Simple heuristic derivation of the universal jump in the superfluid density of two-dimensional superfluids

Chia-Ren Hu
Department of Physics, Center for Theoretical Physics, Texas A&M University, College Station, Texas 77843-4242 and Texas Center for Superconductivity, University of Houston, Houston Science Center, Houston, Texas 77204-5932

(Received 30 January 1992)

A simple heuristic derivation is given for the universal jump in the superfluid density ρ_s at the transition temperature T_c of a two-dimensional superfluid. It is based on the mathematical equivalence of the Hamiltonians of two systems: (i) a superfluid sandwiched between two parallel boundaries that are boosted adiabatically from rest to a finite velocity, and (ii) a two-dimensional two-component Coulomb gas (considered as a set of parallel line charges in three dimensions), sandwiched between two oppositely charged parallel capacitor plates.

The universal jump in the superfluid density ρ_s at the transition temperature T_c of a two-dimensional superfluid such as a ^{4}He film is a most remarkable prediction made in an extension by Nelson and Kosterlitz[1] of the general Kosterlitz-Thouless theory[2–4] of phase transitions in two-dimensional systems with an Abelian continuous symmetry. The Kosterlitz-Thouless theory is based on the notion that vortex-antivortex-pair excitations and their thermally induced unbinding play a unique role in such systems. Since vortices interact with each other via a long-range logarithmic interaction, a system of thermally created vortices and antivortices can be mapped into a two-dimensional two-component classical Coulomb gas. This mapping allows Kosterlitz and Thouless[2] (see also Kosterlitz[3] and Young[4]) to apply renormalization-group theory to predict a "topological" "vortex-unbinding" phase transition in any two-dimensional systems with an Abelian continuous symmetry. Nelson and Kosterlitz[1] invoked a formula

$$K_R^{-1} \equiv M^2 k_B T/\hbar^2 \rho_s(T) = (M^2/\hbar^2) \int d^2r \langle \mathbf{v}_s(\mathbf{r}) \cdot \mathbf{v}_s(0) \rangle \tag{1}$$

(with M the mass of the superfluid particles and k_B the Boltzmann constant) to relate the superfluid density ρ_s of a superfluid to the velocity-velocity correlation function of the superfluid, and evaluated the latter function within the renormalization-group theory of Kosterlitz and Thouless, to arrive at the conclusion

$$\lim_{T \to T_c^-} K_R^{-1} = \lim_{T \to T_c^-} \frac{M^2 k_B T}{\hbar^2 \rho_s(T)} = \frac{\pi}{2} , \tag{2}$$

which is the now-well-known prediction of a universal jump of ρ_s at T_c for a two-dimensional superfluid. Minnhagen and Warren[5] subsequently noted that a more correct starting point for the derivation of Eq. (2) is the formula

$$\rho_s = (1/T) \int d^2r [\langle \mathbf{g}_{s\parallel}(\mathbf{r}) \cdot \mathbf{g}_{s\parallel}(0) \rangle - \langle \mathbf{g}_{s\perp}(\mathbf{r}) \cdot \mathbf{g}_{s\perp}(0) \rangle] , \tag{3}$$

which they reduced to

$$\frac{\rho_s}{\rho_s^0} = \lim_{\mathbf{k} \to 0} \left[1 - \frac{2\pi \rho_s^0 (\hbar/M)^2}{T} V(\mathbf{k}) \langle n(\mathbf{k}) n(-\mathbf{k}) \rangle \right] \tag{4}$$

for a superfluid containing thermally excited vortices which give rise to a transverse contribution to the superfluid current density $\mathbf{g}_s$. In Eq. (4), ρ_s^0 is the "bare" superfluid density not yet renormalized by the thermally created vortices,[6] $V(\mathbf{k})$ is the Fourier-transformed vortex-vortex interaction, and $\langle n(\mathbf{k}) n(-\mathbf{k}) \rangle$ is the Fourier-transformed density-density correlation function of the vortices. After reaffirming the equivalence of the vortex gas to a Coulomb gas they then proved that the right-hand side of Eq. (4) is just equal to the inverse of the $\mathbf{k} \to 0$ limit of the $\mathbf{k}$-dependent dielectric constant of the Coulomb gas:

$$\lim_{\mathbf{k} \to 0} \epsilon(\mathbf{k}) . \tag{5}$$

Identifying this quantity with the $r \to \infty$ limit of the r-dependent dielectric constant $\epsilon_{\mathrm{KT}}(r)$ introduced in the Kosterlitz-Thouless theory, and noticing that that theory has already predicted the relation[2–4]

$$\epsilon_\infty |_{T=T_c} \equiv \lim_{r \to \infty} \epsilon_{\mathrm{KT}}(r) |_{T=T_c} = \frac{\pi}{2} \frac{\rho_s^0(T_c)(\hbar/M)^2}{k_B T_c} , \tag{6}$$

for a two-dimensional superfluid, it follows that Eq. (4) implies Eq. (2).

The proof given by Nelson and Kosterlitz of this remarkable prediction is sufficiently mathematical that it may be difficult for experimentalists and nonexperts to fully understand it. The alternative proof by Minnhagen and Warren reveals that the proof can actually be broken conceptually into two stages: The first stage is to establish that

$$\rho_s / \rho_s^0 = 1/\epsilon_\infty , \tag{7}$$

which is actually true at all temperatures, and the second stage is to invoke the prediction of the Kosterlitz-Thouless theory at $T = T_c$, Eq. (6), in order to arrive at the final result, Eq. (2). The derivation of Eq. (7) by

Minnhagen and Warren is not completely trivial: (i) It begins with Eq. (3), which itself requires a nontrivial proof.[7] The physics contained in that equation is also not transparent. (ii) It invokes the relatively obscure density-density correlation function for the vortices. (iii) It is done in momentum space which is less intuitive than real space. This last point may be the weakest part of the proof by Minnhagen and Warren, since they introduced a *local* dielectric constant in momentum space which is necessarily *nonlocal* in real space, whereas in the original theory of Kosterlitz and Thouless, the dielectric constant is defined to be local in real space and therefore it must be nonlocal in momentum space. In this paper a simple heuristic proof of Eq. (7) is given which bypasses these "shortcomings." Although the present proof is not as rigorous as those of Nelson *et al.* and Minnhagen *et al.*, it is intuitively clear, and therefore should promote a wider understanding of the remarkable prediction of the universal jump in ρ_s at T_c. Note that the present proof differs from that of Minnhagen and Warren at the first stage only; for the second stage one must still invoke the result of the renormalization-group analysis of Kosterlitz, Thouless, and Young, Eq. (6).

The present proof is based on a mathematical equivalence between two physical systems. They are as follows.

System A. A two-dimensional superfluid sandwiched between two parallel boundary lines which are both adiabatically boosted to a velocity v_0 along themselves (relative to a laboratory frame in which the system is initially at equilibrium).

System B. A classical, two-dimensional, two-component, neutral Coulomb gas, considered as a set of parallel line charges of both polarities in a three-dimensional space, sandwiched between two parallel capacitor plates charged to areal charge densities $\pm\sigma$.

More precisely, let the two-dimensional superfluid and the two-dimensional Coulomb gas both be confined in the xy plane (i.e., the equivalent three-dimensional line charges are all parallel to the z axis). The two boundary lines confining the superfluid are assumed to both extend from $y=-\infty$ to $+\infty$, and to be located at $x=-L_x/2$ and $+L_x/2$, respectively. These boundary lines are assumed to be adiabatically boosted to a velocity v_0 in the $-y$ direction. On the other hand, the capacitor plates confining the line charges are assumed to be both parallel to the yz plane, extending to infinity in all directions along these plates, with one plate located at $x=-L_x/2$ charged with a positive areal charge density σ, and the other plate located at $x=+L_x/2$ charged with a negative areal charge density $-\sigma$. We assume that L_x is a macroscopically large length.

For the two-dimensional Coulomb gas confined in the space between the two capacitor plates, the Hamiltonian (per unit length along z) is

$$\mathcal{H}_{\rm CG}^{\mathcal{E}_0}=-\sum_{i\neq j=1}^{2N}\tau_i\tau_j\ln(\rho_{i,j}/a)+\tilde{\mu}\sum_{i=1}^{2N}\tau_i^2-\sum_{i=1}^{2N}\tau_i\mathcal{E}_0x_i\ , \qquad (8)$$

where $\mathcal{E}_0\equiv4\pi\sigma$, $\tau_i=\pm\tau$ is the linear charge density of the ith line charge located at (x_i,y_i) satisfying $\sum_{i=1}^{2N}\tau_i=0$, $\rho_{i,j}\equiv[(x_i-x_j)^2+(y_i-y_j)^2]^{1/2}$ is the direct separation between the ith line charge and the jth one, and $\tilde{\mu}\tau_i^2=\tilde{\mu}\tau^2$ is the creation energy of one such line charge (per unit length along z). This Hamiltonian indicates that the line charges interact not only with each other (the first term), but also with the charges on the capacitor plates via the "external field" $\mathcal{E}_0$ (the last term).

For the two-dimensional superfluid confined between the two moving boundary lines, the Hamiltonian *in the frame of the walls* (per unit thickness if it is a ^{4}He film) is the total kinetic energy of the superfluid plus a term representing the effect of the moving walls. This is a Legendre transformation term common in thermodynamics and statistical mechanics:

$$\mathcal{H}_{sf}^{v_0}=\frac{1}{2\rho_s^0}\int\mathbf{g}_s^2d^2r-v_0(-\hat{\mathbf{e}}_y)\cdot\int\mathbf{g}_sd^2r\ , \qquad (9)$$

where we have not yet included the energy cost associated with the creation of the vortex cores, but this energy will be included in a later stage.[8]

We can immediately write down the following two relations, one for each system (with $\langle\ \rangle$ denoting ensemble average):

$$\langle\mathcal{H}_{\rm CG}^{\mathcal{E}_0}\rangle-\langle\mathcal{H}_{\rm CG}^{\mathcal{E}_0=0}\rangle=\frac{L_xL_y}{8\pi}\left[\frac{1}{\epsilon_\infty}-1\right]\mathcal{E}_0^2 \qquad (10)$$

and

$$\langle\mathcal{H}_{sf}^{v_0}\rangle-\langle\mathcal{H}_{sf}^{v_0=0}\rangle=\frac{L_xL_y}{2}(\rho_s-\rho_s^0)v_0^2\ , \qquad (11)$$

where L_y is the width of the capacitor plates in the y direction in the first formula, and it is the length of the boundary lines of the superfluid in the second formula (i.e., both systems are assumed to obey periodic boundary condition between $y=-L_y/2$ and $y=+L_y/2$).

Equation (10) follows simply from our knowledge of classical electrodynamics: The second term on the right-hand side is there because the Hamiltonian Eq. (8) does not include the electrostatic energy in vacuum when $\mathcal{E}_0$ is not zero. Otherwise the right-hand side is simply the electrostatic energy of a capacitor filled with a dielectric material (which arises from the line dipoles formed by the line charges below T_c due to the Kosterlitz-Thouless transition). The dielectric constant ϵ_∞ appears in the denominator rather than in the numerator because the quantity $\mathcal{E}_0$ should be identified as the displacement vector D rather than the electric field E because it is fixed to the areal charge density σ independent of ϵ_∞. We have used ϵ_∞ instead of, say, $\epsilon_{\rm KT}(r)$, because L_x is macroscopically large so line dipoles of practically all sizes should contribute in this dielectric constant. (To understand this statement fully one must understand the renormalization-group theory of the Kosterlitz-Thouless transition. The easiest place to do so seems to be the paper by Young.[4]) Note that the right-hand side of Eq. (10) is negative since a dielectric substance prefers to enter into a capacitor, so the total energy of the latter must go down as the dielectric substance is introduced inside the capacitor, if only the total charge Q of the capacitor is maintained constant so that the system is closed, as is the case here.

In order to see why Eq. (11) is true, we first rewrite Eq. (9) in the form

$$\mathcal{H}_{sf}^{v_0}=\frac{1}{2\rho_s^0}\int[\mathbf{g}_s-\rho_s^0 v_0(-\hat{\mathbf{e}}_y)]^2 d^2r-\tfrac{1}{2}\rho_s^0 v_0^2\int d^2r \ . \tag{12}$$

The second term of this expression gives immediately the second term on the right-hand side of Eq. (11). The first term in this expression is just the kinetic energy of the superfluid *in the moving frame of the boundary lines.* It should give the first term on the right-hand side of Eq. (11), when the expectation value difference on the left-hand side of Eq. (11) is evaluated. This is because the boundary lines are assumed to be *adiabatically* boosted to the velocity $v_0(-\hat{\mathbf{e}}_y)$, which means that any genuine superfluid component should remain stationary in the laboratory frame, i.e., it must be moving with a velocity $v_0\hat{\mathbf{e}}_y$ in the frame of the moving boundary lines. The thermally created vortices, however, will be able to reach mechanical equilibrium with the boundary lines, effectively acting as part of the normal fluid, which must appear stationary in the frame of the moving boundary lines at thermodynamic equilibrium. This explains why the superfluid density must be renormalized from the "bare value" ρ_s^0 to the yet undetermined "renormalized value" $\rho_s<\rho_s^0$. This accounts for the first term in the right-hand side of Eq. (11).

From another point of view, Eq. (10) may be taken as simply the equation *defining* the macroscopic dielectric constant ϵ_∞, and Eq. (11) may be taken as the equation *defining* the renormalized superfluid density ρ_s. Thus these equations are *by definition* true at all temperatures and whatever other parameter values characterizing the system. (This view also reveals that ρ_s is defined in the frame in which the vortices are statistically at rest.)

We now prove that the two physical systems A and B are mathematically equivalent. That is, we shall establish a one-to-one correspondence chart which maps the Hamiltonian of the first system (i.e., the superfluid in moving boundary lines) to that of the second system (i.e., the Coulomb gas in between the capacitor plates). This proof is only a slight extension of the equivalence proof by Minnhagen and Warren[5] of the same two systems but at $v_0=\mathcal{E}_0=0$, i.e., in their work the whole superfluid system including the thermally created vortices is stationary in the laboratory frame, and the Coulomb gas is subject to no external electric field. As in Ref. 5, we begin by writing

$$\mathbf{g}_s=\mathbf{g}_{s\parallel}+\mathbf{g}_{s\perp} \ , \tag{13}$$

where

$$\mathbf{g}_{s\parallel}=\rho_s^0(\hbar/M)\nabla\theta' \tag{14}$$

is the longitudinal part of $\mathbf{g}$, with $\theta'(\mathbf{x})$ the single-valued, nonsingular part of the phase of the superfluid wave function, which is subject to periodic boundary conditions between $x=\pm L_x/2$ and between $y=\pm L_y/2$ (which are all defined in the laboratory frame),[9] and

$$\mathbf{g}_{s\perp}=\rho_s^0\left[\frac{\hbar}{M}\right]\sum_{i=1}^{2N}m_i f(|\mathbf{x}-\mathbf{x}_i|)\hat{\mathbf{e}}_z\times\nabla\ln\frac{|\mathbf{x}-\mathbf{x}_i|}{a} \tag{15}$$

is the transverse contribution to $\mathbf{g}$ due to the thermally excited vortices, with $m_i=\pm1$ denoting the polarity of the ith vortex, and the function $f(|\mathbf{x}-\mathbf{x}_i|)$ rising from 0 to 1 smoothly as its argument increases from 0 to a few times the coherence length of the superfluid, representing the effect of the core of the ith vortex. We choose to use a discrete representation of the distribution of the vortices, rather than the continuous representation used by Minnhagen and Warren [cf. Eq. (2.10) of Ref. 5], because we do not plan to Fourier transform the current and vortex distribution functions, and our Coulomb gas Hamiltonian, Eq. (8), is also given in such a discrete representation. [We find that the core function f in Eq. (15) is necessary in order to not obtain a divergent self-interaction of the vortices. Such an interaction does not appear in the continuous representation invoked by Minnhagen and Warren. Thus they did not have to introduce this function.]

Substituting Eqs. (13)–(15) into Eq. (9), it is straightforward to obtain

$$\mathcal{H}_{sf}^{v_0}=\tfrac{1}{2}\rho_s^0\left[\frac{\hbar}{M}\right]^2\left[\int|\nabla\theta'|^2d^2x-2\pi\sum_{i\neq j=1}^{2N}m_i m_j\ln\frac{|\mathbf{x}-\mathbf{x}_i|}{a}\right]+\mu\sum_{i=1}^{2N}m_i^2-2\pi\rho_s^0\left[\frac{\hbar}{M}\right]v_0\sum_{i=1}^{2N}m_i x_i \ , \tag{16}$$

where we have now explicitly included the core contribution which is the μ-dependent third term, with μ the core energy of a single vortex.

The first three terms in this equation are not new,[1,5] except that in those earlier derivations a continuous vortex distribution was used. Our addition is the last term, which represents how the vortices are coupled to the moving boundary lines. This term is easier to derive than the second term, so everything we have done so far is relatively easy to follow or reproduce, but we have already done enough to draw our main conclusion.

From Eq. (16) we can see that the longitudinal part of the superflow, which is proportional to $\nabla\theta'$, is not coupled to the moving walls. This is easy to understand since it is well known that a genuine superfluid component cannot be dragged along by moving walls. Thus the first term in this equation may be dropped as far as the evaluation of the left-hand side of Eq. (11) is concerned. The remaining terms in this Hamiltonian may be compared with the Coulomb gas Hamiltonian, Eq. (8). It is then immediately clear that the two Hamiltonians are mathematically equivalent if only one makes the following identifications:

$$\tau_i\leftrightarrow(\pi\rho_s^0)^{1/2}(\hbar/M)m_i \ , \tag{17}$$

$$\tilde{\mu}\tau_i^2\leftrightarrow\mu m_i^2=\mu \ , \tag{18}$$

$$\mathcal{E}_0\leftrightarrow(4\pi\rho_s^0)^{1/2}v_0 \ . \tag{19}$$

This equivalence between the two Hamiltonians in Eqs. (8) and (16) (with the first term dropped in the latter), allows us to equate the right-hand sides of Eqs. (10) and (11), subject to the equivalence chart, Eqs. (17)–(19), and obtain directly Eq. (7). We have thus achieved what we have set out to prove, which, when combined with the prediction of the renormalization-group analysis, Eq. (6), establishes the main goal of this paper, i.e., proving the universal jump of ρ_s at T_c as is given in Eq. (2), which is often written more clearly as

$$\frac{\rho_s(T_c)}{T_c}=\frac{2}{\pi}\left[\frac{M}{\hbar}\right]^2 k_B \ . \tag{20}$$

Before we conclude this paper, we would like to offer another indication of the equivalence of the two systems A and B described above, and represented by the Hamiltonians in Eqs. (8) and (9).

It is well known that as T is increased across T_c, the superfluid system should lose its superfluidity and become a normal fluid, due to the fact that the originally bound vortex-antivortex pairs become unbound free vortices and antivortices, which can then cause the decay of any supercurrent by moving across it in the respectively appropriate directions. More precisely, for the system A, the superfluid component initially has a velocity $v_0\hat{\mathbf{e}}_y$ in the moving frame of the boundary lines which corresponds to a total phase increase of $(M/\hbar)v_0L_y$ in a length L_y along y. Since each positive (negative) vortex moving all the way from $x=-L_x/2$ to $+L_x/2$ (from $x=+L_x/2$ to $-L_x/2$) can cause this phase difference to decrease by exactly 2π, it clearly takes

$$N\equiv[(M/\hbar)v_0L_y]/2\pi L_y=Mv_0/h \tag{21}$$

positive (negative) vortices per unit length along y to move in the $+x$ $(-x)$ direction completely across the supercurrent in order to cause the complete decay of the supercurrent (in the moving frame). Using the correspondence chart, Eqs. (17)–(19), this number may be translated into

$$N=\frac{\mathcal{E}_0}{4\pi\tau}=\sigma/\tau \tag{22}$$

for the system B. This number now has a simple meaning for this second system: As T is increased across T_c, the Coulomb gas changes from behaving as a dielectric to behaving as a conductor, due to the unbinding of the line dipoles formed below T_c out of the positive and negative line charges. Thus at $T>T_c$ the free line charges in the Coulomb gas in between the capacitor plates should completely neutralize the charges on the capacitor plates by moving a certain number of positive (negative) line charges in the $+x$ $(-x)$ direction across the capacitor, per unit width in the y direction. This number is precisely given by N of Eq. (22), since σ is the areal charge of the capacitor, and τ is the linear charge density of each line charge. Their ratio is therefore the number of (oppositely charged) line charges needed to neutralize the surface charges of the capacitor plates. Thus we see another clear indication of the complete equivalence of the systems A and B under the equivalence chart Eqs. (17)–(19).

In summary, we have demonstrated that there is a complete equivalence between the systems A and B defined above under the correspondence chart given in Eqs. (17)–(19), and from this equivalence one can obtain a very simple heuristic derivation of the universal jump in the superfluid density of a two-dimensional superfluid at T_c. In addition, this work also shows how one can effectively apply a pseudoelectric field to the (two-dimensional) pseudoelectric charges played by the vortices in a two-dimensional (neutral) superfluid—an interesting concept which might have additional applications and generalizations (such as to finite wave number and frequency).

The author would like to thank his colleagues, G. Agnolet and W. M. Saslow for very useful comments. This work is partially supported by a grant from the Texas Center for Superconductivity at the University of Houston funded under the prime Grant No. MDA-872-88-G-0002 from the Defence Advanced Research Project Agency and the state of Texas.

[1]D. R. Nelson and J. M. Kosterlitz, Phys. Rev. Lett. **39**, 1201 (1977).

[2]J. M. Kosterlitz and D. J. Thouless, J. Phys. C **6**, 1181 (1973).

[3]J. M. Kosterlitz, J. Phys. C **7**, 1046 (1974).

[4]A. P. Young, Phys. Rev. B **19**, 1855 (1979); J. Phys. C **11**, L453 (1978).

[5]P. Minnhagen and G. G. Warren, Phys. Rev. B **24**, 2526 (1981).

[6]In Ref. 1 ρ_s^0 is written as ρ_0 and in Ref. 5 it is written as ρ. Both of those notations could be mistaken for the total particle density of the system, which would be wrong. In a review article [in *Phase Transitions and Critical Phenomena,* edited by C. Domb and J. L. Lebowitz (Academic, London, 1983), Vol. 9] D. R. Nelson explicitly used the notation ρ_s^0, which clearly indicated that $\rho_s^0(T)\equiv\rho-\rho_n^0(T)$ is already a temperature-dependent quantity, and that it has already taken into account all such contributions to the "bare normal-fluid density" $\rho_n^0(T)$ as the phonons, the rotons, the ^{3}He or any other kinds of impurities, the effect of the substrate and its surface roughness, etc., the only contribution to the normal fluid density not included in ρ_n^0 is the vortices which can be thermally created in the "bare superfluid background." It is then clear why Kosterlitz and Nelson call their prediction of the discontinuity in ρ_s *universal,* since they showed that this *very nonuniversal* ρ_s^0 is *canceled* in the final result of their analysis, Eq. (2) or (21).

[7]P. C. Hohenberg and P. C. Martin, Ann. Phys. (N.Y.) **34**, 291 (1965). See also other references given in Ref. 6 of Ref. 5 above.

[8]Note that the capacitor plates of system A and the boundary lines of system B considered so far are still somewhat idealized, since they are assumed to only produce the last terms appearing in Eqs. (8) and (9), respectively. Actual capacitor plates (which are made of conductors) and boundary lines (which are solid walls to the superfluid) can also produce im-

age charges and vortices, respectively, outside the regions defined by the plates and lines. Fortunately, the laws governing both kinds of images are sufficiently similar that including the effects of these image objects does not spoil the present proof of equivalence of the two systems at all. But for the simplicity of the proof we shall not explicitly include them.

[9]This periodic boundary condition in the laboratory frame may be replaced by the vanishing of θ' of Eq. (14) or its normal derivative at walls which must be not moving perpendicular to themselves in the laboratory frame. Any of these boundary conditions then ensures that the genuine superfluid component of the system is stationary in the laboratory frame, as is required in our definition of the system A. Minnhagen and Warren in Ref. 5 above assumed the boundary condition that θ' is a constant at the walls. We think that that boundary condition is too restrictive.