The Berezinskii-Kosterlitz-Thouless transition

In the last lecture we saw that true long-range order is impossible in 2D and a fortiori in 1D at any finite temperature for a system where the order parameter is a complex scalar object[1]; the reason is simply that long-wavelength phase fluctuations destroy the phase correlation at long distances.

However, this does not prevent 1D and 2D systems behaving, under some circumstances, very much "as if" they possess long-range order. In the following, I will illustrate this point with the specific example of a neutral superfluid (e.g. ^{4}He), where the consequences of (pseudo-) LRO are particularly dramatic, but qualitatively similar considerations apply to other systems described by a complex scalar order parameter, such as the 2D XY model of magnetism ("2D" in the sense that the spins are constrained to lie in a plane), or crystalline order.

In the standard theory of bulk (3D) superfluidity in a neutral system such as ^{4}He, one defines the order parameter $\Psi(\mathbf{r},t)$ as follows: Quite generally, for an arbitrary many-body system, one can define the one-body density matrix $\rho(\mathbf{r},\mathbf{r}',t)$ and find a basis $\chi_i(\mathbf{r},t)$ of single-particle states (which are not necessary energy eigenstates) in which it is diagonal, so that in this basis

$$\rho(\mathbf{r},\mathbf{r}',t)=\sum_i n_i(t)\chi_i(\mathbf{r},t)\chi_i^*(\mathbf{r}',t) \tag{1}$$

The system is said to be Bose condensed if one and only one of the eigenvalues ("occupation numbers") $n_i(t)$ is "macroscopic", i.e. of order of the total number of particles N, while the rest are all "microscopic" (of order 1). If now we denote the single macroscopic eigenvalue $N_0(t)$ and the corresponding eigenfunction $\chi_0(\mathbf{r},t)$, the order parameter is simply defined by the prescription

$$\Psi(\mathbf{r},t)\equiv\sqrt{N_0(t)}\chi_0(\mathbf{r},t) \tag{2}$$

Contrary to an impression given in some of the literature, this definition (which nowhere invokes the idea of "broken $U(1)$ symmetry" or anything of that kind) is perfectly satisfactory for all purposes for which the concept of order parameter has been (legitimately) used.

The superfluid velocity $\mathbf{v}_s(\mathbf{r},t)$ is defined in terms of the order parameter as follows: we write

$$\Psi(\mathbf{r},t)\equiv A(\mathbf{r},t)\ \exp\ i\phi(\mathbf{r},t)\quad (A,\phi\ \text{both real}) \tag{3}$$

and define[2]

$$\mathbf{v}_s(\mathbf{r},t)\equiv(\hbar/m)\nabla\phi(\mathbf{r},t) \tag{4}$$

[1] or more generally whenever the symmetry which is broken by the existence of the order parameter is continuous (and the order parameter is not "pinned", e.g. by random static fields in the magnetic case).

[2] We note that if at some point $\chi_0(\mathbf{r},t)$ and hence $\Psi(\mathbf{r},t)$ is zero, the phase $\phi(\mathbf{r},t)$ and hence the superfluid velocity $\mathbf{v}_s(\mathbf{r},t)$ is undefined.

From this definition follows at once the irrotationality condition

$$\text{curl } \mathbf{v}_s(\mathbf{r},t) \equiv 0 \tag{5}$$

and the Onsager-Feynman quantization condition

$$\oint \mathbf{v}_s \cdot d\mathbf{l} = nh/m \quad (n = 0, \pm 1...) \tag{6}$$

provided the closed circuit in the integral is taken entirely through regions where $\Psi(\mathbf{r},t)$ is nonzero. It is clear in particular that if we take a bulk (3D) annular geometry and assume that $\Psi(\mathbf{r})$ is indeed nonzero everywhere in the annulus, then the number n (the "winding number") which appears in the Onsager-Feynman relation is a *topological invariant*: it is the number of turns in the Argand plane made by the phase of Ψ (or equivalently of χ_0 as we go once around the annulus, and this number cannot change so long as $|\Psi|$ remains everywhere finite ("hula-hoop" analogy). Consequently, provided the system is sufficiently strongly stabilized against fluctuations of $|\Psi|$, a state in which $n \neq 0$ can be highly metastable.

In the standard theory of (bulk) superfluidity, one considers a general configuration of the order parameter $\Psi(\mathbf{r})$ (hence of the superfluid velocity $\mathbf{v}_s(\mathbf{r})$ and assumes that thermal equilibrium has been achieved *subject* to the topological constraints implied by the Onsager-Feynman quantization condition. For the moment we will assume that the equilibrium is with walls which are at rest in the lab frame. Under these conditions it is still in general possible to have a finite mass current $\mathbf{j}(\mathbf{r})$ flowing in the system and we define the *superfluid density* $\rho_s(T)$ by the relation[3]

$$\mathbf{j}(\mathbf{r}) = \rho_s(T)\mathbf{v}_s(\mathbf{r}). \tag{7}$$

In the case of liquid ^{4}He, $\rho_s(T)$ can be measured by various experimental techniques; it is found to tend to zero as $T \to T_c$ and to unity as $T \to 0$ (something which is theoretically expected but surprisingly difficult to prove rigorously, see e.g. AJL, J. Stat. Phys. **93**, 927 (1998)). Along with the current, a finite value of $\mathbf{v}_s(\mathbf{r})$ produces a "flow" (kinetic) term in the total energy, which under the stated conditions is

$$E_{flow} = \frac{1}{2}\int \rho_s \mathbf{v}_s^2(\mathbf{r}) d\mathbf{r} \tag{8}$$

It is useful at this stage to make contact with the Ginzburg-Landau formalism used in lectures 8 and 9. We recall that in that formalism the expression for the "bending energy" of the order parameter $\Psi(\mathbf{r})$ was

$$\mathcal{F}_{bend} = \gamma(T)|\nabla\Psi(\mathbf{r})|^2 \tag{9}$$

If we consider a situation where the amplitude of the order parameter is constant at its equilibrium value $\Psi_0(T)$ while the phase $\phi(\mathbf{r})$ is spatially varying eqn. (9) becomes

$$\mathcal{F}_{flow} = \gamma(T)|\Psi_0(T)|^2|\nabla\phi(\mathbf{r})|^2 \tag{10}$$

[3]Strictly speaking, we should write $\mathbf{j}_s(\mathbf{r}) = \int \rho_s(\mathbf{r},\mathbf{r}')\mathbf{v}_s(\mathbf{r}')d\mathbf{r}'$, but the range of the function $\rho_s(\mathbf{r},\mathbf{r}') = \rho_s(\mathbf{r}-\mathbf{r}')$ is so short that it can usually be safely approximated by $\rho_s\delta(\mathbf{r}-\mathbf{r}')$.

Comparing (10) with eqn. (8) and using (4), we find

$$\rho_s(T) = \frac{2m^2}{\hbar^2}\gamma(T)|\Psi_0(T)|^2 \tag{11}$$

(For a superconductor, $\mathbf{v}_s$ is defined as $(\hbar/2m)\nabla\phi(\mathbf{r})$, so we must make the same replacement ($m \to 2m$) in (11)). From now on it will be generally more convenient to formulate the theory in terms of $\rho_s(T)$.

These considerations can be generalized to the case where the "normal component" (the non-superfluid part of the liquid) is not in equilibrium with stationary walls but rather in internal equilibrium and moving with a mean velocity $\mathbf{v}_n(\mathbf{r})$. Then the total current is

$$\mathbf{j}(\mathbf{r}) = \rho_s(T)\mathbf{v}_s(\mathbf{r}) + \rho_n(T)\mathbf{v}_n(\mathbf{r}) \tag{12}$$

and the kinetic energy is

$$E_{flow} = \int \frac{1}{2}(\rho_s\mathbf{v}_s^2(\mathbf{r}) + \rho_n\mathbf{v}_n^2(\mathbf{r}))d\mathbf{r} \tag{13}$$

An alternative and possibly more useful way of looking at it is that viewed from the frame in which the normal component is locally at rest, the (local) KE density due to the superfluid has the form

$$E_{flow} = \frac{1}{2}\rho_s(\mathbf{v}_s - \mathbf{v}_n)^2 \tag{14}$$

It is useful to note two kinds of topological singularities which can occur in a bulk (3D) superfluid: vortex *lines* and vortex *rings*. A (single)[4] vortex line is a configuration in which there exists a line (in the simplest case straight, so that we can take its locus to be $x = y = 0$) such that the order parameter $\Psi(\mathbf{r})$ vanishes on the line and its phase is just the phase of $x \pm iy$, so that it rotates once in the Argand plane as we go around the line in the physical xy-plane. Thus, according to the Feynman-Onsager relation, the "winding number" with respect to the line $x = y = 0$ is ± 1. Formally,

$$\Psi(\mathbf{r}) = f(r_\perp)(x \pm iy)/r_\perp, \quad r_\perp \equiv \sqrt{x^2 + y^2} \tag{15}$$

where $f(r_\perp)$ is a function which is constant for sufficiently large $r_\perp$ but falls off to zero for $r_\perp \to 0$. We will define the (order of magnitude of the) distance at which $f(r_\perp)$ recovers its uniform value as r_0 (the "core radius"); generally speaking it follows from a more microscopic consideration that r_0 is of the order of an atomic spacing, but close to T_c it is expected to diverge with the correlation (healing) length $\xi(T)$. The energy per unit length of such a vortex is (apart from a small "core" term) just the energy of the uniform system plus the flow energy $\frac{1}{2}\int \rho_s\mathbf{v}_s^2(\mathbf{r})d\mathbf{r}$: this is easily calculated, since $|\mathbf{v}_s| \equiv (\hbar/m)|\nabla\phi| = \hbar/mr$

$$\frac{E_{\text{v.l.}}}{L} = \frac{1}{2}\rho_s(T)(\hbar/m)^2\int_{r_0}^{R}\frac{2\pi r}{r^2} = \pi\rho_s(T)\frac{\hbar^2}{m^2}\ln(R/r_0)$$
$$\equiv (\rho_s(T)/4\pi)(h/m)^2\ln(R/r_0) \tag{16}$$

[4] Multiply quantized lines are unstable.

where R is an upper cutoff of the order of the cell dimensions, so that line energy of a vortex line *diverges* as $R \to \infty$. In the literature the "quantum of circulation" h/m is often denoted κ.

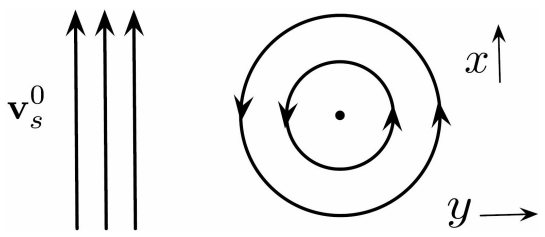
A second 3D configuration which is of interest is a vortex *ring*, which can be visualized approximately as a vortex line bent around into a circle; the radius R_0 of the circle can be arbitrary provided it is $\gg r_0$. It is clear that at large distances ($\gg R_0$) the contribution to $\mathbf{v}_s(\mathbf{r})$ from opposite sides of the ring will tend to cancel, leaving a term of order R_0/r^2 and associated KE $\sim r^{-4}$; thus the total flow energy associated with the ring must be finite. In fact, it is not difficult to see that the ring radius R effectively acts as an upper cutoff, replacing the cell dimension. Thus the total energy of the ring must on essentially dimensional grounds be of the form

$$E_{\text{v.r.}}(R_0) \sim \text{ const. } R_0 \ln(R_0/r_0) \times h^2/m^2 \tag{17}$$

where we do not need for present purposes to calculate the constant. For the 2D system such as ^{4}He films in which we are interested, both vortex lines and vortex rings reduce to a simple excitation, we shall call simply a vortex, in which Ψ is constant in the direction perpendicular to the film and as a function of the in-plane coordinate $\mathbf{r}_\perp$ is given by (15), i.e. it is just a cross-section of a 3D vortex line. Evidently the total energy of such a vortex is just given by the film thickness d multiplied by eqn. (16).

An interesting property of both vortex lines and vortex rings is their interaction with a possible uniform background flow of the superfluid. Suppose that this background flow is (in the lab frame) in the positive x-direction, and we consider a vortex line whose polarity is such that the flow is in the positive x-direction for positive y and vice versa. Then it is clear that the flow energy in the lab frame will be reduced if the vortex moves to the right. This effect is easily calculated[5] quantitatively, and we find for the appropriate energy per unit length

$$(E_{\mathbf{v}.1.}/L)(y) = -(\rho_s(T)(h/m)v_s^0)y \tag{18}$$



where v_s^0 is the magnitude of the background flow. Note that this energy does not depend explicitly on the cutoff. The derivative of this energy with respect to y is the famous *Magnus force*, which is usually written in vector notation in the form

$$\mathbf{F} = \rho_s(T)(\mathbf{v}_s^0 \times \boldsymbol{\kappa}) \tag{19}$$

where $\boldsymbol{\kappa} \equiv (h/m)\hat{\boldsymbol{z}}, \hat{\boldsymbol{z}}$ being a unit vector along the direction of the line, with the "corkscrew sense" of the rotation. These considerations of course apply equally to vortices in a 2D film. In the case of a vortex *ring*, it is intuitively obvious that (for the right polarity) the Magnus force arising from a given segment of the line tends to move

[5] E.g. by putting the system in a slab geometry infinite in the z-direction and using method of images to take account of the boundary conditions.

the segment outwards, so that the net tendency is to expand the ring. Again, it is not difficult to calculate effect quantitatively, with the simple result

$$E(v_s) = \pm\rho_s v_s \kappa \pi R^2 \tag{20}$$

Lets now turn to the application of these ideas to (quasi)-1D and 2D systems. In the following, we will assume that we can define, for the 3D "bulk" geometry, a standard GL free energy functional which is adequate to describe fluctuations on a length scale larger than the (extrapolated) zero-temperature correlation length ξ_0, which we take[6] to have its "default" value of $\sim$ an interatomic spacing a (*not* the temperature-dependent healing length $\xi(T) \sim (T - T_c)^{-1/2}$). We moreover define the bulk 3D superfluid density $\rho_s(T)$ as previously, and recall that it is related to the parameters of GL theory by

$$\rho_s(T) = 2(m/\hbar)^2 \gamma(T) |\nabla \Psi_0(T)|^2 \tag{21}$$

so that the expression for the energy due to superfluid flow (the phase-gradient term) is (cf. above)

$$E_{flow} = \frac{1}{2}\rho_s(T)(\hbar/m)^2 \int |\nabla \phi(\mathbf{r})|^2 d\mathbf{r} \tag{22}$$

We consider the small fluctuations of the phase around its (arbitrary) mean-field value ϕ_0, writing (independently of dimensionality)

$$\phi(\mathbf{r}) = \phi_0 + \Omega^{-1/2} \sum_{\mathbf{k}} \phi_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{23}$$

Then according to the considerations of lecture 9, we have for thermal expectation value

$$\langle \phi_{\mathbf{k}} \phi_{-\mathbf{k}} \rangle = k_{\mathrm{B}} T / [\rho_s(T)(\hbar/m)^2 \mathbf{k}^2] \sim 1/\mathbf{k}^2 \lambda_T^2 \tag{24}$$

(where in the last estimate we suppose $\rho_s \sim \rho : \lambda_T$ is the thermal de Broglie wavelength). The phase correlations in position space are given (up to a constant of order unity) by

$$\langle [\phi(\mathbf{r}) - \phi(0)]^2 \rangle = \frac{\Omega^{-1}}{\lambda_T} \sum_{\mathbf{k}} \frac{\sin^2 \frac{1}{2}\mathbf{k}\cdot\mathbf{r}}{k^2} \tag{25}$$

as previously obtained, where Ω is the (hyper) volume of the system. So far, everything is independent of the precise geometry of the system.

Before embarking on our main topic, the effect of fluctuations and vorticity in thin 2D films, let us first briefly look at some related quantities in a formally 1D (annular) geometry. We can distinguish two limiting cases: In the case when the thickness of the annulus is sufficiently large it follows from eqn. (25) that the effects of fluctuations are negligible (i.e. the system is effectively 3-dimensional) and the magnitude of the order parameter may be roughly approximated by its mean-field value $\Psi_0(T)$ [Problem]. Under

[6] This estimate is correct for ^{4}He. For Fermi superfluids (^{3}He and superconductors), ξ_0 is usually larger, and this needs to be taken into account in some formulae below.

these conditions the superfluid velocity $\mathbf{v}_s$ and thus the winding number $n = \kappa^{-1} \oint \mathbf{v}_s \cdot d\boldsymbol{l}$ are unambiguously defined, and it then follows from topological considerations that states with $|n| \neq 0$ are strongly *metastable*.[7] The exception is very close to T_c, where a non-zero decay rate is observable; this is believed to be due to the "Langer-Fisher" mechanism,[8], in which a vortex *ring* is created and expands in the superfluid flow.

In the present context a more interesting case is that in which the thickness is so small that the 1D correlation length is considerably less than the circumference of the annulus, so that there is no long-range order in the usual sense. (This would in practice require very stringent experimental conditions, *see* Problem). Despite this, however, it is intuitively clear that we can still define a *local* order parameter $\Psi(\mathbf{r})$ (though this would obviously require some generalization of eqn. (2)), and that under most circumstances the thermal expectation value of $|\Psi(\mathbf{r})|^2$ should not be very different from its value in the bulk 3D superfluid at the temperature in question. Moreover, while the *absolute* phase of $\Psi(\mathbf{r})$ is under these conditions strongly fluctuating, the *relative* phase at points differing by $|\mathbf{r} - \mathbf{r}'| \ll \xi_{1D}$, where ξ_{1D} is the 1D correlation length, should still be well-defined, and hence one can still define the superfluid velocity by the standard prescription

$$\mathbf{v}_s(\mathbf{r}) = \frac{\hbar}{m} \nabla \phi(\mathbf{r}) \tag{26}$$

Moreover, the requirement that the phase of the order parameter (condensate wave function) should return to itself mod. 2π when we circumnavigate the annulus still leads to the Onsager-Feynman quantization condition

$$\oint \mathbf{v}_s \cdot d\boldsymbol{l} = n\kappa \quad (\kappa \equiv h/m) \tag{27}$$

The crucial point, now, is that to change the "winding number" n would require, just as in a bulk 3D case possessing genuine long-range order, depression of the order parameter locally to zero. For an annulus sufficiently thin to be in this "1D" limit the relevant process is probably not the nucleation and growth of a vortex ring à la Langer-Fisher, but a simple process ("phase slip") in which $\Psi(\mathbf{r})$ goes uniformly to zero across a cross-section of the annulus; however, at low enough temperatures this process, like the Langer-Fisher one, will be exponentially suppressed by the Gibbs-Arrhenius factor. Hence under appropriate conditions *states of nonzero superflow can be metastable even in the absence of LRO*. The long-wavelength fluctuations which destroy the LRO are completely irrelevant to this argument ("slinky" analogy). However, one thing worth noting (with an eye to the 2D case) is that the free energy barrier which enters in Gibbs-Arrhenius factor tends to zero with the cross-sectional area A of the annulus, so that in the truly 1D limit ($A \to 0$), at any nonzero temperature, phase slips always become infinitely frequent and completely destroy metastable superflow.

[7] It should be strongly emphasized that this result holds only for a *complex scalar* order parameter, such as is realized in ^{4}He or the classic superconductors. For a more complicated order parameter the situation is more interesting, see Ho, PRL **49**, 1837 (1982).

[8] PRL **19**, 560 (1967).

We can now turn to the case of our primary interest, quasi-2D thin films. The considerations below apply to any 2D phase transition for which the order parameter is a complex scalar object, including the XY model, superconductivity, CDW's (in so far as the band structure is sufficiently close to 1D to allow those to exist) and even the formation of a crystalline lattice; however, for definiteness I shall explicitly discuss, as above, the case of neutral superfluid such as ^{4}He (a system which has in fact played a major role in experimental tests of the theory). It should be emphasized that (most of) the considerations below (like those above on the 1D ring) do *not* apply to an order parameter which has either a less complicated structure than a complex scalar (e.g. the Ising model of ferromagnetism) or a more complicated one (e.g. Heisenberg ferromagnetism). The case of superconductivity involves a special complication due to the long range of the Ampère force, which will be discussed in lecture 11; however, at the end of the day it turns out that for a sufficiently thin films it can be handled in a way exactly analogous to that of a neutral superfluid.

We first return briefly to our previous considerations concerning small phase fluctuations around the mean-field equilibrium. Recall once more that the general expression of the (complex scalar) order parameter $\Psi(\mathbf{r})$ arising from phase fluctuations is (remembering that $\rho_s = 2\gamma(T)|\Psi_0|^2$)

$$\langle \Psi(0)\Psi^*(\mathbf{r})\rangle = \exp - Q(r)$$
$$Q(r) \equiv \frac{2k_{\mathrm{B}}T}{(2\pi)^d \rho_s(\hbar/m)^2}\int_0^{k_c} d^dk \frac{\sin^2 \frac{1}{2}\mathbf{k}\cdot\mathbf{r}}{k^2} \tag{28}$$

where k_c is an upper cutoff, provisionally taken as $\sim \xi_0^{-1}$. A little thought shows that in 2D the integral has (for $k_c r \gg 1$) the value $\pi \ln k_c r -$ const. Hence, we find that the order parameter correlations decay according to a *power law*:

$$\langle \Psi(0)\Psi^*(\mathbf{r})\rangle \sim r^{-\eta(T)} \tag{29}$$

where the exponent $\eta(T)$ is given by

$$\eta(T) \equiv k_{\mathrm{B}}T/(2\pi\rho_s(T)(\hbar/m)^2) \tag{30}$$

We need not be too concerned about the prefactor of $r^{-\eta}$, which depends on exactly how the upper cutoff is treated.

Eqn. (30) has an interesting consequence. We recall that the "susceptibility" (in the case of superfluidity, this is the response to a fictitious field which breaks $U(1)$ symmetry) is proportional to the space integral of the correlator $\langle \Psi(0)\Psi^*(r)\rangle$. We see that in 2D this is infinite or finite according as $\eta(T)$ is less than or greater than 2. Hence, we expect a subtle change in the behavior of the system at the temperature such that

$$\rho_s(T) = \frac{1}{4\pi}\left[\frac{m}{\hbar}\right]^2 k_{\mathrm{B}}T \tag{31}$$

(note that here $\rho_s(T)$ is the (mean-field) superfluid density *per unit area*, so the dimensions are right). We shall see, however, that before this point is reached something quite different happens.

In fact, *small* fluctuations of the phase are by no means the whole story. As we have seen, in 2D, for a complex scalar order parameter (symmetry group $U(1)$) there also exist *topological* excitations, namely vortices. We already calculated above the energy per unit length of a vortex *line* in a 3D system, and it is immediately clear that we can generalize this result to a 2D system:

$$E_{\rm v} = \pi\rho_s^0(T)(\hbar/m)^2(\ln(R/r_0) - \text{const.}) \tag{32}$$

Here $\rho_s^0(T)$ is the (mean-field) superfluid (mass) density *per unit area*, r_0 is a short-distance cutoff corresponding roughly to the "size" of the vortex core (which we expect to be of the order of the GL healing length $\xi(T)$), R is a measure of the linear dimensions of the cell and the constant depends on the shape of the cell (and precise definition of R).

Consider now the $free$ energy associated with the formation of a vortex. This is the energy $E_{\rm v}$ minus $TS_{\rm v}$, where $S_{\rm v} \equiv k_{\rm B}\ln W_{\rm v}$ and $W_{\rm v}$ is the number of possible different ways of creating the vortex. It is clear that the order of magnitude of $W_{\rm v}$ (which does not need to be precisely defined for present purposes) is $R^2/r_0'^2$, where r_0' is of the order of r_0. Hence we have

$$S_{\rm v} = 2k_{\rm B}\ln(R/r_0) - \text{const.} \tag{33}$$

and the total free energy $F_{\rm v}$ is

$$F_{\rm v} \equiv E_{\rm v} - TS_{\rm v} = \left\{\frac{\pi\rho_s^0(T)}{(m/\hbar)^2} - 2k_{\rm B}T\right\}\ln(R/r_0) - \text{const.} \tag{34}$$

In the thermodynamic limit ($R/r_0 \to \infty$) the constant can be neglected, and we see that $F_{\rm v}$ *changes sign* at a characteristic temperature T_0 given by the formula[9]

$$k_{\rm B}T_0 = \frac{\pi}{2}\left[\frac{\hbar}{m}\right]^2\rho_s^0(T_0) \tag{35}$$

Note that T_0 is (trivially) *below* the mean-field transition temperature T_c where ρ_s^0 vanishes. However, it is clear that for a 2D system (film) of thickness $d \gg \lambda_T$ (where λ_T is the de Broglie wavelength at T_c) T_0 clearly approaches T_c. In fact, if we make an order-of-magnitude estimate of $\rho_s^0(T)$ by the GL-type formula $\rho_s^0(T) \sim \epsilon\rho_{\rm 2D}$ where $\epsilon \equiv 1 - T/T_c$, we find that the value ϵ_0 of ϵ corresponding to T_0 is of order

$$\epsilon_0 \sim \left(\frac{a}{\lambda_T}\right)^3\left(\frac{\lambda_T}{d}\right) \sim \frac{\lambda_T}{d} \tag{36}$$

(since typically $\lambda_T \sim a$, the inter-particle spacing). Hence for films more than a few atoms thick T_0 is very close to T_c, and to see the effects associated with the former clearly we need to use rather thin films.

[9]Note that the value of ρ_s defined by this equation is 8 times that defined by the disappearance of infinite susceptibility (see above) so the latter point is never reached.

What is the significance of T_0 (often denoted T_{KT} in the literature)? Within the naive "single-vortex" calculation given, we expect that for $T < T_0$ the density of vortices should be zero in the thermodynamic limit, while for $T > T_0$ there should a large number (proportional to some power of R) present. Intuitively, we should expect the presence of this large number of vortices not only to destroy any LRO in the technical sense that has survived the effect of small phase fluctuations, but more importantly to destroy the whole concept of a topologically constructed winding number and thus the phenomenon of persistent currents (superfluidity). However, the above naive calculation, while it actually turns out (subject to one provision, see below) to give the correct answer for the temperature T_0 of this "topological" transition, is not adequate quantitatively, inter alia because it neglects the effects of vortex-vortex interactions. The classic theory of the topological transition, which takes these interactions fully into account, was given by Kosterlitz and Thouless,[10] and in the rest of this lecture and the next I will describe it.

A first and essential step in the KT theory is to obtain the energy of interaction of a pair of vortices. This requires quite careful definition. If we simply add the flow *energies* corresponding to the two vortices separately, then as we have seen we get $2A\ln R/r_0$, where R is the cell dimension and r_0 the size of the vortex "core"; for our case the constant A is $\pi\rho_s(T)(\hbar/m)^2$. The energy of interaction between the vortices comes from the cross-terms in the kinetic energy of flow: if the cores of vortices 1 and 2 are located at $\mathbf{r}_1$ and $\mathbf{r}_2$ respectively then we have

$$E_{\rm int} = \rho_s^0(T)\int d^2\mathbf{r}\ \mathbf{v}_{s1}(\mathbf{r})\cdot\mathbf{v}_{s2}(\mathbf{r}) = \pm\rho_s^0(T)(\hbar/m)^2\int d^2\mathbf{r}\frac{(\hat{\boldsymbol{z}}\times(\mathbf{r}-\mathbf{r}_1))(\hat{\boldsymbol{z}}\times(\mathbf{r}-\mathbf{r}_2))}{|\mathbf{r}-\mathbf{r}_1|^2|\mathbf{r}-\mathbf{r}_2|^2} \tag{37}$$

where $\pm$ signs refer respectively to parallel and antiparallel orientation. After a little calculation using various trigonometric identities, we find that if we choose the origin so that the vortices are centered at $\pm\mathbf{r}_{12}/2$ (so that their distance apart is r_{12}) the integral reduces to the expression

$$2\pi\int_0^R \frac{r\ dr}{r^2+(r_{12}/2)^2} \approx 2\pi\ln(R/r_{12}) - \text{const.} \tag{38}$$

Thus, the mutual interaction energy is

$$E_{\rm int} = \pm 2A\ln(R/r_{12})(-\text{const}) \tag{39}$$

If we assume that the total "vorticity" in the sample (i.e. the number of $+n$ vortices minus the number of n ones) is zero, than the terms proportional to $\ln R$ will cancel, and if we define the "charge" on a single vortex to be $\pm q$ where $q \equiv \sqrt{A}$, then we write the total energy of a system of vortices in the form

$$E = \frac{1}{2}\sum_{ij} U(r_{ij}) \tag{40}$$

[10] J. Phys. C **6**, 1181 (1973).

where[11]

$$U(r_{ij}) = -2q_i q_j \ln(r_{ij}/r_0) + 2\mu \text{ for } r_{ij} > r_0, \text{ and } U(r_{ij}) = 0 \text{ for } rij < r_0 \tag{41}$$

which is the form usually used in the literature. The "chemical potential" term 2μ is necessary because it takes a finite energy to create a pair of antiparallel vortices at $r_{ij} = r_0$, because (inter alia) of the necessity to force the *amplitude* of the order parameter to zero (something not taken into account in the above calculation). Note that in our case the magnitude of the "charge" q on a single vortex is

$$|q| = (\hbar/m)\sqrt{\pi\rho_s^0(T)} \tag{42}$$

and so depends on temperature. It is important to note that in this formula $\rho_s^0(T)$ is the "mean-field" superfluid density (and thus vanishes when $T \to T_{c0}$, the mean-field transition temperature); as we shall see, this is not in fact a directly measurable quantity.

A very useful aid to one's intuition is that the form (41) of the interaction energy is exactly that of the so-called "2D Coulomb gas" (a 2D plasma of $+n$ and $-n$ electric charges interacting by the Coulomb potential, $e^2 \ln r_{ij}$, which would be obtained from the exchange of photons strictly confined to the plane); although this model has probably no simple physical realization, it is a favorite subject of study for many-body theorists. (Hence, of course, the name "charge" for the q_i's). We will exploit this analogy in the next lecture.

[11] Strictly speaking we should write $E = \frac{1}{2}\sum_{ij}(-2q_i q_j \ln r_{ij}/r_0) + \sum_i \mu$.