Journal of Statistical Physics, Vol. 62, Nos. 1/2, 1991

On the Limiting Distribution of Pair-Summable Potential Functions in Many-Particle Systems

Kevin A. O'Neil[1] **and Richard A. Redner**[1]

Received June 18, 1990

For systems with finite phase space volume, the density of states can be viewed as a multiple of the probability density of the energy, when the phase space variables are independent uniformly distributed random variables. We show that the distribution of a random variable proportional to the sum of pairwise interactions of independent identically distributed random variables converges to a limiting distribution as the number of variables goes to infinity, when the interaction satisfies certain homogeneity requirements. The moments of this distribution are simple combinations of cyclic integrals of the potential function. The existence of this limit gives information about the structure function of some systems in statistical mechanics having pair-summable interactions, even in the absence of a thermodynamic limit. The result is applied to several examples, including systems of two-dimensional point vortices.

KEY WORDS: Convergence in distribution; pair-summable potential; point vortices; statistical mechanics.

1. INTRODUCTION

If a thermodynamic system is isolated (the energy constant), the thermodynamic properties of the system are determined by the microcanonical ensemble. The fundamental object is the structure function, which gives the density of states of the system as a function of energy. Identifying volume with probability, the structure function can be viewed (up to a multiplicative constant) as the probability density function for the potential energy when the particle positions are regarded as independent, uniformly distributed random variables.

Many-particle systems in which the potential energy can be expressed as the sum of interactions between pairs of particles, and having finite

[1] Department of Mathematical and Computer Sciences, University of Tulsa, Tulsa, Oklahoma 74104.

822/62/1-2-26

0022-4715/91/0100-0399$06.50/0

configuration space volume, are common in physics. The purpose of this paper is to show that for a certain class of interaction functions, the probability distribution of the potential energy *per particle* approaches a limiting distribution as the number of particles n goes to infinity. The moments of the limiting distribution are given explicitly, in terms of cyclic integrals of the interaction function. (The integrals are reminiscent of the cluster integrals which appear in the theory of imperfect gases,[1] but here the potential need not be short ranged.) Those functions associated with the system which depend on this distribution thus tend to limiting forms as $n \to \infty$, even in the absence of a thermodynamic limit. If, however, the entropy density tends to a limiting function of the energy density when the size of the system is adjusted with n to keep the mean particle density constant, then these limits coincide with the thermodynamic limit.[2]

The restrictions placed on the interaction function are quite mild. They essentially require that the interaction function be homogeneous and have zero mean, and that the moments of the energy distribution of the n-particle system be defined for all n.

The next section of the paper is largely devoted to the statement and proof of Theorem 1. The proof is accomplished by computing the moments of the probability distribution of the n-particle potential function and finding the limits of these moments. A convergence test then establishes that these moments uniquely determine a limiting distribution. The application of the theorem to the statistical mechanical problem mentioned above is presented as Theorem 2.

In Section 3 several examples are discussed. The first is a simple quadratic interaction for points on a circle, for which the rate of convergence to the limiting distribution is explored. Next the logarithmic pairwise potential, as appears, for example, in point vortices, line charges, and guiding center plasmas, is shown to satisfy the hypotheses of the theorem (in the periodic case). This system is particularly interesting because the limiting distribution has features which do not appear in the thermodynamic limit.[3] A variation of the first example which is relevant to the distribution of eigenvalues of random matrices is also mentioned. The last example is a problem of integral geometry to which the theorem applies.

2. LIMIT THEOREMS

The main result of this paper is the following theorem.

Theorem 1. Let M be a manifold and g a probability measure on M. Suppose $f: M \times M \to \mathbf{R}$ is a function with the following properties:

(a) $f(x, y) = f(y, x)$; (b) $0 = \int_M f(x, y)\, dg(x)$ for all $y \in M$; (c) $\sup_{y \in M} (\int_M f^2(x, y)\, dg(x)) < \infty$; (d) every integral of the following form exists:

$$\int_{M^n} f(x_{i_1}, x_{j_1}) \cdot \ldots \cdot f(x_{i_r}, x_{j_r})\, dg(x_1) \cdots dg(x_n) \tag{2.1}$$

Let $x_1, \ldots, x_n$ be independent, identically distributed random variables with distribution g on M, and let U_n denote the random variable

$$\frac{1}{n} \sum_{1 \leqslant i < j \leqslant n} f(x_i, x_j) \tag{2.2}$$

Then, as $n \to \infty$, the sequence U_n converges in distribution to a random variable U with moments

$$\mu_m = m! \sum [(I_{n_1}/2n_1)^{e_1} \cdot \ldots \cdot (I_{n_r}/2n_r)^{e_r}]/[e_1! \cdot \ldots \cdot e_r!] \tag{2.3}$$

where

$$I_k = \int f(x_1, x_2) \cdot \ldots \cdot f(x_{k-1}, x_k)\, f(x_k, x_1)\, dg(x_1) \cdot \ldots \cdot dg(x_k)$$

and the sum is taken over all collections of positive integers $(n_1, \ldots, n_r)$, $(e_1, \ldots, e_r)$ which satisfy $1 < n_1 < \cdots < n_r \leqslant m$ and $n_1 e_1 + \cdots + n_r e_r = m$.

Remarks.

Assumption (a) indicates that the "interaction" f depends only on unordered pairs of points; (b) is required to prevent the means of the variables U_n from diverging; (d) ensures that the moments of the U_n exist; and (c) is a condition which suffices to establish that the moments μ_m determine a distribution. The last two conditions are satisfied by all bounded functions f; however, they are weak enough that many interaction functions for which the energy of a collision is infinite will also satisfy them. An example is given in the next section.

At first glance, one might not expect U_n to converge to a distribution, since it is essentially $(n-1)/2$ times the average of a number of values of f. Condition (b) of the hypothesis, however, requires that f have zero mean in a homogeneous way. If the evaluations of f were at independent points, the limiting distribution would be normal by the central limit theorem. Since the summands are not independent, the limiting distribution is in general not normal (and not even symmetric), and is determined by the interplay of f with the distribution of the points (x_i, x_j) in M^2.

The random variable U_n is an example of a U-statistic of order two.[4,5] U-statistics have been widely studied by statisticians, but the explicit formulas for the moments of the limiting distribution are new.[5] Furthermore, the proof technique can be applied to random variables which are not U-statistics (e.g., the vortex lattice systems of Section 3).

Proof of Theorem 1. The moments of the random variable U_n are the expected values

$$\mu_m(n) = \langle U_n^m \rangle = n^{-m} \left\langle \left[\sum_{i<j} f(x_i, x_j) \right]^m \right\rangle \tag{2.4}$$

A proof (combinatorial in nature) of the relation $\mu_m(n) = \mu_m + O(1/n)$ is presented in Appendix A. It remains to show that these limiting moments μ_m determine a distribution. A sufficient condition[6] is that the series $\sum \mu_{2m} t^m/(2m)!$ converges for all t in some nonempty interval. We will establish that, for some constant K, the following inequality is satisfied:

$$\mu_{2m}/(2m)! \leqslant (16K^2)^m \tag{2.5}$$

We begin by finding bounds for the integrals I_k. It is convenient to abbreviate $f(x_i, x_j)$ as f_{ij}, $dg(x_i)$ as dg_i, and to adopt the convention that when the variables of integration in an integral are not specified, all variables present in an expression are integrated. Similarly, the domain of integration will be suppressed.

Let $K^2 = \sup_y \int |f(x, y)|^2 \, dg(x)$. It is immediate that $|I_2| \leqslant \int |\int f_{12}^2 \, dg_1| \, dg_2 \leqslant K^2$. Using the Schwartz inequality,

$$|I_3| \leqslant \int \left| \int f_{12} f_{23} \, dg_2 \right| \cdot |f_{13}| \, dg_1 \, dg_3 \leqslant K^2 \int |f_{13}| \, dg_1 \, dg_3 \tag{2.6}$$

Writing f_{13} as $f_{13} \cdot 1$, the Schwartz inequality gives $\int |f_{13}| \, dg_1 \leqslant K$, and hence $|I_3| \leqslant K^3$.

The integral I_k is bounded in a similar way for higher k:

$$\begin{aligned} |I_k| &\leqslant K^2 \int |f_{34} \cdot \ldots \cdot f_{n1}| \\ &\leqslant K^4 \int |f_{56} \cdot \ldots \cdot f_{n1}| \leqslant \cdots \leqslant K^k \end{aligned} \tag{2.7}$$

(When k is odd, a factor of 1 must be inserted at the last step.) This shows that the summand $I_{n_1}^{e_1} \cdot \ldots \cdot I_{n_r}^{e_r}$ which appears in the formula for μ_m is bounded by $K^{n_1 e_1 + \cdots + n_r e_r} = K^m$. Hence $\mu_{2m}/(2m)!$ is bounded by K^{2m} times the number of terms in the summation for μ_{2m}. Each term corresponds to a distinct partition of the integer $2m$ into positive parts. Using a standard recursive formula[7] for the number of partitions, it is easy to show that this number is bounded by 4^{2m}. This establishes the desired inequality and proves that the limit of the distributions U_n is a distribution. This completes the proof of Theorem 1.

In applying the theorem to statistical mechanics, we interpret $f(x_i, x_j)$ to be the energy of the interaction between particles at x_i and x_j, and U_n to be the energy per particle.

Suppose now that the phase space M has finite volume. If $\phi(E)$ denotes the volume of states having energy no greater than E, then the function $\Omega(E) = \phi'(E)/\phi(\infty)$, the statistical weight,[2] or density of states at energy E, coincides with the probability density of the random variable $\sum_{i<j} f(x_i, x_j)$, when the particle positions are independent, uniformly distributed random variables.

Theorem 2. Let M be a compact manifold, g the uniform probability measure on M, and suppose M, g, and f satisfy hypotheses (a)–(d) of Theorem 1. If the potential energy of the n-particle system with state space M is given by $\sum_{i<j} f(x_i, x_j)$, then in the limit of large n, the energy-density statistical weight function $\Omega_n(E/n)$ of the microcanonical ensemble equals the probability density with moments μ_m.

3. APPLICATIONS

The result of Section 2 shows that the distribution of U_n is essentially independent of n for large n. As a simple example, let M be the unit interval with ends identified (circle). Figure 1 shows the graph of a quadratic inter-

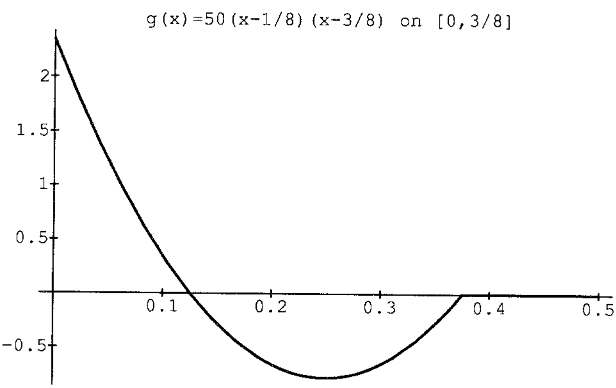

Fig. 1. Quadratic potential function, $f(x, y) = g(d(x, y))$; d = distance along the circle (i.e., arc length).

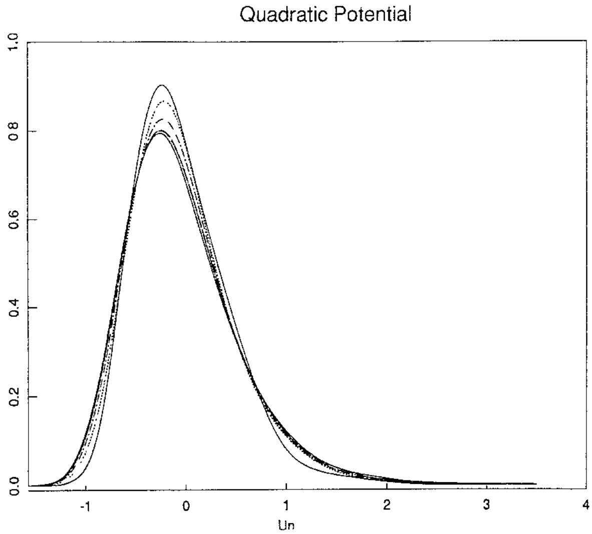

Fig. 2. Density estimates for the quadratic potential for $n = 3$, 5, 10, 20, 50, and 200. The height of the peak of the graph decreases with increasing n. The graphs for $n = 20$ and $n = 50$ are indistinguishable on this plot.

action function (resembling a Lennard-Jones potential) which depends only on the distance (arc length) between points in M. Figure 2 shows the estimated distribution of U_n for several values of n. The curves were generated by a nonparametric density estimation technique based on splines.[8] Notice the rapid rate of convergence in n of the distribution curves. We also give a table of estimates $\hat{\mu}_m$ of the moments of the distributions for various n derived from the simulations (Table I). The estimates are based on a sample size of 200,000 points for each n. By grouping the data into 20 groups of 10,000, estimates of the variance have been obtained. As the table shows, good estimates of the moments higher than the fourth require a much larger sample size. This can be very time consuming for large n, as

Table I. Moment Estimates for the Quadratic Potential[a]

n	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$	$\hat{\mu}_5$	$\hat{\mu}_6$
5	0.222 ± 0.004	0.138 ± 0.007	0.28 ± 0.02	0.48 ± 0.06	1.2 ± 0.2
10	0.246 ± 0.006	0.159 ± 0.015	0.35 ± 0.04	0.62 ± 0.12	1.5 ± 0.4
20	0.260 ± 0.006	0.172 ± 0.014	0.39 ± 0.05	0.8 ± 0.2	2.1 ± 1.0
50	0.268 ± 0.005	0.180 ± 0.013	0.42 ± 0.04	0.7 ± 0.1	1.9 ± 0.4
200	0.273 ± 0.004	0.185 ± 0.010	0.42 ± 0.04	0.8 ± 0.1	2.15 ± 0.6

[a] Moment estimates ± 1 standard deviation.

the computational effort of the simulation is $O(n^2)$. In contrast, the moment formula in Theorem 1 require the evaluation of a limited number of integrals for each moment, which in some cases can be evaluated exactly. [An example is $f(x, y) = \cos 2\pi(x - y)$, for which all integrals are powers of two.]

The conditions (c) and (d) of the preceding theorem are obviously met by bounded interaction functions f, but many systems of interest are governed by potentials which are unbounded for collisions. A logarithmic singularity is an example which occurs in many physical systems; we concentrate here on the periodic case.[9] If point vortices of unit strength are arranged in a pattern in the plane periodic in two directions, the configuration space becomes that of n points in a torus (plane modulo lattice) and the energy density of the configuration is proportional to $E = (1/2) \log n + (1/n) \sum_{i<j} f(x_i, x_j)$, with $f(x, y) = C \log|x - y| +$ (an even analytic function of $x - y$).[10] Clearly, Theorem 1 can be used to approximate the distribution of E for large n if the hypotheses (a)–(d) can be justified. Condition (a) is immediate, and (b) is established by a simple symmetry argument (see Appendix B). Because M is a torus, every integral $\int f^2(x, y)\, dg(x)$ in (c) has the same value. Finally, (d) can be shown to hold using the fact that the function $r|\log r|^\alpha$ has a removable singularity at $r = 0$ for positive α. Thus, a limiting distribution exists for $E - (1/2) \log n$, which can be used to approximate the distribution of energy in large but finite systems. This of course is not the same as the thermodynamic limit, which because of the $\log n$ term depends only on the values of the limiting distribution near $-\infty$. As a result, the two limits are qualitatively different. In particular, all finite systems at energies above $(\text{const} + \frac{1}{2} \log n)$ are in a negative temperature state, while no such states exist in the thermodynamic limit.[3]

There are other systems related to this vortex lattice system which also satisfy the conditions of Theorem 2. If the vortices are given either positive or negative unit circulations $\Gamma_1, \ldots, \Gamma_n$, then the interaction becomes $\Gamma_i \Gamma_j f_{ij}$. Interestingly, the limiting distribution exists and is independent of the choice of circulations, since the moments are computed in terms of integrals over cycles, which involve only the circulations squared. If the logarithm in the interaction function is replace by the modified Bessel function K_0 (which still has a logarithmic singularity at the origin), a system of flux lattices in a type II superconductor is modeled.[9]

Nonperiodic configurations of point vortices must be constrained in some way in order to obtain a configuration space of finite volume before Theorem 1 can be applied. For example, if the vortices are taken to lie on a circle, one obtains a system that has been investigated in conjunction with the distribution of eigenvalues of matrices with random entries.[11]

Finally, consider the following geometric construction. Let $x_1,..., x_n$ be n points on the sphere M of radius r in R^3. Connect these points with great circle arcs to form a complete graph on the sphere. Let p_i denote the product of the lengths of the arcs extending from x_i, and g the geometric mean of the p_i:

$$g = (\prod p_i)^{1/n} = \left[\prod_{i \neq j} d(x_i, x_j) \right]^{1/n} \tag{3.1}$$

If the positions of the vertices are independent uniformly distributed random variables, then the moments of the distribution of log g can be found for large n by applying the theorem when the radius r takes that value for which the mean of log g is zero.

APPENDIX A

Let α denote a collection of integers $i_1,..., i_m, j_1,..., j_m$ taken from the set $S_n = \{1,..., n\}$ with $i_k < j_k$, and

$$f_\alpha = f_{i_1, j_1} \cdot ... \cdot f_{i_m j_m} = f(x_{i_1}, x_{j_1}) \cdot ... \cdot f(x_{i_m}, x_{j_m})$$

We compute the moments

$$\mu_m(n) = \left\langle \left(\frac{1}{n} \sum_{1 \leqslant i < j \leqslant n} f_{ij} \right)^m \right\rangle = n^{-m} \sum_\alpha \langle f_\alpha \rangle \tag{A1}$$

By property (b) of the hypothesis of the theorem, $\langle f_\alpha \rangle = 0$ if α contains any integer value only once. Moreover, if J' is the collection of those α in which every integer value that appears, appears more than once and at least one value appears more than twice, then $n^{-m} \sum_{\alpha \in J'} \langle f_\alpha \rangle$ is $O(1/n)$: all the integrals $\langle f_\alpha \rangle$ are bounded by hypothesis (d), and the cardinality of J' is the number of ways of distributing fewer than m values from S_n among the $2m$ integers which constitute α, a quantity which is $O(n^{m-1})$.

Consider now the sum $n^{-m} \sum_{\alpha \in J} \langle f_\alpha \rangle$, with J the collection of α in which each integer value that appears, appears exactly twice; the argument above shows that $\mu_m(n) = n^{-m} \sum_{\alpha \in J} \langle f_\alpha \rangle + O(1/n)$. Each f_α can be identified with an edge-valued graph with m ordered nodes and two edges per node, that is, a collection of cycles with orders totalling m. The edges represent the integer values in α, and the nodes represent the factors $f(x_{i_k}, x_{j_k})$. (See Fig. 3.) Note that if f_α and f_β have the same graph but different integers assigned to the edges, then $\langle f_\alpha \rangle = \langle f_\beta \rangle$. The integral depends only on the graph identified with it, in particular, only on the

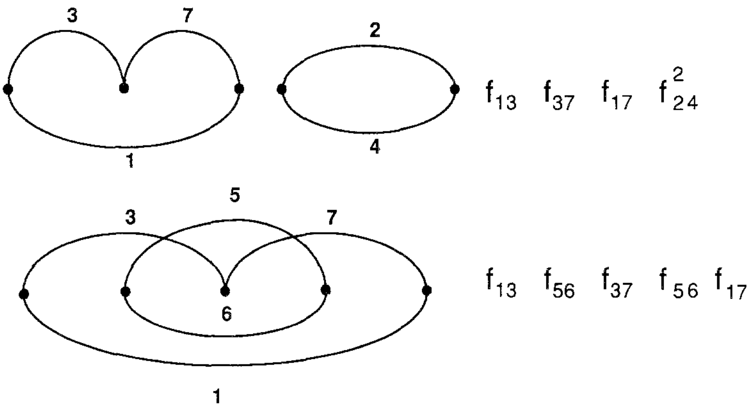

Fig. 3. The association of f_α with a graph having m ordered vertices and integer-valued edges. Note that $\langle f_\alpha \rangle = I_2 I_3$ in both instances.

number of cycles of various orders. Thus, $\langle f_\alpha \rangle$ can be expressed in terms of the cyclic integrals

$$I_k = \int f_{12} f_{23} \cdot \dots \cdot f_{k-1,k} f_{k1} \tag{A2}$$

Specifically,

$$\langle f_\alpha \rangle = I_{n_1}^{e_1} \cdot \dots \cdot I_{n_r}^{e_r} \tag{A3}$$

for some integers n_k, e_k satisfying $n_1 e_1 + \cdots + n_r e_r = m$. Consequently,

$$n^{-m} \sum_{\alpha \in J} \langle f_\alpha \rangle = \sum C_m(n, e)\, I_{n_1}^{e_1} \cdot \dots \cdot I_{n_r}^{e_r} \tag{A4}$$

The sum is taken over all collections of positive integers such that $1 < n_1 < \cdots < n_r \leqslant m$ and $n_1 e_1 + \cdots + n_r e_r = m$. The coefficient $C_m(n, e)$ is equal to n^{-m} times the number of $\alpha \in J$ having the graph consisting of exactly e_k copies of the cycle of order n_k, $1 \leqslant k \leqslant r$. We now compute this coefficient.

Given a graph, one obtains an associated α by choosing m integers from S_n, assigning these integers to the edges of the various cycles of the graphs, and then assigning the nodes to the m factors of f_α. This can be done in $N = n!\, m!/(n-m)!$ different ways, but not all resulting in distinct α. (See Fig. 4.) For each cycle of order $k > 2$, N overcounts by a factor of $2k$, the number of rotations and reflections of the cycle. Furthermore, the e cycles of order k are indistinguishable, so that N overcounts by a factor

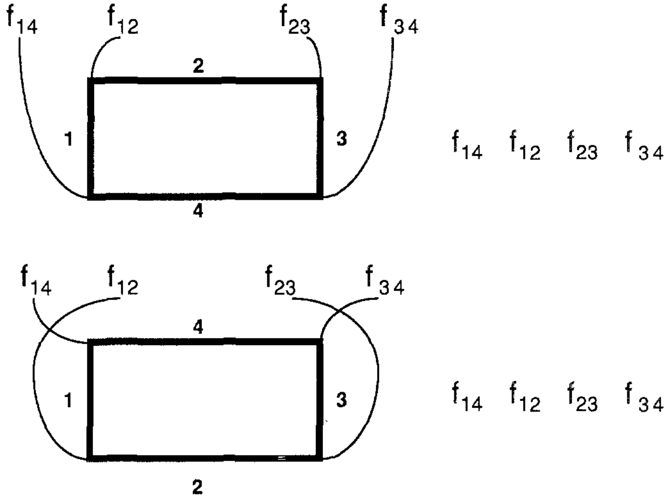

Fig. 4. Two ways of assigning integers to edges and nodes to factors which yield the same f_α.

of $e!$. Thus, N must be divided by $(2k)^e e!$. Similar considerations when $k=2$ result in the same factor.

We conclude that

$$\begin{aligned} C_m(n, e) &= \frac{n!\, m!}{n^m (n-m)!} \left[(2n_1)^{e_1} e_1! \cdot \ldots \cdot (2n_r)^{e_r} e_r!\right]^{-1} \\ &= m! \left[(2n_1)^{e_1} e_1! \cdot \ldots \cdot (2n_r)^{e_r} e_r!\right]^{-1} + O(1/n) \end{aligned} \tag{A5}$$

Thus, $\mu_m(n) = \mu_m + O(1/n)$.

APPENDIX B

This Appendix establishes property (b) for the vortex lattice potential of Section 3.

Given a configuration of vortex lattices with positions $z_1, \ldots, z_n$ and circulations $\Gamma_1, \ldots, \Gamma_n$, the expression for the total kinetic energy in a rotating reference frame[10] can be divided by n and rearranged to give the following formula for the energy per vortex:

$$\frac{E}{n} = \frac{1}{n} \sum_{1 \leqslant i < j \leqslant n} \Gamma_i \Gamma_j f_{ij} + \bar{\Gamma}^2 \left(E_1 - \frac{\ln \varepsilon}{4\pi}\right) \tag{B1}$$

with

$$f_{ij} = f(z_i, z_j) = -\frac{1}{4\pi}\,[\ln|\sigma_0(z_i - z_j)|^2 - \pi\rho\,|z_i - z_j|^2] + 2E_1 \qquad \text{(B2)}$$

The cutoff radius ε is referenced to a unit cell of unit area. An expression for energy density is obtained by replacing ε by $\varepsilon n^{-1/2}$ and suppressing the ε term,

$$\frac{E}{n} = \frac{1}{n}\sum \Gamma_i \Gamma_j f_{ij} + \bar{\Gamma}^2\left(E_1 + \frac{\ln n}{8\pi}\right) \qquad \text{(B3)}$$

Henceforth we assume all circulations are unity.

The energy density so computed has the same value for all arrangements of vortices into sublattices L/n containing n^2 vortices per unit cell, $n \geqslant 1$. Thus, $n^{-2}\sum f_{ij} = -(\ln n^2)/8\pi$. But by the symmetry of the configuration, $n^{-2}\sum f_{ij} = (1/2)\sum_{k \neq 1} f_{1k}$. This yields the identity $\sum f_{1k} = -(\ln n^2)/4\pi$, which can be used to determine the (improper) integral of f:

$$\int_M f(x, y)\,dx = \int_M f(x, 0)\,dx = \lim_{n \to \infty} n^{-2}\sum f_{1k} = 0 \qquad \text{(B4)}$$

ACKNOWLEDGMENTS

This work had its beginning in joint work of Kevin O'Neil and Laurence J. Campbell on negative temperature states of vortex systems.[12] Figure 2 was produced with the help of software written by Kevin Gehringer.

REFERENCES

1. J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (Wiley, New York, 1954), Chapter 3.
2. S. F. Edwards and J. B. Taylor, Negative temperature states of two-dimensional plasmas and vortex fluids, *Proc. R. Soc. Lond. A* **336**:257–271 (1974).
3. J. Frohlich and D. Ruelle, Statistical mechanics of vortices in an inviscid two-dimensional fluid, *Commun. Math. Phys.* **87**:1–36 (1982).
4. H. Rubin and R. A. Vitale, Asymptotic distribution of symmetric statistics, *Ann. Stat.* **8**:165–170 (1980).
5. Robert J. Serfling, *Approximation Theorems of Mathematical Statistics* (Wiley, New York, 1980), Chapter 5.
6. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. II, 3rd ed. (Wiley, New York, 1968), p. 228.
7. A. P. Hillman, G. L. Alexanderson, and R. M. Grassl, *Discrete and Combinatorial Mathematics* (Dellen, San Francisco, 1987).

8. K. Gehringer, Nonparametric probability density estimation using normalized B-splines, Masters thesis, Department of Mathematical and Computer Sciences, University of Tulsa (1990).
9. L. J. Campbell, Vortex lattices in theory and practice, in *Mathematical Aspects of Vortex Dynamics*, R. Caflisch, ed. (SIAM, Philadelphia, 1989), pp. 195–204.
10. K. O'Neil, On the Hamiltonian dynamics of vortex lattices, *J. Math. Phys.* **30**:1373–1379 (1989).
11. F. J. Dyson, Statistical theory of the energy levels of complex systems. I., *J. Math. Phys.* **3**(1):140–156 (1962).
12. L. J. Campbell and K. A. O'Neil, Statistics of 2D point vortices and high Reynolds number fluids, preprint.