the Lennard-Jones or Ising models. In addition to being general, the other advantage of this argument is that it does not require sophisticated mathematics. The disadvantage of this argument is that it is relatively abstract. In Sec. III, we introduce a second method for understanding the BDF that is somewhat less abstract in that it involves actual numbers not just equations. Then in Sec. IV, we introduce an even more concrete method that involves actual physical degrees of freedom—i.e., positions and velocities.

III. SIMPLE COMPUTATIONAL EXPERIMENT

In this section, we introduce a simple computational experiment that allows one to accomplish two things without any sophisticated mathematics or any sophisticated programming:

(1) obtain the exponential form of the Boltzmann distribution function (BDF), given in Eq. (1); and
(2) observe how the precise shape of this distribution depends on the size of the bath.

To do this, we simulate N particles, where the system of interest S consists of a *single* particle; and the other $N-1$ particles serve as the bath B. We observe the distribution of single-particle energies and explore how this distribution depends on the value of N. This simulation is available online, both as an executable Java program that includes built-in student exercises[31] (created using *Easy Java Simulations*[32]), and as Python code,[33] which was used to generate Figs. 3–5.

A key assumption of this simulation is that the total energy of the N-particle super-system (S + B) is a constant: E_N. Each iteration of the simulation then consists of generating $N-1$ random numbers r in the range $0 < r < E_N$, which will be used to partition the total energy among the N particles. The partitioning is shown graphically in Fig. 3, which displays sample results for $N = 5$ particles with total energy $E_N = 1.0$ (using arbitrary units). The vertical red lines represent one set of four random numbers, $r = \{0.19, 0.29, 0.66, 0.72\}$, and the numbers near the top of Fig. 3 represent the energies of the five particles. These energies are obtained by subtracting the values of the consecutive (sorted) random numbers. The algorithm that we use for this simulation is then as follows:

```
(1) Store N-1 random numbers in an array, A
(2) Sort the array from smallest to largest
(3) Append the value 0 to the beginning of A
(4) Append the total energy to the end of A
(5) Compute the energy of each particle by
    subtracting the consecutive values of A.
```

For the example shown in Fig. 3, the array would contain the values $A = \{0, 0.19, 0.29, 0.66, 0.72, 1\}$. This process is then repeated many times, each time using a different set of random numbers, to generate an ensemble of many microstates for these N particles, and all single-particle energies are recorded.

Figure 4 shows a histogram of results that were obtained using $N = 5$ particles, repeated 10,000 times, for a total of 50,000 single-particle energies. (It does not matter which of the N particles is chosen as the "system" particle, so all N of the particle energies can be included.) The plot is clearly a decreasing function of E that looks like it might be

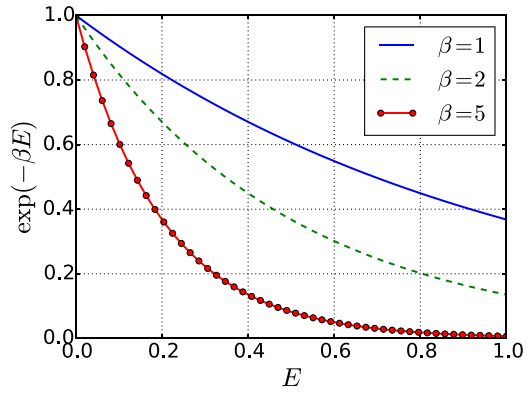

Fig. 2. Plot of Eq. (11) for three different values of β (with $A_i = 1$).

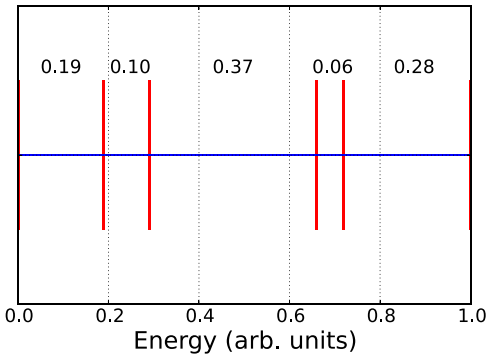

Fig. 3. Partitioning of energy among $N = 5$ particles in a microcanonical ensemble. This requires $N - 1 = 4$ random numbers.

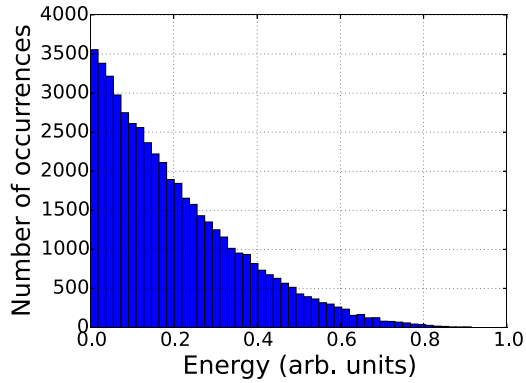

Fig. 4. Histogram resulting from the partitioning of energy E_N among $N = 5$ particles, repeated $R = 10,000$ times.

exponential—consistent with the BDF. In particular, Fig. 4 looks very similar to the $\beta = 5$ curve from Fig. 2. To test whether or not, this histogram is actually an exponentially decreasing function of E, the same data can be plotted using a semi-log scale, so any exponential function will appear linear.

The top curve in Fig. 5 shows the same data as in Fig. 4, but now using a semi-log scale. Clearly for $N = 5$, these data are not an exponential function of E. This is because the (four-particle) bath is only four times larger than the (one-particle) system of interest, whereas the Boltzmann distribution requires that the number of particles in B is much larger than the number of particles in S. The other data in Fig. 5 show that as N increases—i.e., as the bath of $N - 1$ particles gets larger—the results become more linear on this semi-log scale. For $N = 20$ (the bottom curve in Fig. 5), the data appear quite linear, which is to be expected when the bath is much larger than the system.

It is easy to understand—at least qualitatively—why these histograms decrease more rapidly with increasing N. For $N = 2$, there is a single value random number r, so all energies are equally probable (in the range $0 < r < E_N$), resulting in a flat histogram. For $N = 3$, when there are two random numbers, the histogram becomes a decreasing function for a very simple reason: In order to observe *one* particle with a very large energy, the other *two* particles must have very small energies. Hence, small energies will be observed more often than large energies. For larger values of N, this effect becomes even more pronounced. Moreover, for large values

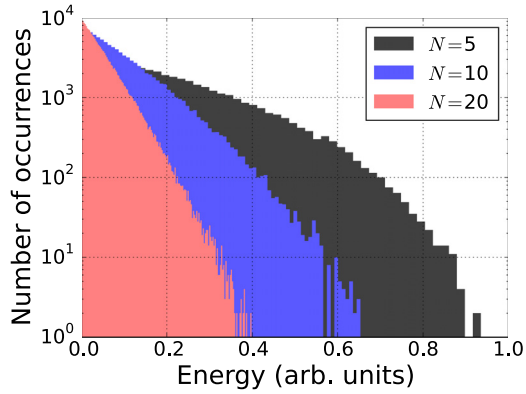

Fig. 5. Histogram showing how many times different ranges of energies were encountered. For each value of N, the partitioning was repeated $R = 10,000$ times (i.e., the ensemble size is 10,000).

of N, it becomes extremely improbable that a large energy (E close to E_N) will *ever* be observed. This is because all $N - 1$ random numbers would need to simultaneously have a value very close to 0 or E_N, like rolling 100 dice and observing all ones or all sixes.

The strengths of the simulation described in this section are that it clearly demonstrates how randomness is central to the emergence of the BDF and that it is *computationally* simple. The disadvantage is that it does not represent a particularly simple *physical* system. Specifically, the process of generating uniformly distributed random numbers to partition the energy assumes a uniform density of states, which is not a feature of most physical systems. One physical system with a uniform density of states is the one-dimensional ultra-relativistic ideal gas, with $E = pc$. Another is the one-dimensional quantum harmonic oscillator, where $E = (n + \frac{1}{2})\hbar\omega$, but this quantum system has discrete—rather than continuous—degrees of freedom. The energy partitioning for a system of quantum harmonic oscillators should therefore use discrete random numbers; a discrete version of the simulation is also provided online.[31]

The partitioning of a one-dimensional number line is by no means a new idea, but we believe the treatment presented here is able to provide new insights, as discussed both above and in Sec. V. For example, this process—when applied to money instead of energy—provides a model of exponentially decaying wealth distribution, as discussed in Ref. 34. The partitioning of discrete energy quanta is also discussed in Sec. 4.2 of Ref. 35, but that discussion focuses on the enumeration of microstates as opposed to the random partitioning of the (one-dimensional) energy line. A similar process is carried out in Ref. 36 using a spreadsheet to compute the entropy of a system of quantum harmonic oscillators; this approach also appears as Problem 6.1 of Ref. 37.

IV. A TWO-LEVEL SYSTEM: EXPERIMENT AND SIMULATION

In this section, we present a simple physical system that can be used to experimentally and computationally verify the Boltzmann distribution function for a two-level system. Prentis[38] proposed a simple mechanical experiment consisting of a ping-pong ball that is allowed to move between two steps of area a_0 and a_1 with height difference h, as shown in Fig. 6. A motorized *Squiggle Ball* is placed in each level and serves the role of the bath B for the ping-pong ball (the system S) by re-establishing thermal equilibrium: As long as the Squiggle Balls both vibrate equally energetically and collide with the ping-pong ball several times before it switches levels, the average kinetic energy of the ping-pong ball will be the same on both levels.

The time t that the ping-pong ball spends on each of the levels of this system can be related to the probability p of finding the ping-pong ball on a particular level.[39] Students can understand this intuitively by considering the *random picture* metaphor as suggested by Bao and Redish:[40] students imagine taking pictures of the system at random times and then using those pictures to make predictions of where the ping-pong ball is most likely to be found. Following-up on the reasoning presented in Secs. II and III, here we first assume that these probabilities are described by a Boltzmann distribution function, and then we obtain data (shown in Fig. 8) to verify this assumption.

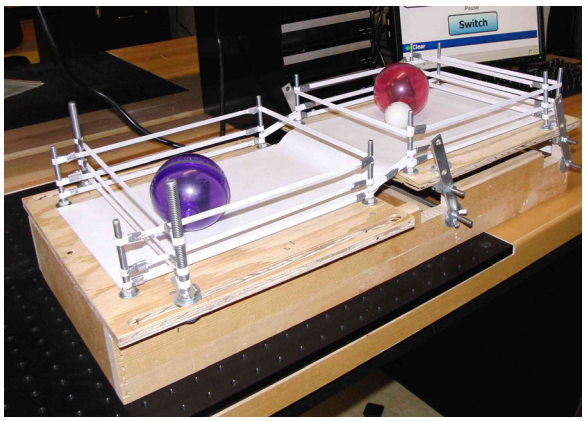

Fig. 6. A Boltzmann machine as proposed by Prentis (Ref. 38).

The connection between time and probability can be expressed mathematically as

$$t \propto p = \frac{\Omega\, e^{-E/kT}}{Z}, \tag{13}$$

where t corresponds to the time the ping-pong ball spends in the level with energy E, p represents the probability of finding the ping-pong ball in the top (or bottom) level, and Ω is the degeneracy of the top (or bottom) level. All locations within a given level have the same potential energy, so the degeneracy of a level is proportional to its area a. (A more detailed discussion can be found in Ref. 38). Thus, the ratio of the times that the ping-pong ball spends on each of the levels is

$$\frac{t_1}{t_0} = \frac{p_1}{p_0} = \frac{\Omega_1 e^{-E_1/kT}/Z}{\Omega_0 e^{-E_0/kT}/Z} = \frac{a_1}{a_0} e^{(E_0-E_1)/kT}. \tag{14}$$

Since the average kinetic energies are equal, the energy difference between the two levels is $E_0 - E_1 = -mgh$, where m is the mass of the ping-pong ball, g is the gravitational field strength near the Earth's surface, and h is the height difference between the two levels. Thus,

$$\frac{t_1}{t_0} = \frac{a_1}{a_0} e^{-mgh/kT}. \tag{15}$$

This experiment is very illustrative but is also time-consuming because one must track the ping-pong ball for several minutes for each data point. We therefore propose the following computational experiment that allows for quick data collection and exploration of different parameter values. A virtual "ball" is placed in a two-level system, it is given an initial random position and velocity, and it is then allowed to move according to classical mechanics. To simulate the effect of the bath (or the Squiggle Balls in the experiment), the velocity of the particle is again randomized every X number of steps. We will refer to each randomization of the velocity as a "kick," though be aware that *for simplicity of the model*[41] the previous velocity is replaced, not modified, in this process. Two versions of this simulation are provided in the online supplement:[33] an *Easy Java Simulations* version[42] that includes built-in student exercises and was used to produce Fig. 7, and a MATLAB version that was used to produce Fig. 8. Pseudocode for the simulation method is given below.

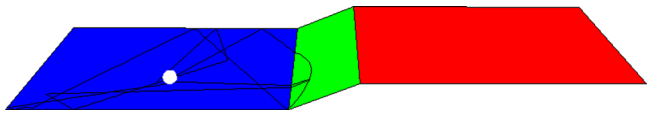

Fig. 7. Screenshot from the *Easy Java Simulations* version of the two-level (Boltzmann machine) simulation.

```
(1) Inputs:
     Width of box, W
     Length of lower level, L0
     Length of upper level, L1
     Height difference between levels, h
     Gravitational acceleration, g
     Time-step, dt
     Number of time-steps, N
     Maximum speed of the kick, vmax
(2) Generate a random initial position
    vector:
     x(0) = [(L0 + L1)*random, W*random]
(3) Generate a random initial velocity
    vector:
     r1 = random
     r2 = random
     v = [r1*cos(2*pi*r2), r1*sin(2*pi*r2)]
(4) Loop: for i from 1 to N
    (a) Modify position:
          x(i) = x(i-1) + v*dt
    (b) If statements to check for collisions
        with walls, modify velocity and
        position accordingly
    (c) If statement to check if particle has
        crossed step going up
        (i) If 0.5*m*vx*vx > m*g*h,
            modify velocity
        (ii) Else, reflect particle back
    (d) If statement to check if particle has
        crossed step going down
            Modify velocity
    (e) Every X time-steps, randomly modify
        v:
        if mod(i,X) is 0
         r1 = random
         r2 = random
         v = vmax*[r1*cos(2*pi*r2),
         r1*sin(2*pi*r2)]
    (f) Record number of time-steps the
        particle
        spends in each level
(5) Calculate ratio of time spent in each
    level.
```

With this simulation, one can easily vary the temperature of the bath, the dimensions of the two steps, and the height difference between the steps. Data collection is reduced to a few seconds per data point, allowing students to explore a wide range of parameter values and see how these parameters affect the time ratio.

Figure 8 shows the results of running the simulation for three different ratios of step area, a_1/a_0, and two different temperatures. For each ratio the simulation was run for 10^6 time steps, with random changes to the particle's velocity every 100 time steps. The y-intercept in Fig. 8 corresponds to the ratio of the areas of the steps, as is expected from Eq. (15). Taking the natural logarithm of both sides of Eq. (15), we obtain

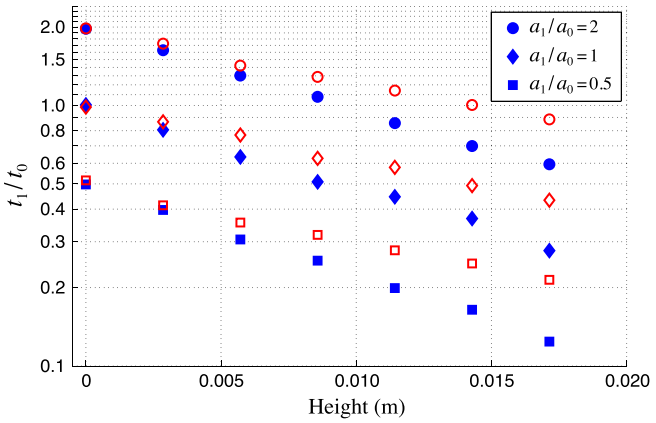

Fig. 8. Results of the two-level simulation for different ratios of step area. The open (red) markers correspond to increasing v_{max} by a factor of $\sqrt{2}$.

$$\ln\left(\frac{t_1}{t_0}\right) = \ln\left(\frac{a_1}{a_0}\right) - \frac{mg}{kT}h, \tag{16}$$

so the slope of each line is equal to $-mg/(kT)$. From the equipartion theorem, $kT = \frac{1}{2}m\overline{v^2}$ in two dimensions, where $\overline{v^2}$ is the average of the square of the speed. Hence, the temperature of the bath—and thus the slope of the plots in Fig. 8—can be changed by changing the maximum speed v_{max} of the random distribution that is periodically used to reset the particle's velocity. It is particularly illustrative to increase the value of v_{max} by a factor of $\sqrt{2}$. This doubles the temperature and thus decreases the slope of $\ln(t_1/t_0)$ vs h by a factor of 2, as can be seen by comparing the filled and open symbols in Fig. 8.

To verify that the simulation produces slopes that have the correct numerical values, we have also provided two alternate implementations[33,42] in which velocities are generated using the following nonuniform distributions: (1) using speeds v generated using the Maxwell-Boltzmann distribution, which allows the user to directly specify the temperature; and (2) using values of v^2 drawn from a Gaussian distribution with a user-defined peak (most probable) value for v^2, which, since the peak of the distribution corresponds to $\overline{v^2}$, allows the user to specify the value of $\overline{v^2}$ and hence specify $kT = \frac{1}{2}m\overline{v^2}$. However, in order to get numerically correct slopes, one must be careful to check for convergence, as we discuss in Sec. V. Finally, it is interesting to note the robustness of this method. Regardless of which distribution is used for randomly generating the speeds, one will still obtain an exponential distribution function for this two-level system.

V. DISCUSSION

Let us now discuss some of the connections among our three different ways of approaching the Boltzmann distribution function (BDF), particularly with regard to the bath and how this bath achieves randomization.

The simple approaches of Secs. II (theory) and III (computational) serve to quickly liberate the BDF from the more mathematically intense derivations that were mentioned in Sec. I, and that we feel should be avoided for a student's first introduction to the BDF. Instead, we advocate introducing the BDF to beginning students by justifying its exponential character using the simple logical argument presented in Sec. II and the simple computational experiment presented in Sec. III. Then students can validate the exponential character of the BDF by examining the relative particle populations in a two-state system as described in Sec. IV (simulation). We have used this simulation, in conjunction with the physical experiment that it simulates, as a modern physics laboratory experiment; it could also be used as an experiment in a solid state (or condensed matter) physics course.

Our three approaches described above provide different, complementary perspectives on the "heat bath" that must always be present for the BDF to apply.

First, the logical argument of Sec. II demonstrates the necessity of the bath from a mathematical perspective. Specifically, in order for Eq. (4) to be correct (that the probabilities multiply), it is necessary for the individual probabilities, $p(E_X)$ and $p(E_Y)$, to be independent. This independence is achieved via a large bath, which physically separates systems X and Y, as shown schematically in Fig. 1. If the two systems were not well separated (e.g., if the bath were not sufficiently large), then Eq. (4) would not be valid and we could not conclude that the probability distribution is exponential.

Next, the computational partitioning experiment of Sec. III provides an opportunity to explore the effect of the bath size by varying the number of particles N. We immediately see that the probability distribution is not exponential unless the bath is much larger than the system. If the bath is too small, then high-energy states are over-represented, while low-energy states are underrepresented.

Finally, the two-level system described in Sec. IV provides insight into the physical origin of the BDF by incorporating the physical dynamics of a macroscopic system. Here, the bath is represented by random velocity "kicks" given to the particle and can be varied either by (1) changing the parameters of the random velocity distribution, or (2) changing how often a kick occurs. In case (1), if the kicks are too small—so that the ball does not receive enough energy from the bath to climb the ramp—the ball will be stuck in the lower level. To explore this phenomenon for yourself, simply decrease the value of the maximum kick speed v_{max}; you will see that the data from Fig. 8 are only linear in the regime $\frac{1}{2}mv_{max}^2 \gg mgh$, and the time spent on the top surface is zero if $\frac{1}{2}mv_{max}^2 < mgh$. In case (2), the ball will also spend too much time in the lower level if the kicks do not occur sufficiently often. Thus, in addition to checking the convergence of the simulation with respect to the time step (that it is small enough) and the number of time steps (that it is large enough), one needs to check for convergence with respect to the frequency of kicks. For the system to reach equilibrium, there must be many kicks at one level before a transition to the other level occurs. In both cases, we see that if the bath is too small (in frequency or intensity), then the higher-energy states are underrepresented compared with the BDF.

Although Sec. IV involves a *classical* system, the Boltzmann distribution is also a good approximation to *quantum* distributions in thermal equilibrium when the occupancy of each state (the average number of particles per state) is much smaller than 1. For this reason, the two-level system from Sec. IV provides a useful model for discussing a variety of quantum systems: the ratio of atoms in a gas that are in an excited state vs the ground state (for example, for hydrogen atoms in the Sun or atoms in a laser); the number of conduction electrons in a semiconductor; and the fraction of nuclei whose spin is anti-aligned with the magnetic field in a nuclear magnetic resonance experiment, among others.

The juxtaposition of experiment, theory, and simulation presented in this paper illustrates the synergy among these

three facets of physics for understanding statistical systems. This "three-fold way" has changed the way science is conducted and is changing the way physics is taught. Given the importance and myriad applications of the Boltzmann distribution in modern physics, students should develop a conceptual understanding of this function along with a more general facility with statistical reasoning. The simplified theoretical approach of Sec. II aims to give students this conceptual understanding without getting bogged down in complex mathematics. The experiment in Sec. IV gives students a visual representation of a simple system that they can connect to their prior knowledge of kinematics. Finally, the computer simulations in Secs. III and IV enhance this understanding and build intuition by allowing students to easily and quickly explore all important system parameters and to make discoveries of their own.

ACKNOWLEDGMENTS

This work was supported in part by NSF-TUES grant DUE-1140034. The authors thank Jeremiah Bartz for useful comments related to Ref. 29. The authors also thank the anonymous referees for their careful reading and useful comments.

[a]Electronic mail: lengelhardt@fmarion.edu
[b]Electronic mail: mlpuerto@stthomas.edu
[c]Electronic mail: norman.chonacky@yale.edu

[1]R. P. Feynman, *Statistical Mechanics: A Set Of Lectures*, 2nd ed. (Westview Press, Boulder, CO, 1998).
[2]More precisely, this is the probability of the system occupying a particular microstate with this value of energy. If there is more than one microstate with the same energy, then the probability of the system having this energy is the product of this probability times the number of microstates that have that energy (i.e., the "degeneracy").
[3]Feynman used Q for the partition function in 1972, but today most sources use Z, as we do here.
[4]R. K. Pathria, *Statistical Mechanics*, 3rd ed. (Academic Press, Boston, 2007).
[5]F. Reif, *Fundamentals of Statistical and Thermal Physics* (Waveland Press, Long Grove, IL, 2008).
[6]L. E. Reichl, *A Modern Course in Statistical Physics*, 3rd ed. (Wiley-VCH, Weihheim, 2009).
[7]R. H. Swendsen, *An Introduction to Statistical Mechanics and Thermodynamics* (Oxford U.P., New York, 2012).
[8]F. T. Wall, "Alternative derivations of the statistical mechanical distribution laws," Proc. Nat. Acad. Sci. **68**, 1720–1724 (1971).
[9]G. D. Hoyt, "A more concise derivation of the Maxwell distribution," Am. J. Phys. **39**, 845–845 (1971).
[10]R. C. Dunbar, "Deriving the Maxwell distribution," J. Chem. Ed. **59**, 22–23 (1982).
[11]D. K. Russell, "The Boltzmann distribution," J. Chem. Ed. **73**, 299–300 (1996).
[12]S. A. C. McDowell, "A simple derivation of the Boltzmann distribution," J. Chem. Ed. **76**, 1393–1394 (1999).
[13]R. B. Shirts and M. R. Shirts, "Deviations from the Boltzmann distribution in small microcanonical quantum systems: Two approximate one-particle energy distributions," J. Chem. Phys. **117**, 5564–5575 (2002).
[14]C. W. David, "Boltzmann without Lagrange," J. Chem. Ed. **83**, 1695–1697 (2006).
[15]R. López-Ruiz and X. Calbet, "Derivation of the Maxwellian distribution from the microcanonical ensemble," Am. J. Phys. **75**, 752–753 (2007).
[16]V. Gurarie, "The equivalence between the canonical and microcanonical ensembles when applied to large systems," Am. J. Phys. **75**, 747–751 (2007).
[17]R. López-Ruiz, J. Sañudo, and X. Calbet, "A geometrical derivation of the Boltzmann factor," Am. J. Phys. **76**, 780–781 (2008).
[18]A. Walstad, "On deriving the Maxwellian velocity distribution," Am. J. Phys. **81**, 555–557 (2013).
[19]M. Campisi and D. H. Kobe, "Derivation of the Boltzmann principle," Am. J. Phys. **78**, 608–615 (2010).
[20]M. A. M. Versteegh and D. Dieks, "The Gibbs paradox and the distinguishability of identical particles," Am. J. Phys. **79**, 741–746 (2011).
[21]N. J. Giordano and H. Nakanishi, *Computational Physics*, 2nd ed. (Addison-Wesley, Upper Saddle River, NJ, 2005).
[22]H. Gould, J. Tobochnik, and W. Christian, *An Introduction to Computer Simulation Methods: Applications to Physical Systems*, 3rd ed. (Addison-Wesley, San Francisco, 2006).
[23]R. H. Landau, J. Páez, and C. C. Bordeianu, *A Survey of Computational Physics: Introductory Computational Science* (Princeton U.P., Princeton, 2008).
[24]M. A. Omar, *Elementary Solid State Physics: Principles and Applications*, 4th ed. (Addison-Wesley, Reading, MA, 1994).
[25]C. Kittel, *Introduction to Solid State Physics*, 8th ed. (Wiley, Hoboken, NJ, 2005).
[26]N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Holt, Rinehart and Winston, New York, 1976).
[27]A. B. Downey, *Think Bayes; Bayesian Statistics Made Simple* (O'Reilly Media, Sebastopol, CA, 2013), <http://greenteapress.com/thinkbayes/thinkbayes.pdf>.
[28]A similar argument was also made on page 312 of S. O. Kasap, *Principles of Electronic Materials and Devices*, 3rd ed. (McGraw-Hill, New York, 2006).
[29]For a proof that Eq. (8) is indeed a unique solution of Eq. (7), provided that we restrict the possible solutions to continuous, real-valued functions—i.e., physically sensible solutions—see Exercise 18.46 from E. Hewitt and K. Stromberg, *Real and Abstract Analysis: A Modern Treatment of the Theory of Functions of a Real Variable* (Springer, Berlin, 1965).
[30]For systems with continuous degrees of freedom—e.g., an ideal gas—this sum is replaced with an integral.
[31]L. Engelhardt and N. Chonacky, "Boltzmann distribution from a microcanonical ensemble model" (Java software), <http://www.compadre.org/osp/items/detail.cfm?ID=11563> (2011).
[32]W. Christian and F. Esquembre, "Modeling Physics with Easy Java Simulations," Phys. Teach. **45**, 475–480 (2007); L. Engelhardt, "Introduction to EJS Video Tutorial," <http://www.compadre.org/OSP/tutorials/EJS_Tutorial/> (2010).
[33]See supplementary material at http://dx.doi.org/10.1119/1.4923204 for all of the simulations and source code associated with this article.
[34]V. M. Yakovenko and J. B. Rosser, "Statistical mechanics of money, wealth, and income," Rev. Mod. Phys. **81**, 1703–1725 (2009).
[35]H. Gould and J. Tobochnik, *Statistical and Thermal Physics: With Computer Applications* (Princeton U.P., Princeton, NJ, 2010).
[36]T. A. Moore and D. V. Schroeder, "A different approach to introducing statistical mechanics," Am. J. Phys. **65**, 26–36 (1997).
[37]D. V. Schroeder, *An Introduction to Thermal Physics* (Addison-Wesley, San Francisco, 1999).
[38]J. J. Prentis, "Experiments in statistical mechanics," Am. J. Phys. **68**, 1073–1083 (2000).
[39]This is known as the "ergodic hypothesis," which allows us to relate the parameters of the experimental system to the parameters in the statistical distribution, $p(E)$. See P. Ehrenfest and T. Ehrenfest, *The Conceptual Foundations of the Statistical Approach in Mechanics* (Dover, New York, NY, 1990), pp. 21–26; Shang-keng Ma, *Statistical Mechanics* (World Scientific, Philadelphia, PA, 1985), pp. 440–445; C. J. Thompson, *Mathematical Statistical Mechanics* (Macmillan, New York, NY, 1972), pp. 214–217.
[40]L. Bao and E. F. Redish, "Understanding probabilistic interpretations of physical systems: A prerequisite to learning quantum physics," Am. J. Phys. **70**, 210–217 (2002).
[41]To model the bath using random impulses would require that a damping force also be included to keep the energy from growing indefinitely. Such a process—referred to as "Langevin dynamics"—would be more physically realistic, but it would lack the simplicity of the method presented here. See D. S. Lemons and A. Gythiel, "Paul Langevin's 1908 paper "On the Theory of Brownian Motion" ["Sur la théorie du mouvement brownien," C. R. Acad. Sci. (Paris) **146**, 530–533 (1908)]," Am. J. Phys. **65**, 1079–1081 (1997).
[42]L. Engelhardt and M. Lopez del Puerto, "Boltzmann machine model" (Java software), <http://www.compadre.org/osp/items/detail.cfm?ID=13330> (2014).