RICE UNIVERSITY

Theoretical Investigation of Possibilities for Generating Coherent Helium-4 Beams

by

Robert B. O'Connor, Jr.

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS

Thesis Director's Signature:

Houston, Texas
(May, 1972)
Theoretical Investigation of Possibilities for Generating Coherent Helium-4 Beams

Robert B. O'Connor, Jr.

Abstract

As part of an effort to find a suitable illuminant for the holographic microscopy of single biological molecules, a survey has been made of various possible methods for generating neutral He-4 beams. The required energies are in the range from $10^2$ to $10^3$ ev which correspond to particle velocities of about $10^5$ cm/sec. Coherence requirements for lensless holography set the maximum acceptable velocity bandwidth at about $10^2$ cm/sec. Continuous liquid beams, droplet beams, thermal effusion beams, supersonic nozzle beams, ionized beams, and excited beams have been considered. Of these the supersonic nozzle approach appears to be the most promising. However, filtering of the beam to achieve the necessary bandwidth appears to be a major obstacle. Droplet beams might also be possible if the liquid flux is not too great or if a suitable means for gasification can be found. Development of detection technology is required in order to produce a satisfactory recording of the scattered helium atoms. Fundamental questions still remain as to the ultimate suitability of helium as an illuminant for microscopy.

The original idea and starting point was that since both He-4 atoms and photons were Bose particles, it might be possible to construct the helium analog of an optical laser. After a review of laser theory and of the fundamental properties of He-4 below the $\lambda$-point, the principal problem appears to be that the energies at which many of the interesting superfluid phenomena exist are two to three orders of magnitude below the required beam energies. Accelerating superfluid droplets seems to be the most practical way of bridging the gap, but there are the problems mentioned above. One possible analog model still appears to merit further work.
# Table of Contents

## Chapter I: Introduction
- References

## Chapter II: Photon Beams
- 2.1 Photon Properties
- 2.2 Coherence
- 2.3 Imaging
- 2.4 Lasers
- References

## Chapter III: Properties of Helium-4
- 3.1 General
- 3.2 Ground State
- 3.3 Phonon Excitations
- 3.4 Roton Excitations
- 3.5 Two Fluid Model
- 3.6 Flowing Helium
- References

## Chapter IV: Helium-4 Beams
- 4.1 Imaging Requirements
- 4.2 Beam Types and Properties
- 4.3 Beam Filtering
- References

## Chapter V: Summary and Conclusions

## Acknowledgment

## Appendix A: Photon Beams
- Photon Properties
  - A2.1a Bose Properties of Photons
  - A2.1b Photon Statistics
  - A2.1c Fluctuations
- Coherence
  - A2.2a Classical Theory of Temporal Coherence
  - A2.2b Wiener-Khintchine Theorem
  - A2.2c Classical Theory of Spatial Coherence
- Imaging
  - A2.3a Helmholtz-Kirchoff Theorem
  - A2.3b Fresnel-Kirchoff Integral
  - A2.3c Fresnel Transformation
  - A2.3d Alternate Form for Fresnel Transformation
  - A2.3e Basic Principles of Lensless Holography
  - A2.3f Abbe Sine Condition
  - A2.3g Rayleigh Resolution Criterion
- Lasers
  - A2.4a Absorption and Emission Probability Rates
  - A2.4b Transition Probability Rates
  - A2.4c Amplification
  - A2.4d Cavities
  - A2.4e Threshold Condition
  - A2.4f Output Bandwidth
- References
## Appendix B: Properties of Helium-4

### Ground State
- B3.2a Bose-Einstein Distribution  
- B3.2b Chemical Potential  
- B3.2c Ground Orbital Occupancy  
- B3.2d Feynman Theory of the Ground State

### Phonon Excitations
- B3.3a Lattice Vibrations  
- B3.3b Thermodynamic Relations  
- B3.3c Feynman Theory of Phonon States

### Roton Excitations
- B3.4a Thermodynamic Relation for Rotons

### Two Fluid Model
- B3.5a Superfluidity  
- B3.5b Equation of Continuity  
- B3.5c Adiabatic Equation  
- B3.5d Euler's Equation  
- B3.5e Momentum Flux  
- B3.5f Navier-Stokes Equation  
- B3.5g Equations of Motion  
- B3.5h First and Second Sound Propagation

### Flowing Helium
- B3.6a Classical Rotational Flow  
- B3.6b Minimal Energy of Classical Irrotational Flow  
- B3.6c Helmholtz Vortex Theorem  
- B3.6d Classical Velocity Field  
- B3.6e Velocity Field of a Classical Ring Vortex  
- B3.6f Energy Dissipation in a Classical Fluid  
- B3.6g Classical Poiseuille Flow  
- B3.6h Irrotational Flow in Helium  
- B3.6i Rotational Flow in Helium  
- B3.6j Critical Velocity and Flow Resistance  
- B3.6k Rotons as Ring Vortices

### References
Chapter I

Introduction

Our ability to examine individual atoms by any means whatsoever is limited at best. Usually we deduce the properties of individual atoms and molecules by their aggregate response to a given set of macroscopic conditions. For example, much of what we know about the structure of complex molecules has been obtained from x-ray, neutron and electron diffraction studies of macroscopic crystals formed from such molecules. Occasionally, however, we are tantalized by crude images of very large molecules or very heavy ones; as for instance, the DNA strand of a bacteriophage,\(^1\) or a single uranium atom.\(^2\) These few examples suggest that with the advance of modern technology, it should be a relatively simple step to increase the resolution of our probes to the point that any atom in any molecule can be imaged and the structural configuration worked out from that molecule alone. This speculation might be valid except for one basic fact; that the energy required to "see" at the molecular or atomic scales is usually as great or greater than the energy that binds the molecules together. We can generally expect to destroy the molecule in the process of imaging it.

Recently Breedlove and Trammell\(^3\) in examining the fundamental limitations of microscopy concluded that with electrons, protons or photons as illuminants, it would not be possible to image individual biological molecules without inflicting severe radiation damage on them; so heavy, in fact, that the molecules at the end of the imaging process would almost certainly not be the same as at the start. Furthermore, any attempt to increase the resolving power would only make matters worse, since increased resolving power with a given illuminant usually
implies greater energy. The ratio of inelastic to elastic scattering increases with energy, contrary to some widely held earlier beliefs, so that the expected damage would be proportionately greater. Therefore, with the above illuminants there are very stringent restrictions on the ultimate feasible resolution.

Other illuminants discussed in that study were neutrons and helium atoms. In principle the former would make an ideal illuminant but because of their extremely small cross sections, enormous fluxes would be required; for example, on the order of $10^{20} - 10^{22}$ neutrons/cm$^2$-sec. would be required to produce an image in a reasonable length of time. Helium-4, on the other hand, suffers in the opposite direction as a potential illuminant. Its cross sections are extremely large due to its being an atom with an effective "hard sphere" diameter on the order of angstroms. Except for this limitation, however, it should be a possible illuminant. It is inert, spinless, and deBroglie wavelengths on the order of 1 Å are attained with energies of about .02 eV; less by a factor of almost $10^3$ than the binding energies of many molecules of interest.

The object of this thesis is to explore various possible approaches toward generating a suitable beam of helium-4 atoms for molecular imaging. An explicit assumption has been made that the beam should be coherent so that lensless holographic techniques could be applied. This is not a necessary condition, however. It should be possible to ionize or excite the post-scattered helium atoms and form images through conventional electrostatic or magnetic lenses.

The original idea and starting point was that since helium-4 atoms are Bose particles, it might be possible to construct the analog of a laser where the emitted beam, instead of being photons, would be atoms. The possibility of such a device has not been ruled out and further work
is required to establish or rule out its feasibility.

This thesis is organized basically into three parts; a survey of photon beams and their properties, a survey of the properties of helium-4 below the Λ-point, and an analysis of various methods of beam generation. Most of the detail and background material are relegated to corresponding appendices.

References


Chapter II

Photon Beams

Photons and helium-4 atoms are Bose particles. Because of the extensive knowledge which exists about photon beams, it is appropriate to review some of their properties in the hope of furthering our understanding of helium-4 beam generation, beam properties, and imaging systems.

In this chapter we discuss the basic Bose properties of photons and their statistical consequences. This leads into the concept of coherence and its implications in the formation of images. Next, the basic principles of holography are reviewed and the factors affecting resolution discussed. Finally, we analyze the formation of coherent photon beams by lasers with particular emphasis on the threshold condition and output bandwidth.

2.1 Photon Properties. Most macroscopic electromagnetic phenomena can be described satisfactorily in terms of Maxwell's equations. These lead to the wave equation and a wave description for the transmission of electromagnetic energy in some medium. Even at the microscopic level, the classical wave viewpoint leads to results which are consistent with experiment. However, it is a basic postulate in quantum mechanics that particles can be described in terms of waves and conversely. Therefore an electromagnetic wave has a particle description; not only is it a wave but it is also a set of traveling photons, the number and energy of which are related to the energy and frequency of the wave, and the angular momenta of which are related to the polarization.

The particle description can be developed using the macroscopic electromagnetic wave as the starting point. \(^*\)

\(^*\) Spin considerations are omitted since they are not relevant for this study.
One supposes that the wave is contained in a rectangular box and imposes the boundary condition that it vanishes at the walls. This leads to the specification of allowable standing waves or modes. Each mode can be thought of as the superposition of two traveling waves of common frequency but of opposite direction. Therefore, if we identify photon states with modes, we have an alternate description of an electromagnetic wave. The frequency of the mode determines the energy carried by a photon; the number of photons in the mode determines the energy carried in the corresponding frequency interval of the macroscopic wave, etc. The explicit wave-particle connection is given by the deBroglie relationship, $\lambda = \frac{h}{p}$; wavelength being a wave concept and momentum, a particle concept. For a small box the photon states are well separated. If the box is allowed to become larger, the photon states come closer together and eventually, if the box is carried far enough, approach a continuum.

If a photon is known to be in a precise state, it follows from the uncertainty principle that its location is completely unknown. Through the connection of state and mode, this is equivalent to saying that if a photon has definite momentum and energy, it can also be described as an infinite plane wave with a definite wave vector.

Photons, when viewed as particles, are bosons. This says two things: they are indistinguishable and they are symmetric under exchange. The first leads to the result that, as a consequence of some interaction, the probability for a photon to enter an $n$-occupied state is $n+1$ times larger than it is for it to enter an unoccupied state (A2.1a). A corollary is that the probability of knocking a photon out of an $n$-occupied state is $n$ times larger than knocking one out of a singly occupied state. It is clear from this that one should expect large fluctuations in the occupancy of photon states. The second property leads to the possibility of more than one photon to occupy a single state in the
first place. The first property could not exist without the second.

The fact that a photon is a Bose particle leads to significant consequences in the statistical distribution of states. If we call a single particle state an orbital, the orbital occupancy, instead of being given by the usual Boltzmann distribution,

\[ n(\epsilon) = C e^{-\beta \epsilon}; \quad C = \text{normalization constant}, \]

is given by Planck's distribution (A2.1b),

\[ n(\epsilon) = \frac{1}{e^{\beta \epsilon} - 1}. \]

For large energies the results are indistinguishable but for low energies the difference is profound.

Next, let us consider the fluctuations in orbital occupancy. From statistical mechanics we have

\[ \Delta \bar{n}^2 = \bar{n}^2 - \bar{n}^2 = -\frac{1}{\beta} \frac{\partial \bar{n}}{\partial \epsilon}. \]

Substituting Planck's distribution for \( \bar{n} \),

\[ \Delta \bar{n}^2 = -\frac{1}{\beta} \frac{\partial}{\partial \epsilon} [e^{\beta \epsilon} - 1] = e^{\beta \epsilon} [e^{\beta \epsilon} - 1]^{-2} \]

or

\[ \Delta \bar{n}^2 = \bar{n} (\bar{n} + 1). \]

In the limit as \( \bar{n} \) becomes larger, we obtain the relation for fluctuations in state occupancy for a Bose gas,

\[ \Delta \bar{n}^2 \propto \bar{n}^2 \]

or

\[ \Delta \bar{n}^2 / \bar{n}^2 \propto 1. \]

This relation can also be derived (A2.1c) from an electromagnetic wave viewpoint, which again indicates the equivalence of the wave and particle descriptions of electromagnetic radiation.

2.2 Coherence. Coherence is a concept that expresses the degree of phase correlation between space-time points
in a wave. There are two aspects of it that, while not independent, are most easily discussed separately; temporal coherence and spatial coherence.

Temporal coherence can be illustrated by means of a Michelson interferometer, figure 1. Light from source \( S \) is split into two beams by the half silvered mirror, reflected off the mirrors \( M_1 \) and \( M_2 \), and recombined along the path leading to the observation point \( P \). Depending upon the relative lengths of paths 1 and 2, interference fringes may or may not be seen at \( P \). If 1 and 2 are equal, maximum contrast fringes will be observed and the component beams are said to be completely coherent. If the lengths of 1 and 2 are sufficiently different, no interference pattern will be seen at \( P \) and the two components are said to be completely incoherent. Intermediate path differences lead to intermediate fringe visibilities and the component beams are said to be partially coherent.

Any real source has a finite bandwidth which has the effect of causing the phase between frequency components of a wave to change over some interval of time. We can see this quite easily if we let \( \nu \) be the effective minimum frequency of the band and \( \nu + \Delta \nu \) the maximum. If the two extreme components are in phase at one instant of time, they will be phase shifted with respect to each other at a later time. We may define coherence time \( \tau_c \) as the time required for the two phases to shift by \( 2\pi \) radians. That is,

\[
(V + \Delta V) \tau_c - V \tau_c = 1
\]

or

\[
\tau_c = \frac{1}{\Delta \nu},
\]
where $\tau_c$ is the coherence time.

Another way to view coherence time is illustrated in figure 2. Here we think of the source as radiating a set of overlapping wave trains, the superposition of which defines the signal. Each train has a finite time (and spatial) length. We may think of the mirror splitting each train into components which are then recombined with a phase shift determined by the path difference. Coherence time corresponds to the length of the individual trains. If the recombined trains overlap, there is at least partial coherence; if they do not, the resultant signal at $P$ is incoherent.

The two views of coherence are clearly related. It is well known and easily demonstrated by Fourier analysis that a time signal of finite length $T$ has an effective bandwidth of approximately $\Delta \nu \sim \frac{1}{T}$.

The essence of temporal coherence is that for time intervals less than the coherence time, there is a more or less well defined phase relationship between two points on the time trace from a single source. The phase relationship is better defined for intervals which are short compared to the coherence time and less well defined for relatively longer intervals. Under normal conditions we would expect there to be no temporal coherence between the wave trains of independently emitting sources.

We can relate the wave view of temporal coherence to the particle view through the Heisenberg uncertainty principle. Refer again to figure 2 and suppose that we are located at the observation point $P$. We should like to
determine which path the photon traveled, either 1 or 2, based upon its arrival time at \( P \). From the uncertainty principle we have

\[
\Delta E \cdot \Delta t \geq \hbar
\]

or

\[
h \Delta \nu \cdot \Delta t \geq \hbar
\]

The branch taken by the photon can be determined if \( \Delta t \geq \frac{1}{\Delta \nu} \) for a given \( \Delta \nu \) or bandwidth. But \( \tau_c = \frac{1}{\Delta \nu} \), so we see that the coherence time \( \tau_c \) is a measure of the maximum possible resolution in time for a particle having a specified energy resolution. Equivalently, it is the maximum time in which a wave can interfere with itself; i.e., have temporal coherence.

Spatial coherence is usually illustrated by means of a Young's double-slit configuration as shown in figure 3. It consists of an extended luminous source where \( S \) and \( S' \) are two extreme points. \( P_1 \) and \( P_2 \) are slits in an otherwise opaque mask and \( O \) is the observation point. We wish to compare the phases of the radiation at \( P_1 \) and \( P_2 \) as a function of the length of the source, the distance between source and mask, and the slit separation. Note that if only a single source point, say \( S \), were used, the experiment would be equivalent to the Michelson interferometer experiment discussed in connection with temporal coherence.

Consider first the single source point \( S \). The condition for the coherence of signals at \( P_1 \) and \( P_2 \) is that the path difference between \( P_1 S \) and \( P_2 S \) be less than \( \lambda \), where \( \lambda \) is the wavelength corresponding to the highest effective frequency in the emission band of \( S \). From the geometry and
symmetry of the source placement, any element of source \( S-S' \) will also satisfy this condition. Using the notation and geometry in the figure, we obtain for the spatial coherence condition
\[
I \cdot \frac{L}{R} \leq \lambda
\]
We define coherence length by the equality in this expression. That is,
\[
L_c \equiv \frac{\lambda R}{\ell}
\]
where \( L_c \) is the coherence length. Generalizing we may define coherence area by
\[
A_c \equiv \frac{\lambda^2 R^2}{A_s}
\]
where \( A_c \) is the coherence area, and \( A_s = l^2 \) is the source area.

The essence of spatial coherence is that for points separated by less than the spatial coherence length in the transverse plane, there is a more or less well defined phase relationship; the relationship being better defined for separation distances which are short compared to the coherence length, etc.

We may relate wave and particle viewpoints of spatial coherence through the uncertainty principle. Consider again Young's double-slit experiment, figure 3. Suppose we receive at point \( O \) a photon emitted from the extended source and suppose we should like to determine which slit the photon passed through. From the uncertainty principle,
\[
\Delta p_x \cdot l \geq \hbar,
\]
but
\[
\frac{\Delta p_x}{p_x} = \frac{L}{R} \quad \Rightarrow \quad \Delta p_x = \frac{L}{R} \cdot p_x,
\]
and
\[
p_x = \frac{h}{\lambda} \quad \Rightarrow \quad \Delta p_x = \frac{L}{R} \cdot \frac{h}{\lambda},
\]
so
\[
\frac{L}{R} \cdot \frac{h}{\lambda} = \Delta p_x \approx \frac{h}{\lambda},
\]
or
\[
L \approx \frac{\lambda R}{\ell}.
\]
Provided this last condition holds, we can specify the slit traversed by the photon. But $L_c = \frac{\lambda}{k}$ so that the coherence length $L_c$ is a measure of the limiting position resolution for a photon with specified momentum resolution. Equivalently, $L_c$ is the maximum spatial dimension over which a wave can interfere with itself; i.e., exhibit spatial coherence.

The theory of coherence can be developed much further than has been done above. The classical theory of temporal coherence is outlined in appendix A2.2a, the Wiener-Khintchine theorem is derived in appendix A2.2b, and the classical theory of spatial coherence is developed in appendix A2.2c. The important result of this theory is that quantitative measures of coherence are developed. For example, the measure of temporal coherence is the auto-correlation function

$$\gamma(\tau) = \frac{\langle E(t) E(t-\tau) \rangle}{\langle E(t)^2 \rangle}$$

and the measure of spatial coherence is the cross-correlation function

$$\gamma_{xz}(\tau) = \frac{\langle E_x(t) E_z(t-\tau) \rangle}{\sqrt{\langle E_x(t)^2 \rangle \langle E_z(t)^2 \rangle}}.$$

Each can be expressed in terms of the spectral properties of the signals through the Wiener-Khintchine theorem. In holographic imaging the degree of coherence is important because ultimately it determines whether or not fringes are formed on a hologram when a reference beam is made to interfere with an object beam.

2.3 Imaging. For the purposes of this investigation it is sufficient to consider only scalar diffraction theory in the analysis of image formation. The starting point for this theory is the scalar wave equation; that is, a single component of the more general vector wave equation derived from Maxwell's equations. Because our detectors have finite response times, we only measure
time-averaged values of the electromagnetic field, for which a scalar wave equation is sufficient. Furthermore, we are not concerned with polarization effects.

The basic theorem in scalar diffraction theory is the Helmholtz-Kirchhoff theorem (A2.3a) which states that the field at some point in a region can be expressed as the integral of a function of the field on the boundary of the region (excluding the observation point) and its normal derivatives. This theorem follows directly from the scalar wave equation. Assuming that the field drops off to zero at sufficiently large distances from the observation point, we can reduce the surface area over which the integration is carried out to the much smaller region that corresponds to the aperture of some optical system. The resulting integral (A2.3b) is known as the Fresnel-Kirchhoff integral and is given by

\[ E(\xi) = \frac{i}{\lambda} \iint_{S_0} E_{inc}(\xi') \frac{e^{ik|\xi' - \xi|}}{|\xi' - \xi|} \left[ \frac{\cos \theta + \cos \theta'}{2} \right] d^2\xi' \]

where \( \xi \) is the position of the observation point, \( \xi' \) is the variable position vector to the aperture surface, \( E_{inc}(\xi') \) is the incident field on the aperture surface, and \( \theta, \theta' \) are the angles between \( \xi \) and \( \xi' \) with the surface normal respectively. From its form and the factor \( \exp[\frac{ik|\xi' - \xi|}{|\xi' - \xi|}] \) in the integrand, it is evident that the Fresnel-Kirchhoff integral is essentially a mathematical statement of Huygen's principle.

In application this equation can frequently be reduced to (A2.3c)

\[ E(\xi) = \frac{i}{L} \int_{S} e^{-\frac{ik}{11} [(x-x_s)^2 + (y-y_s)^2]} E(x_s y_s z_s) \, dx_s dy_s \]

where \( S \) corresponds to the aperture plane and \( \xi \) corresponds to the observation point on some transverse plane. This equation is known as the Fresnel transformation and is
valid for optical systems whose rays make only small angles with the optical axis. It expresses how the field is transformed from one plane to the next and can be generalized to include the transformations due to lenses, masks, etc. An alternate frequently used form (A2.3d) is

$$E(z) = A \iint E(x_s y_s z_s) e^{ik(u x_s + v y_s)} dx_s dy_s ,$$

where $A$ is a constant and $u, v$ are direction cosines from the point $(x_s, y_s)$ in the aperture surface to the observation point. The Fresnel transformation is basic to the performance and resolution limitations of optical systems. It forms the underlying theoretical justification for holographic imaging which we next discuss.

Holography is also known as "wavefront reconstruction" and this expresses in a general way the basic idea. If one could capture on a photographic plate, say, the instantaneous amplitude and phase information contained in a light beam, then one would have the necessary information to reconstruct that beam at a later time and form an image that would be in every way equivalent to the original image. Since most detectors have long response times compared to the periods of ordinary light, the phase information is lost in the recording process unless special precautions are taken. Obviously if the phase information is lost, there is in general no way of reconstructing the beam. In holography, the phase information is retained by the ingenious technique of beating the main light beam with a reference beam and recording the resultant interference pattern. A necessary condition for holography is coherence between the main beam and reference beam so that the phase relationship at any given point between the two is essentially constant. It can be shown (A2.3e) that when a developed hologram is reilluminated with a new incident beam both a real image and a virtual image of the original object are formed. Depending upon how closely
the new beam approximates the original reference beam, the rays corresponding to the virtual image become the "reconstructed" beam.

Resolution limitations in optical systems depend upon the geometry of the optics, the nature of the source, the frequency and bandwidth of the illuminating beam, the properties of the detection system, and so forth. This is true not only for ordinary optical systems but also for those adapted to holography. We shall discuss only the most general resolution theory here which is that based on the Rayleigh resolution criterion as applied to microscopes. Whether or not the magnification and imaging of some object is carried out in a one step process as with an ordinary light microscope or in a multiple step process employing holography, say, the principles which determine resolution are the same. It is likely, however, that the fewer the number of steps used, the greater the ultimate resolution.

The Rayleigh resolution criterion (A2.3g) can be expressed as

\[ \alpha = 0.61 \frac{\lambda}{a_0} \]

where \( \alpha \) is the half angle subtended by the image at the final principal plane of the lens system, \( \lambda \) is the wavelength, and \( a_0 \) is the radius of the aperture. If we assume that the system obeys the Abbé sine condition (A2.3f), then the resolution criterion can be expressed as

\[ \delta x = 0.61 \frac{\lambda}{\sin \alpha_0} \]

where \( \delta x \) is the minimum resolvable distance in the object and \( \alpha_0 \) is the half angle subtended at the object by the entrance aperture of the system. From this one may expect that the ultimate resolution in any microscope cannot be much better than the wavelength of the illuminating beam.

2.4 Lasers. A laser in its simplest aspects consists of an amplifying medium, a source of input energy, a cavity,
and an egress for the output radiation. It is essentially
an oscillator for which the feedback mechanism is provided
by the cavity. Energy is supplied to the atoms of the
medium to excite the majority of them to a higher electronic
state. In the process of reverting to the ground state,
these atoms radiate certain characteristic frequencies. Be-
cause of the cavity, only a very limited number of modes of
this radiation will be retained. The rest are absorbed by
the cavity walls. The modes that are retained, however,
traverse the length of the cavity and stimulate other atoms
to radiate. This stimulated radiation will, because of the
fundamental Bose properties discussed earlier, enter the
same mode with increasing probability. Thus, there is a
tendency for modes preferred by the cavity to be amplified.
The amplification continues until various losses offset
the gain, and steady state operation ensues. It might
seem that the of the entire laser system should become
infinite and that the bandwidth of the output should be
essentially zero. This does not happen because there is
a finite amount of noise generated which effectively sets
a limit on the bandwidth; nevertheless it can be very narrow.

In a typical gas laser the interaction between the
atoms and the radiation field is an electric dipole inter-
action and because of the discreteness of the atomic
states, only transitions involving well-defined amounts
of energy are possible. This means that only photons of
certain frequencies are emitted or absorbed and that the
original relatively broadband input energy is converted
to an essentially discrete spectrum within the cavity. By
considering the cavity radiation field to act as a per-
turbation on the free atom Hamiltonians (A2.4a) we can
derive an expression for the absorption and induced
emission probability rates. The two probability rates are
equal and are proportional to the density of the radiation.
They are also proportional to the square of the expectation
value of the perturbation energy between the two states involved. Taking into account that the radiation field is discrete, that the energy levels may be degenerate, and that there is also spontaneous emission not related to the radiation field, we can derive expressions for upward and downward transition probability rates (A2.4b). Next, if we multiply these probability rates by the populations of the respective levels, we obtain the absolute transition rates. Recognizing that a radiation mode corresponds to a pair of oppositely traveling waves we may express the transition rates in terms of energy flux along the direction of wave travel. We obtain for the flux

\[ I_\nu(\xi) = I_\nu(o) e^{-\alpha \xi} \]

where

\[ \alpha = (n_k - n_\xi) f(o) \]

is the gain factor. \( n_k \) and \( n_\xi \) are the population densities (assumed for the moment non-degenerate) of the lower and upper levels, respectively, and \( f \) is a function of the frequency. Now \( f \) is positive definite, so if amplification is to take place, then \( n_\xi > n_k \); that is, we must have a population inversion (A2.4c). This is a necessary condition for oscillation but not a sufficient one. There are losses in the cavity (A2.4d) which cause the flux to decay so that the population inversion must be larger than it would otherwise be. Expressing the loss by the decay time parameter \( \tau_c \), we may obtain the general oscillation threshold condition (A2.4e)

\[ \left[ n_k - \frac{g_k}{g_\xi} n_\xi \right] > \frac{8 \pi v_o^2}{c^2 g(o)} \cdot \frac{t_{spon}}{\tau_c}. \]

\( \frac{g_k}{g_\xi} \) accounts for level degeneracies, \( v_o \) is the oscillation frequency, \( g(o) \) is the amplitude of the radiation distribution at \( \nu = v_o \), and \( t_{spon} \) is the spontaneous emission time from state \( \xi \) to state \( k \). Provided this condition is satisfied, amplification of the mode will take place and con-
tinue until losses equal gain. At this point the mode is said to be saturated.

The output bandwidth of a laser is related to the total $Q$ of the laser system (A2.4f). The $Q$ in turn is related to the energy stored in the cavity and the steady state power loss. Part of the power loss is, of course, the output power and the rest is related to cavity losses, which we can describe by a characteristic decay time. Assuming these form a complete set of variables and arguing from a dimensional point of view, we obtain the proportionality

$$\Delta \nu_{\text{out}} \propto \frac{(\Delta \nu_{\text{cav}})^2 h \nu}{\rho_0}$$

where $\Delta \nu_{\text{out}}$ is the output bandwidth, $h \nu$ the cavity energy, $\Delta \nu_{\text{cav}} = 1/\tau_c$ is the reciprocal of the cavity characteristic decay time, and $\rho_0$ is the output power. From this we conclude: the faster the output beam can deplete the stored energy, the more narrow the bandwidth, and the smaller the cavity losses the narrower the bandwidth.

References


9. Maitland and Dunn, Chapter 4.
Chapter III
Properties of Helium-4

The common isotopes of helium are He-3 and He-4. Because of the odd number of nucleons, He-3 acts like a fermion and He-4, with an even number of nucleons, acts like a boson. It is He-4 in which we are interested. To be as apparently simple a substance as it is, He-4 has a surprising number of unusual properties. The properties of particular interest for this study are those related to excitations and flow. In particular we shall be interested in the properties below the $\lambda$-point, or below about $2.2^\circ K$. In this temperature range, superfluidity is manifested. Superfluid flow is characterized by apparently zero friction and by macroscopic quantum behavior. In some respects He-4 at these temperatures acts like an ideal Bose gas and it is here, perhaps, that the analogy between coherent photons and helium is closest.

3.1 General. The excitation properties of He II (that is, He-4 below the $\lambda$-point) can be summarized by a plot of energy versus momentum as shown in figure 1. The heavy curve is experimentally observed and represents the energy distribution of helium atoms as a function of momentum. The graph can be divided into three regions which correspond to the rather distinct regions suggested by the curve. The $E$-$p$ relationship in the lowest
momentum regime is interpreted as being due to the excitation of phonons. The wavelengths are long and the motion is collective; that is, involves many individual atoms. A consequence is that such excitations produce long range density variations in the He II. The E-p relationship in the middle region is interpreted as being due to stirring or local rotational motions of small numbers of atoms. The wavelengths are on the order of the atomic spacing and do not involve long range variations in the density. The E-p relationship for the highest momentum region is interpreted as being due to excitations of essentially free atoms. The wavelengths are shorter than the interatomic spacing and the mean free paths are long.

Landau\(^8,9\) considered the phonon and roton excitations as gases that move through the helium essentially without interaction. This led him to argue that He II could be considered as a composite fluid consisting of a superfluid component that flows without viscosity, and a normal component that gives rise to viscosity and other normal fluid properties. The former would obey the Euler equations of motion and the latter, the Navier-Stokes equations of motion. The two-fluid model, as this concept is called, is useful in explaining at least semi-quantitatively many properties of He II. For example, not only can one derive the equations of motion of ordinary sound but one also predicts the existence of an entropy wave, called second-sound by Landau. The model breaks down at higher momenta because it is then no longer possible to regard the "two fluids" as being non-interacting.

Based on the phonon and roton excitation theories, one would predict that excitations could not occur for E/p ratios of less than about 6.0 \(\times\) 10\(^3\) cm/sec. It turns out, however, that some sort of excitations take place for E/p ratios two orders of magnitudes or so less than that. Since E/p corresponds to a velocity, this is equivalent to
saying that superflow stops at flow rates much smaller than would be predicted by the theory so far. The answer is that in flowing He II, vortices can form at relatively low flow rates and absorb the kinetic energy of flow. This effectively destroys the superflow. Such vortices are quantized and are strongly affected by the geometries of flow. For example, flow rates through capillaries depend critically on capillary diameters, etc.

3.2 Ground State. One of the first attempts to develop a theory to account for the unusual properties of liquid helium was by F. London. He assumed that liquid helium could be considered as an ideal Bose-Einstein gas (B3.2a-B3.2c) and that below some characteristic temperature $T_c$, macroscopic numbers of the helium atoms would begin to "condense" into a lowest energy state. This lowest energy state could not really be zero since in order for the He-4 to remain liquid at absolute zero, zero-point energy is required ($\sim 30 \text{ cal/mole}$). The specific heat relationship for such a model, figure 1, shows a discontinuity in slope at $T_c$ which is qualitatively related to the actual discontinuity observed. Below $T_c$, which works out to about $3.1^\circ K$, the model specific heat is given by

$$C_v = 1.98 R \left( \frac{T}{T_c} \right)^{3/2},$$

where $R$ is the gas constant. The specific heat thus obeys a $T^{3/2}$ law whereas at very low temperatures the actual $C_v$ obeys a $T^3$ law. Refinements of the model which take into account an average interaction

![Figure 1](image-url)
with the rest of the gas do not alter the incorrect specific heat behavior.

The first approach to the theory of the ground state from a microscopic viewpoint was made by Feynman\textsuperscript{10} in 1955. On qualitative grounds (B3.2d) he argued that the ground state wave function was everywhere positive and that it had maximum amplitude for a uniform density distribution of the atoms.

3.3 Phonon Excitations. Landau\textsuperscript{8,9} suggested that the lowest energy excitations could be explained by assuming the equivalent of a Debye solid model for He II. This model considers the liquid as a lattice and relates the excitations to phonon modes (B3.3a). The various thermodynamic properties are easily deduced (B3.3b) and it is found that the specific heat now obeys a $T^3$ law.

Feynman\textsuperscript{10} extended this model to obtain the form of a wave function which could account for the lowest level excitations (B3.3c). He further was able to show that these had to be phonons.

3.4 Roton Excitations. The Landau model\textsuperscript{8,9} also included the theory of a second distinct type of excitation which Landau believed was related to rotational modes of motion. The energy-momentum spectrum he proposed was

$$E_{ro} = \Delta + (p-p_0)^2/2\mu,$$

where $\Delta$ is an energy gap, $p_0$ the momentum for which the roton spectrum is minimum, and $\mu$ an effective mass. By adjusting the values of these parameters, a good fit can be made to the roton portion of the spectrum, figure 3.1-1. From this relation the roton contribution to the specific heat and the entropy of He II can be determined (B3.4a).

Feynman\textsuperscript{10} deduced the form of the roton wave function and obtained an expression for the roton energy spectrum,

$$E = \hbar^2 k^2 / 2m S(\kappa),$$
where $S(k)$ is a form factor. Furthermore, he suggested a neutron scattering experiment from which it should be possible to obtain the entire excitation spectrum quite accurately. Subsequently a number of such experiments were done and the spectrum was indeed obtained.

3.5 Two Fluid Model. The two fluid model originally suggested by London and Tisza and extended by Landau, has been quite useful in explaining, at least semi-quantitatively, a number of the He II properties. Landau used it to explain why the phenomenon of superfluidity existed (B3.5a). As discussed before, the two fluid model considers He II as composed of two fluids; one which has zero viscosity and carries no entropy, and the other which has the usual properties of an ordinary fluid. The first is governed by an adaptation of the Euler equation of motion

$$\rho_s \frac{D}{Dt} \frac{\nu_s}{c_s^2} = \rho_s S \nabla T - \frac{\rho_s}{c_s} \nabla p$$

and the second by an adaptation of the Navier-Stokes equation of motion

$$\rho_n \frac{D}{Dt} \frac{\nu_n}{c_n^2} = -\rho_s S \nabla T - \frac{\rho_n}{c_n} \nabla p - \eta_n \nabla \times (\nabla \times \nu_n) .$$

The derivation of the corresponding classical equations is developed in appendices (B3.5b) through (B3.5f). A combined form which shows that a temperature gradient produces a separation in the velocities of the two components is given by

$$\rho_n \frac{D}{Dt} (\nu_n - \nu_s) = -\rho_s \nabla T .$$

(B3.5g)

It is beyond the scope of this study to go very far into the ramifications of two-fluid flow and the extensive experimental work which has been done. However, the equations of motion can be used to derive the propagation velocities of ordinary sound and second sound, and this is done in appendix (B3.5h).

The difficulty with the two-fluid theory is that the ground state atoms and the atoms in the higher momentum states do interact and the equations of motion are coupled.
Furthermore, at the higher flow energies, one can expect the usual onset of turbulence for which no adequate theory exists.

3.6 Flowing Helium. In its ground state helium can flow without friction or loss provided that the relative velocity of one element of the fluid moves quite slowly with respect to an adjacent element or a boundary. As soon as this relative velocity exceeds, say, $10^{-1} - 10^{1} \text{cm/sec}$, depending on the geometry, losses develop and superflow ceases. What happens is that rotational motion begins and extracts kinetic energy from the flow. The precise mechanics of how it begins is not known. From the solution of the classical Navier-Stokes equation with cylindrical boundary conditions, one would expect either rigid body or vortex motion (B3.6a). For identical boundary conditions the latter is energetically more favorable (B3.6b) and experiments seem to confirm that the loss of superflow at low velocities is probably due to the formation of vortices. Since vorticity is conserved (B3.6c), it follows that vortex lines either terminate at the boundaries or else close on themselves to form rings. The velocity field in a fluid due to a vortex line is given by the fluid mechanical equivalent of the law of Biot-Savart (B3.6d),

$$\mathbf{v} = \frac{\mu}{4\pi} \int \frac{dl \times r}{r^3},$$

where $\mu$ is the "strength" of the vortex. This relation can be used to derive the translational velocity of ring vortices (B3.6e) and deduce their motional properties in the presence of boundaries, etc.

In most practical cases in fluid mechanics, it is very difficult to make use of the equations of motion and derive analytically the properties of a flowing fluid. The few instances for which analytical methods yield complete solutions usually involve laminar flow and very simple boundary conditions; for example, low Reynolds number flow in pipes,
etc. Some of the background theory of this is developed in appendices (B3.6f) through (B3.6g). At high Reynolds numbers full turbulence is developed and some problems can again be treated in a reasonable satisfactory manner; however, for our purposes we shall not be concerned with these.

Feynman$^{10}$ in his attempts to develop a microscopic quantum theory of liquid helium considered the problems of flow. He assumed essentially that, except at a finite number of points, the velocity distribution could be expressed as the gradient of a scalar velocity field; that is, would mostly be irrotational (B3.6h). This led to the proposal that circulation was quantized and given by

$$\oint \mathbf{v} \cdot d\mathbf{l} = \iint (\nabla \times \mathbf{v}) \cdot dA = 2\pi \frac{h}{m} \cdot n \quad ; \quad n = 0, 1, 2, \ldots$$

He then considered the rotational aspects of helium flow (B3.6i) and concluded that vortex motion for macroscopic bodies was energetically slightly more favorable than rigid body motion. Next, to explain the loss of superflow at velocities well below $6.0 \times 10^3 \text{cm/sec}$, he proposed that as soon as the kinetic energy of flow exceeded the energy of a minimum quantum vortex, vortices would begin to form and extract energy from the flow (B3.6j). In this manner the superflow properties would effectively be destroyed. Finally, in an attempt to obtain a more physical explanation for rotons, he suggested, as did Onsager$^{15}$ earlier, that rotons might be the smallest vortices (B3.6k). A vortex of radius $2.2\AA$ had the momentum $p_\circ$ of a roton and an energy of the right order.
References

Chapter IV

Helium-4 Beams

The practical objective of this study is to determine if a suitably dense, coherent, and energetic beam of helium-4 atoms can be generated for use in a microscope capable of resolving the structure of single organic molecules. In this chapter we shall attempt to touch on most of the pertinent aspects of the objective. All of the analyses will be qualitative or order-of-magnitude, and any instrumental configurations, conceptual.

The imaging system we have assumed to be lensless. As mentioned in the introduction, this is not a necessary condition. It does, however, allow us to avoid many problems related to the microscope design which, with helium atoms instead of photons, might prove quite difficult. Assuming lensless holography, the first question is how to record the scattered He-4 beam and retain the necessary information to reconstruct a three dimensional image of the target. This leads to questions of wavelength, geometry, flux and coherence. Next, the general approaches to beam generation must be considered. Beam properties must be compatible with the imaging requirements. Attainment of adequate coherence in the beam is the principal problem and methods for filtering any beam must be considered.

4.1 Imaging Requirements. An imaging system consists of an illuminating beam, a target, a lens system, and a detection system. It is not necessary that the formation of an image be carried out in a one-step process. As a matter of fact, it is assumed here that the image formation will be at least a two-step one, with the detection immediately following the scattering of the beam. In the discussion that follows we shall not be concerned with how the final
image is formed; we shall be concerned only with the problems of effecting the initial detection.

The most fundamental factors affecting ultimate resolution are the wavelength of the illuminating beam and the effective aperture of the imaging system. These are expressed by the Rayleigh resolution criterion,

\[ \Delta \xi = \frac{1.22 \lambda}{\sin \alpha} \]

where \( \alpha \) is the effective angular aperture as seen by the object. Assuming a required resolution on the order of \( 1 \text{Å} \), and a beam energy of about \( 0.024 \text{eV} \), the detection system must subtend an angle at the target of \( 35^\circ \) or more.

In order to record a hologram, we must form an interference pattern between the beam scattered by the target and a mutually coherent reference beam. There are at least three different ways of doing this as illustrated in figure 1. The first is an in-line geometry originally suggested by Gabor; the second is the common off-axis geometry originally proposed by Lohmann; and the third is the lensless Fourier transform method of Stroke. We wish to look at the resolution limitation imposed by the "graininess" of the detector system in the above configurations. For this purpose the first two are essentially equivalent.

Consider figure 2 which represents two plane waves pro-
ducing fringes. In order for the received wave at the detector to be unambiguously interpretable

\[ \alpha_o = \sin^{-1} \frac{\lambda}{d} \approx \frac{\lambda}{d} \]  

It follows that the minimum resolvable distance in the object is given by

\[ d \approx 2l \]  

Let us assume that there exists a photographic-type emulsion that can record the impacts of helium atoms. It is unlikely that the emulsion resolution will be greater than a line spacing of 5 microns. Using the above relation the resolution limit in the object is on the order of \( 10^6 \) ft, which is obviously unacceptable.

The crux of the problem lies in the angular relationship between the object and reference waves and is not fundamental. Suppose \( \Delta(\sigma) \) represents the phase shift between reference and object waves as a function of the position coordinate on the detection surface. The fringe spacing \( s \) is then found from the relation,

\[ \frac{d\Delta}{ds} = \frac{\lambda}{s} \]  
or

\[ s = \lambda \frac{d\sigma}{d\Delta} \]  

The ratio of \( s \) to \( 2l \) is the factor by which the resolution limit can be increased. To see how much this can be, consider the lensless-Fourier-transform geometry shown in figure 3. In this case
The situation in the previous case can be represented as shown in figure 4. Here

\[ \Delta = R \sin i' \]

and

\[ \frac{d\Delta}{di'} = R \cos i' \]

or

\[ \frac{d\Delta}{d\sigma} \cdot \frac{d\sigma}{di'} = R \cos i' \]

But

\[ \frac{d\sigma}{di'} = R \]

so

\[ \frac{d\Delta}{d\sigma} = \frac{R}{R} \cos i' \]

The situation in the previous case can be represented as shown in figure 4. Here

\[ \Delta = R (\sec i' - 1) \]

and

\[ \frac{d\Delta}{di'} = R \sec i' \tan i' \]

But

\[ \frac{d\Delta}{di'} = \frac{d\Delta}{d\sigma} \cdot \frac{d\sigma}{di'} = \frac{d\Delta}{d\sigma} \cdot R \sec^2 i' \]

so

\[ \frac{d\Delta}{d\sigma} = \sin i' \]

It follows then that the resolution of the geometry of figure 3 is greater than that of figure 4 by the factor

\[ \frac{R \lambda \sin i'}{h \lambda \cos i'} = \frac{R}{h} \tan i' \]

By choosing \( R \gg h \) or \( i > \frac{\pi}{4} \) or both, the resolution can be improved to approach the Rayleigh resolution limit.

From the above analysis we conclude that the lensless-Fourier-transform geometry is the general type required for imaging with a helium beam.

Next let us consider possible types of detectors or
detector systems. Two common methods of helium detection are ionization followed by a photomultiplier detector, say, or directly by means of a very sensitive pressure detector such as a Stern-Pirani detector. The latter converts changes in pressure into changes in resistance of wires. These being part of a bridge network can be calibrated in terms of flux. The sensitivity of such devices is on the order of \(10^{15} \text{ atoms/sec.-sr.}\) which for our purposes is much too low. The highest density gaseous beams of helium that we can expect to achieve will probably not have fluxes in excess of \(10^{20} \text{ atoms/sec.-sr.}\). The total scattering cross sections of typical atoms with helium will be on the order of \(10^{14} \text{ cm}^2\) or less so that something like \(10^6 \text{ atoms/sec.}\) will be scattered. This of course is many orders of magnitude below the threshold sensitivity. It is probable, therefore, that the helium atoms will have to be ionized immediately after scattering and then detected, either with or without acceleration. It is likely, though, that acceleration would simplify the detection problem. If sufficiently large energies were imparted to the atoms, it might then be possible to use photographic emulsions. A conceptual detection scheme is shown in figure 5. The higher the field and the shorter the distance between screen and plate, the less the influence of scattering off the screen.

Beyond what we have already said, it is difficult to make many

![Figure 5](image-url)
quantitative statements about the ultimate resolution of imaging with He-4 beams. A great deal depends on details of construction and geometry that at this point are completely unknown. We can, however, make some general statements and/or suggest areas where further analysis is still feasible.

The resolution of an object depends, as we have said, upon the deBroglie wavelength of the scattered helium. That does not necessarily mean however that the internal atomic structure of a molecule can be resolved. The problem is that helium is an atom with an effective diameter which is determined by its electrons. In a typical scattering by a carbon or a nitrogen atom, one may expect the distance of closest approach to be on the order of $2-3\text{Å}$. Qualitatively, this means that the best possible theoretical image of a molecule would have the appearance of an ordinary space-filling model used by chemists. One might expect that the holes in some ring type molecules would be visible. Effective hole diameters in a typical cyclic molecule such as benzene are on the order of $2-3\text{Å}$ also.

Noise factors that will strongly affect resolution are target motion during scattering, variation in phase between reference and object beams as a result of vibration, unwanted scattering within the beam and elsewhere in the instrument, etc. How severe some of these factors will be is not known but it is likely none will be insignificant.

Coherence requirements are simple. The coherence of the incident beam must be such that its coherence volume is greater than the volume occupied by the object. Furthermore at all points on the detector surface, the reference and object beams must be coherent. This latter requirement essentially sets the minimum temporal coherence required, assuming the spatial coherence is adequate. If points $d$ apart along the beam are to be coherent, then it is easy to see that
\[ \Delta v \leq \frac{\lambda}{\Delta \sigma} v, \]

where \( \Delta v \) is the velocity bandwidth. If we require \( \Delta l \sim 1 \text{cm} \) then

\[ \Delta v \approx \frac{10^{-8}}{1} \times 10^5 \sim 10^{-2} \text{cm/sec}. \]

This is an extremely narrow bandwidth and will be very difficult to achieve.

Flux requirements depend critically upon the detection system used. The minimum flux will be determined by the time required to image with a given detector sensitivity. Maximum fluxes will be determined either by detector saturation (overexposure, for example) or by target damage. A molecular structure can be considered as a set of coupled harmonic oscillators having characteristic frequencies. If the incident helium atoms have an effective frequency corresponding to one of the characteristic molecular frequencies, then successive scatterings will probably excite the vibrational modes into successively higher states, resulting in the rupture of the molecule. This will happen if the time interval between collisions is short compared to the lifetimes of the vibrational states.

4.2 Beam Types and Properties. There are two basic classes of helium beams, liquid and gaseous, and several approaches for generating each. Liquid beams may be continuous or they may be in the form of droplets. Gaseous beams may be neutral, excited, or ionized. It seems probable that liquid beams will have to be gasified prior to scattering to prevent excessive damage to the target and to avoid interatom scattering.

Droplets of liquid helium are easily formed with the size being determined principally by the surface tension, the size of the capillary on which they are formed, and the external force field. Rates of formation are of course a function of capillary size, which determines the critical flow velocity of the helium for a given
temperature. A numerical example will give a feel for the orders of magnitude involved.

The volume of a falling drop is on the order of

\[ V = \frac{\kappa (2\pi r V)}{(\rho g + f)} \]

where \( r \) is the external capillary radius, \( V \) is the surface tension, \( \rho \) is the density, \( g \) is the gravitational acceleration, \( f \) is any other external force besides gravity, and \( \kappa \) is a numerical constant on the order of about 0.8. Letting \( f=0 \), \( V = \frac{.37 \text{erg/cm}^2}{\rho} = \frac{.145 \text{gm/cm}^2}{\rho} \), we obtain for \( V \) as a function of \( r \),

\[ V \sim 1.25 \times 10^{-3} r. \]

The expression for flow rate of superfluid helium through a capillary is assumed to be

\[ R = \left( \frac{\rho_s}{\rho} \right) V_{sc} \pi \frac{d^2}{4}, \]

where \( \rho_s/\rho \) is the ratio of superfluid to normal fluid, \( V_{sc} \) is the critical velocity and \( d \) is the inside capillary diameter. At 10K, \( \rho_s/\rho \sim 1 \) and \( V_{sc} \sim 25 \text{cm/sec} \) for \( d \sim 3 \times 10^{-6} \text{cm} \). Therefore

\[ t_{\text{formation}} = \frac{V}{R} \sim 210 \text{sec}. \]

A drop falling under the influence of gravity alone will not reach velocities of the order of \( 10^6 \text{cm/sec} \) in any reasonable size apparatus. Therefore it must be subjected to an external force, and an obvious way to do this is to charge the drop and accelerate it through an electrostatic field. Positive helium ions or alpha particles probably present the most attractive approach since these, through polarization, attract clusters of other helium atoms about them. \(^9\)-\(^11\) When the ion is subjected to an electrostatic field, it tends to stay with the liquid instead of being extracted from it because of the relatively strong attraction to 50-100 of its neighbors. It should therefore be
possible to subject the drop to considerable fields in order to reach full beam velocity in a relatively short distance. From the data of Careri et al.,\textsuperscript{11} at temperatures below 2\textdegree K we would estimate the probability per second of extracting a positive ion in a field of intensity 300 \text{V/cm} as being less than \text{0.05}. A 1 cm drop has a mass of about 10\textsuperscript{-4} grams. If we assume an acceleration distance of 1 meter and a field of 300 \text{V/cm}, the charge that we have to impart to the drop for it to reach 10\textsuperscript{5} \text{cm/sec} at the end of 1 meter is roughly 10\textsuperscript{4} esu. The time to traverse the distance is on the order of 10\textsuperscript{-3} sec. This is a rather large charge but it is quite likely that the field intensity could be substantially increased resulting in a smaller charge per drop.

Ideally one would hope to have no flow inside a drop so that there would be no dispersion of velocity. This might be achievable provided vortex motion is not set up during the drop formation and provided a certain amount of evaporation takes place during flight. Evaporation would extract energy from the kinetic energy of motion and lower the drop temperature. Since circulation is quantized in liquid helium, it is possible to make a crude estimate of the maximum rate of drop formation to avoid the formation of vortices. Consider figure 1. Assume a ring vortex is formed of diameter \(a\). From equation (B3.6k-2), we have for the minimum angular momentum of a ring vortex of radius

\[ \mathcal{L} = 2\pi \rho \frac{\pi}{m} a^3 = \frac{3}{2} \pi \frac{\pi}{m} M, \]

where \(M\) is the total drop mass. Now assume that each element of entering mass contributes to the angular momentum in the amount \(\delta M a^2\) so that the total angular momentum imparted during drop formation is \(\frac{1}{2}Mva\). Equating this to the above expression for the angular momentum of a ring vortex and

![Figure 1](image)

\(8Mva/2\)
solving for the velocity we get

\[ V = \frac{3}{2} \frac{\pi \frac{h}{2}}{M} = \frac{3\pi \frac{h}{2}}{M}. \]

Assuming \( 2a \sim 1 \text{cm} \), we obtain

\[ V \sim 3 \times 10^{-2} \text{ cm/sec}. \]

Since this result is \( 10^3 \) times smaller than the velocity used previously, it suggests that the formation of vortices in drops might be a problem.

Another factor that should be analyzed when considering droplet beams is evaporation. As a drop becomes smaller the surface area to volume ratio becomes larger and the lifetime of the drop (assuming constant temperature) decreases significantly. An upper limit to the mass loss can be obtained if we assume that the evaporation process is adiabatic; that is, if we assume that all the heat of vaporization must be supplied by the thermal energy of the drop. The thermal energy is given by \( E = \int C_v \, dT \) where \( C_v \) is the specific heat. Using experimentally determined values for \( C_v \), the maximum available energy at \( 2^\circ K \) is approximately \( 1.87 \times 10^7 \text{ erg/gm} \). Thus

\[ E \sim 1.87 \times 10^7 \times M \text{ erg}, \]

where \( M \) is the drop mass. Since the latent heat at \( 2^\circ K \) is approximately

\[ 90 \times 10^7 \text{ erg/mole} \sim 25 \times 10^7 \text{ erg/gm}, \]

we have,

\[ \delta m \times 2.5 \times 10^8 \leq 1.87 \times 10^7 \times M \]

or

\[ \frac{\delta m}{M} \leq \frac{1.87}{25} \sim 7\%, \]

where \( \delta m \) is the mass loss. Thus the maximum possible loss at \( T \sim 2^\circ K \) under adiabatic conditions is on the order of 7%.

The really major problem connected with liquid beams, continuous or droplet, is the probable need to gasify the beam. It is not at all clear how this should or can be done while at the same time maintaining a high degree of coherence.
Next let us consider the possibilities of generating a coherent continuous liquid beam. The first and most critical problem is turbulence which, for the normal fluid, sets in at Reynolds numbers of 1200 or less. While some vorticity might be acceptable, at least in the beam formation stage, full turbulence is almost certainly not. The concept of Reynolds number as applied to the two-fluid model of liquid helium is a complex and uncertain subject and is beyond the scope of this thesis. However, we can make an estimate of the maximum normal fluid velocities beyond which the flow is turbulent. We have the expression for Reynolds number

$$R_e = \frac{\rho v d}{\eta},$$

where $\rho$ is the density, $d$ is the pipe diameter and $\eta$ is the viscosity. Assuming $d \sim 10^{-3} \text{ cm}$, $\rho \sim 1.15$, $\eta \sim 25 \times 10^{-6} \text{ poises}$, and $R_e 1200$, we obtain for the critical velocity

$$v \sim \frac{1200 \times 25 \times 10^{-6}}{1.15 \times 10^{-1}} \approx 2 \text{ cm/sec}.$$ 

While this calculation is necessarily rough, it does give a critical velocity that is of the same order as experimental results.

The onset of vorticity in the superfluid component can be estimated from Feynman's theory, (B3.6j). Essentially, the theory says that when the kinetic energy of flow equals the minimum vortex energy, critical velocity has been reached. This velocity is approximately given by

$$v_c \approx \frac{\eta}{md} \ln \frac{d}{a},$$

where $d$ is the orifice diameter and $a$ is the minimum vortex core radius. For $d \sim 10^{-3} \text{ cm}$ and $a \sim 10^{-6} \text{ cm}$ we obtain

$$v_c \sim \frac{10^{-3}}{10^{-7}} \sim 1.6 \times 10^{-1} \text{ cm/sec}.$$ 

Again this is on the order of critical velocities observed in experiments and about a factor of 10 lower than the critical velocity for turbulence of the normal fluid.
Regardless of which velocity we accept as the maximum, it is exceedingly low compared to the required beam velocity. It is obvious that the beam must be accelerated. Two ways in which this could be done is to put the source in relative motion with respect to the target, or to implant ions in the beam and try to accelerate it as was done in the case of the droplet beams. If the source were rotated relative to the target the first would require, for example, an angular velocity of about \(10^3\) radians per sec if the orifice were one meter distant from the center of rotation. The second might be possible but would almost surely result in breakup of the stream into droplets. Furthermore, the subjecting of positive ions in the cylindrical stream to high electrostatic fields would almost surely cause vortices to form with the attendant dispersion of the beam velocity field. Altogether the continuous stream approach does not appear practicable.

There are at least two ways of forming initially neutral gas beams; thermally and mechanically. Thermal methods are well known and involve heating a gas to some temperature in an oven containing an exit slit. The gas molecules effuse from the slit and the beam is formed by means of collimating slits. The velocity distribution in such a beam can be expected to be Maxwell-Boltzmann.

Mechanical methods involve putting a gas under pressure and allowing it to expand from a nozzle. This technique has been used increasingly in the past ten years to produce very high intensity beams of relatively narrow velocity dispersion. Applications include studies of surfaces, chemical reactions in colliding beams, aerodynamic studies and so forth. The basis for the technique is the Kantrowitz-Grey supersonic nozzle shown in

![Diagram of a supersonic nozzle](image)
figure 2. Gas expands as it comes out of the nozzle causing the temperature to drop in the beam. Only the central portion of the beam is passed by a skimmer. The velocity distribution in such a device is proportional to
\[ e^{-\frac{m(u-U)^2}{kT_i}} \]
where \( u \) is the usual thermal velocity, \( U \) is the mass flow velocity, and \( T_i \) is the temperature at the skimmer entrance. This last is given by
\[ T_i = T_o \left[1 + \frac{c}{k} M^2 \right]^{-1} \]
where \( T_o \) is the temperature at the nozzle, \( M \) is the Mach number at the skimmer, and \( \gamma \) is the ratio of specific heats at constant pressure and constant volume. From these relations it is clear the lower \( T_i \) is, the narrower the velocity spread. For high Mach numbers this can be very small. In the case of beam velocities of \( \sim 2 \times 10^5 \text{ cm/sec} \), \( M \sim 10 \) and \( T_i \sim 0.03 T_o \). This implies a reduction in width of the velocity distribution by a factor of 30. Vyse et al. have actually constructed a low temperature helium source of the type described above and found for a beam velocity of \( \sim 3 \times 10^4 \text{ cm/sec} \) a velocity spread \( \sim 2.5 \times 10^3 \). This compares with about \( \sim 2.2 \times 10^4 \) for a thermal distribution at the actual nozzle temperature of 5.9\( ^\circ \)K.

Beam intensities of such sources as high as \( 10^{19} - 10^{20} \text{ particles/sec-sr} \) have been achieved; about 500 times, at Mach 10, the flux of the best thermal sources. The limitation is condensation of the particles in the formation of dimers, trimers, etc. It is reasonable to suppose that such condensation would not be acceptable for the purposes of microscopy.

Other ways of forming gaseous beams with velocities of \( \sim 10^5 \text{ cm/sec} \) include ionization and excitation. The former involves ionization of the helium either in an RF discharge tube or by means of an electron beam, then accelera-
tion to full beam speed, followed by neutralization in a charge exchange process. The latter involves exciting the helium to some higher state, accelerating it in a magnetic field, and having a sufficiently long path length such that it can decay back to the ground state prior to collision with the target. This last technique has not yet been explored in detail but it looks like a rather unlikely approach, principally because lifetimes in a specific and very narrow range would be required and extremely large magnetic fields would be needed. Another problem common to both is the fact that the velocity dispersion would be Maxwell-Boltzmann. Also ionized beams tend to spread unless they can be channeled in magnetic fields, for example.

Many unorthodox approaches to beam generation have been considered during the course of this study. The starting point for all has been He II because of its inherent coherence properties and related macroscopic quantum phenomena. Generally speaking, the difficulty in obtaining beam velocities of the order of $10^5 \text{ cm/sec}$ from He II is that the energies involved correspond to temperatures in the range $10^3 - 10^4 \text{ K}$, not to $1-2 \text{ K}$. This means that there has to be center of mass motion of the low temperature helium on the order of $10^5 \text{ cm/sec}$ to take advantage of its attractive properties; hence, the consideration of accelerated liquid beams and drops.

One variant approach that might have possibilities but which has not yet been explored in detail is suggested by analogy to the laser. The output particle of the laser is the photon; the output particle of the helium device is a neutral atom. The phonon is generated by an electron changing orbits (energy levels) in an atom. The closest analogy that has been thought of so far for the atom is a lattice, and for the electron, a phonon. Thus one might envision a helium atom trapped in the lattice, a phonon being annihilated in the process of elevating the atom to the former's energy and linear momentum state, and the
atom passing out through the lattice. A necessary require¬
ment on the lattice would be that the energy barrier between
cells as seen by the helium atom correspond to velocities
on the order of $10^5$ cm/sec or perhaps, less. If such a
lattice (crystal) could be found, then the device might just
possibly work. One can think of many ramifications and
conditions such as cavity geometries, density of lattice
filling, etc. in order to achieve a monochromatic linear
momentum state, but these thoughts are vacuous unless a
suitable lattice can be found. The first step, therefore,
is to specify theoretically the necessary properties of the
lattice and then to search for possible materials.

4.3 Beam Filtering. All methods of beam generation
discussed so far require strong filtering to reduce the veloc-
ity spread to $\Delta v \approx 10^2$ cm/sec. There are several ways of
approaching the problem. One is to use a mechanical chopper,
a second is to diffract the beam off the surface of a
crystal such as LiF, and a third is to velocity select by
the deflection of an ionized beam. It is easy to see that
a chopper is likely to be impractical for several reasons.
Aside from mechanical vibrations which are undesirable in
a microscope, rotational speeds on the order of $10^6-10^7$
revolutions per second would be required.

Diffraction off the surface of a crystal at first
glance appears to be a possible approach. From the grating
equation we have

$$\sin \theta_k - \sin \theta_i = \frac{k \lambda}{d} \quad k = 1, 2, \ldots,$$

where $k$ is the diffraction order, $d$ the grating spacing,
$\theta_i$ the incident angle, and $\theta_k$ the $k^{th}$ order diffraction
angle. To find the sensitivity of $\theta_k$ with respect to $v$,
differentiate and make use of the deBroglie relation to
obtain

$$\delta \theta_k = \left[ k \frac{m \lambda}{h d \cos \theta_k} \right] \delta v$$

Since $\theta_i$ is known and assuming $k=1$, the grating equation
can be used to find $\theta_{ir}$. However, let us assume the reasonably favorable case where $\theta_{ir} \approx 85^\circ$ so that $\cos \theta_{ir} \approx 0.1$.

For $d \approx 2 \times 10^{-3}$ cm we get

$$\delta \theta_{ir} \approx \left[ \frac{6.65 \times 10^{-24} \times 10^{-16}}{6.62 \times 10^{-27} \times 2 \times 10^{-1}} \right] \delta v$$

$$\approx \left[ 5 \times 10^{-4} \right] \delta v.$$

Assuming that we can accept $\delta v \approx 10^{-2}$ cm/sec and that the beam diameter on the target should be $\approx 1$ cm, we find for the grating-target distance

$$L \approx 10^{1.5} \times 10^{-16} \approx 2 \times 10^5 \text{ cm}.$$

Thus it appears that surface diffraction will not be practicable either. The above calculation neglects diffraction effects due to the finite diameter of the beam. Using the Rayleigh criterion one can make a crude estimate of this as follows. If $d = 1$ cm then

$$\alpha \approx \sin \alpha = \lambda / d$$

$$\approx 10^{-8} / 10^{-1} \approx 10^{-7}.$$

Thus uncertainty in the diffraction angle is of the same order as $\delta \theta_{ir}$, and unless a wider beam were used, any size apparatus would not yield the narrow velocity width required.

The third approach appears to be the most promising although it is beset with all the problems associated with handling high density ionized beams. One technique would be to direct the charged beam between electrostatic deflection plates and require that it exit through a small diameter hole which is so placed that it would pass only the desired band of velocities. Aside from problems of preventing the beam from expanding laterally, there is considerable uncertainty about how much widening of the band would take place during the beam neutralization process. Furthermore the selection of a band of velocities of width
\( \Delta v \sim 10^2 \text{cm/sec} \) from an initial band of width \( \Delta v \sim 10^3 \text{cm/sec} \) means that the flux would be reduced by a factor on the order of \( 10^5 \). Thus if the initial flux were \( 10^{19} \text{atoms/cm}^2\text{-sec} \) the final flux would be on the order of \( 10^{14} \text{atoms/cm}^2\text{-sec} \).

Assuming \( 10^2 \) scatterings per \( \AA^2 \) are required to form an image, imaging times on the order of \( 10^3 \) seconds would be needed.

The conversion of a continuous liquid or droplet beam into a gas beam is at this point wholly unresolved. It is conceivable that deflecting the liquid off a diffracting surface might tend to produce the phase transition but it seems possible, if not likely, that it would introduce a large spread in the velocity distribution due to scattering and internal collisions.

References


Chapter 5

Summary and Conclusions

This has been a broad ranging, but not deep, analysis of the feasibility of producing coherent He-4 beams for use in microscopy. The choice to make it broad has been deliberate in an attempt to avoid overlooking crucial aspects of the problem.

The principal conclusion is that while all possibilities for the generation of suitably coherent beams of He-4 have not been ruled out, it will be very difficult to produce such beams. The essential requirements for a microscope employing holographic imaging are estimated to be:

- Particle Velocities $\sim 10^5 \text{cm/sec}$
- Velocity Bandwidth $\sim 10^{-2} \text{cm/sec}$
- Flux $\geq 10^{14} \text{atoms/cm}^2 \cdot \text{sec}$

The most difficult requirement to meet is probably the narrow velocity bandwidth.

Of the various methods of beam generation considered, liquid beams are probably too dense. Gas beams with un-filtered fluxes in the range of $10^9 - 10^{20} \text{ atoms/cm}^2 \cdot \text{sec}$ are feasible with the higher fluxes being limited mainly by condensation. The most likely method for generating gas beams makes use of the Kantrowitz-Grey nozzle. Thermal methods of beam generation might be possible but presently the fluxes are lower and the velocity bandwidths wider than those of the supersonic nozzle beams.

Any of the above beams will have to be filtered substantially to reduce the velocity bandwidth. The only possibly feasible method considered so far requires that the beam be ionized. Much further work needs to be done to explore all the problems presented by such an approach.

In this thesis the helium atom has been viewed as an essentially non-interacting hard sphere and the imaging
properties analyzed accordingly. In reality there is a Van der Waals attraction between it and the molecules being imaged. This leads to large effective cross sections at low energies. A fully quantum mechanical analysis of the imaging process is required in order to determine how effective an illuminant He-4 can be in principle.

The detection of low energy helium atoms is a very real one. A practical microscope will almost surely require a continuous detector such as a photographic emulsion. Means must therefore be explored for effecting such recording.

Finally the laser analog would appear to merit further study. The required lattice properties must first be established. If this should result in specifications that seem potentially realizable a search for suitable materials can follow.
Acknowledgment

I should like to thank Professor George Trammell for proposing this investigation and for critical advice along the way.
Appendix A: Photon Beams
A2.1a Bose Properties of Photons. Suppose we have a set of \( n \) identical atoms which can emit photons. Assume that we are concerned only with monoenergetic emissions. Suppose, further, that we wish to determine the probability of emission into photon states having a given average direction and contained within a small solid angle \( \delta \Omega \). Initially we assume that the photons are distinguishable, and that each is emitted into a slightly different state. Then if \( a_1 \) is the probability amplitude per unit solid angle for photon \( a \) to go into state \( 1 \), \( b_2 \) the probability amplitude per unit solid angle for photon \( b \) to go into state \( 2 \), etc., we have for the probability that all \( n \) photons go into their respective states,

\[
|a_1|^2 |b_2|^2 \ldots |n|^2 \delta \Omega, \delta \Omega_2 \ldots \delta \Omega_n.
\]

Next we assume \( |a_1|^2 = |a_2|^2 = |a_3|^2 = \ldots, |b_1|^2 = |b_2|^2 = |b_3|^2 = \ldots, \ldots \), which is valid if all \( n \) states lie in a small enough element of solid angle \( \Delta \Omega \), and we integrate to find the probability for all photons to enter any of the close lying states in \( \Delta \Omega \). Thus,

\[
(1) \quad |a_1|^2|b_1|^2 \ldots |n|^2 (\Delta \Omega)^n = |a_1b_1 \ldots n|^2 (\Delta \Omega)^n.
\]

Now, recognizing that photons are really indistinguishable, we must add up the various probability amplitudes before squaring. In this way we obtain the probability for the photons to occupy all close lying states. Thus

\[
\text{Sum of all combinations of photons } a, b, \ldots, N \text{ to go into states } 1, 2, \ldots, n.
\]

Assuming again that \( |a_1|^2 = |a_2|^2 = |a_3|^2 = \ldots, |b_1|^2 = |b_2|^2 = |b_3|^2 = \ldots, \ldots, \ldots \) since all states in \( \Delta \Omega \) are close lying, we can write

\[
|n! a_1b_1 \ldots n|^2 \delta \Omega, \delta \Omega_2 \ldots \delta \Omega_n.
\]

Integrating each \( \delta \Omega \) over \( \Delta \Omega \) and taking into account the repeated summings of the same surface elements, we obtain the total probability of \( n \) photons being emitted into the set of close lying states defined by \( \Delta \Omega \).
Comparing this result with equation (1), we see that the probability of entering the same state is enhanced by a factor of \( n! \). The indistinguishability property of Bose particles has had a pronounced effect.

The second Bose property, that of symmetry under exchange, has also had a pronounced effect. This property was used when we considered the sum of all combinations of photons \( a, b, c, \ldots, N \) into the states \( 1, 2, 3, \ldots, n \). For every term in the sum there is a single second term whose only difference is the exchange of two photons. The sum may then be grouped according to such pairs of terms. For Bose particles we may reorder the second term in each pair without affecting the sign. For Fermi particles such reordering reverses the sign of the second term in each pair so the sum vanishes. If photons were Fermi particles, then the probability of emitting more than one photon into a close lying set of states is virtually zero.

The Bose property of probability enhancement described above can be stated in a slightly different way. Suppose we know that there are already \( n \) photons in a set of close lying states and we wish to know the probability that the next photon to be emitted will enter the same set of close states. From equation (2) we can write

\[
P_{n+1} = \frac{1}{n!} |n!ab\ldots N|^2 (\Delta \Omega)^n = n! |ab\ldots N|^2 (\Delta \Omega)^n.
\]

where \( P_i \) is the probability of putting a single photon in an unoccupied set of close lying states. Since we know by hypothesis that there are \( n \) photons already in the set, \( P_n = 1 \); therefore the probability of putting yet another photon in the same set is \( n+1 \) times the probability of putting the
first photon in the set. Thus

\[ P_{n+1} = (n+1) P_n \]

We can write this basic result alternatively as follows. If we let |n\> be a state having \( n \) photons, |n+1\> be a state containing \( n+1 \) photons, and \( a^\dagger \) be the creation operator, then the probability amplitude for the state to increase its complement from \( n \) to \( n+1 \) photons is given by \( \langle n+1|a^\dagger|n\rangle \). Thus

\[ \langle n+1|a^\dagger|n\rangle = \sqrt{n+1} \langle n|n\rangle \]

Taking the hermitian conjugate of this and replacing \( n \) by \( n-1 \), we find that the probability amplitude for a state with \( n \) photons to lose one is

\[ \langle n-1|a_0|n\rangle = \sqrt{n} \langle n-1|n-1\rangle \]

where \( a_0 \) is the annihilation operator.

A2.1b Photon Statistics. Suppose we have a cavity containing \( N \) atoms, any one of which may be in the ground state |\( \alpha \rangle \) or an excited state |\( \beta \rangle \). Assume that the energy separation between states is \( \hbar \omega \) and that the numbers of atoms in the two states obey the Boltzmann equation

\[ \frac{N_\beta}{N_\alpha} = e^{-\beta \hbar \omega} \]

Assume further that the cavity also contains a number \( n \) of photons with energy \( \hbar \omega \). Since the atoms may absorb or emit photons at this frequency, we ask what is the relative probability of each? It is clear that

\[ P_{\text{absorp.}} = C n N_\alpha \]

and

\[ P_{\text{emission}} = C(n+1)N_\beta \]

where \( C \) is a constant of proportionality. At equilibrium we expect the two probabilities to be equal so

\[ n N_\alpha = (n+1)N_\beta \]

or

\[ n/(n+1) = N_\alpha / N_\beta \].
Substituting from equation (1),

\[ \frac{n}{n+1} = e^{-\alpha \omega} \]

and solving for \( n \), we obtain Planck's Law,

\[ n = \frac{1}{e^{\alpha \omega} - 1} \]  

(2)

An interesting thing about equation (2) is that not only is it a consequence of the Bose properties of photons but also it is a consequence of the transition properties of the atoms which are described by Boltzmann's equation.

Since the energy of each photon is \( \hbar \omega \), the total energy of an \( n \)-occupied photon state is

\[ E_n = n \hbar \omega \]

or

\[ E_n = \frac{\hbar \omega}{[e^{\alpha \omega} - 1]} \]

The number of photons is an integer so there are discrete steps of \( \hbar \omega \) between energy levels. Such a property is characteristic of the quantum mechanical harmonic oscillator and suggests that an electromagnetic wave can also be interpreted as a collection of elementary harmonic oscillators of different frequencies, the corresponding numbers of which determine the energy spectrum of the wave.

The state of a photon is determined by the wave equation and the boundary conditions. For periodic boundary conditions this leads to photon wave numbers

\[ k_\alpha = \frac{2\pi}{L_\alpha} n_\alpha ; \alpha = x, y, z \text{ and } n_\alpha \text{ integral}, \]

where \( L_\alpha \) is the spatial distance between nodes of the \( \alpha \)-component. The number triplet \((n_x, n_y, n_z)\) specifies the photon mode. In \( k \)-space the volume corresponding to a single mode is given by

\[ \frac{1}{2} \left( \frac{2\pi}{L_x} \right) \left( \frac{2\pi}{L_y} \right) \left( \frac{2\pi}{L_z} \right) = \frac{1}{2} \left( \frac{2\pi}{L} \right)^3 \]

The \( \frac{1}{2} \) factor is introduced to account for the two possible photon polarizations.
We wish now to determine the number of modes lying between $k$ and $k+\delta k$. The total volume in $k$-space is given by \( \frac{4}{3} \pi k^3 \). The number of modes corresponding to this $k$-volume is

\[
N_k = \frac{4}{3} \pi k^3 \frac{1}{V} \frac{(2\pi)^3}{V} = \frac{\theta}{3} \pi k^3 V/(2\pi)^3.
\]

The corresponding number of modes per unit volume in configuration space is

\[
\eta_k = \frac{N_k}{V} = \frac{\theta}{3} \pi k^3/(2\pi)^2.
\]

The corresponding number of modes in the range $k$ to $k+\delta k$ is

\[
\delta n_k = [\frac{\theta \pi k^2}{(2\pi)^3}] \delta k
\]

or

\[
\rho_k = \frac{\delta n_k}{\delta k} = \frac{\theta \pi k^2}{(2\pi)^2}
\]

where $\rho_k$ is the $k$-space mode density per unit volume of configuration space. By making use of the relation $k = \frac{2\pi}{c} \nu$, we can express this result as

\[
(3) \quad \rho_\nu = \frac{\delta n_\nu}{\delta \nu} = \frac{\theta \pi \nu^2}{c^3}
\]

where $\rho_\nu$ is the mode density in the frequency interval $\nu$ to $\nu+\delta \nu$ per unit volume in configuration space.

Finally, the number of modes in the frequency interval $\nu$ to $\nu+\delta \nu$ per unit volume is found by forming the product of equations (2) and (3),

\[
\eta_\nu = \frac{\delta n_\nu^2}{c^3} \frac{1}{e^{\lambda \nu} - 1}.
\]

The energy density in the same interval is obtained by multiplying by $\hbar \nu$,

\[
(4) \quad \rho(\nu) = \frac{\delta \pi \hbar \nu^3}{c^3} \frac{1}{e^{\lambda \nu} - 1}.
\]

This last equation is known as Planck's distribution and gives the electromagnetic energy density in configuration space as a function of frequency.
A2.1c Fluctuations. Suppose we have an electric field $E$ which is the result of randomly phased emissions from many identical atoms. Assume that the phase distribution is stationary so that the statistical properties of $E$ are independent of time. The energy of the field is proportional to $|E|^2$ which also is proportional to the number of photons $n$. If we let the contribution to the field by an individual emission be given by $E_j = E_j e^{i\phi_j}$, then the total energy of the field is proportional to

$$|E|^2 = E^2 \left( \sum_j e^{i\phi_j} \right)^* \left( \sum_k e^{i\phi_k} \right)$$

(1)

Averaging over time, all terms average to zero except those for which $j = k$. If there are $N$ emitting atoms, then

$$\langle |E|^2 \rangle_t = E^2 N.$$

To find the fluctuations in energy we need to find $\langle (|E|^2)^2 \rangle_t$. Starting with equation (1) we can write

$$|E|^2 = E^2 \left( N + \sum_{j \neq k} e^{i(\phi_k - \phi_j)} \right).$$

Squaring and taking the time average

$$\langle (|E|^2)^2 \rangle_t = \langle E^4 \left[ N^2 + 2N \sum_{j \neq k} e^{i(\phi_k - \phi_j)} + \sum_{j \neq k} e^{i(\phi_k - \phi_j)} \sum_{m \neq l} e^{i(\phi_m - \phi_l)} \right] \rangle_t$$

$$= E^4 \left[ N^2 + N(N-1) \right]$$

$$= E^4 [2N^2 - N].$$

Finally,

$$\langle (\Delta |E|^2)^2 \rangle_t = \langle (|E|^2)^2 \rangle_t - \langle |E|^2 \rangle_t^2$$

$$= E^4 \left[ 2N^2 - N^2 \right]$$

$$= E^4 N(N-1).$$

As $N$ becomes very large

$$\langle (\Delta |E|^2)^2 \rangle_t \approx E^4 N^2.$$
and the fractional fluctuation becomes

\[ \frac{\langle (d|E|^2)^2 \rangle}{\langle |E|^2 \rangle^2} \approx 1. \]

Thus, the results from both the particle and wave viewpoints are equivalent.

A2.2a Classical Theory of Temporal Coherence.⁴

It is an experimental fact that most detectors with which we observe signals have responses that are proportional to the time averages of the square of the sums of all electric fields at the points of detection. Under ordinary circumstances this leads to a recording proportional to the sum of the squares of the amplitudes of the individual signals, all phase information about the signals being lost in the process.

As an example, consider two electric fields having different frequencies being recorded by a single detector. We have for the time-averaged squared total field,

\[ \langle E^2 \rangle_t = \langle [E_1 + E_2]^2 \rangle_t = \langle E_1^2 \rangle_t + \langle E_2^2 \rangle_t + 2 \langle E_1 E_2 \rangle_t. \]

If \( E_1 = A_1 \cos(\omega_1 t + \varphi_1) \) and \( E_2 = A_2 \cos(\omega_2 t + \varphi_2) \), then

\[ \langle E^2 \rangle_t = \frac{1}{2} A_1^2 + \frac{1}{2} A_2^2 + 2 A_1 A_2 \langle \cos(\omega_1 t + \varphi_1) \cos(\omega_2 t + \varphi_2) \rangle_t \]

\[ = \frac{1}{2} A_1^2 + \frac{1}{2} A_2^2 + 2 A_1 A_2 \left\{ \langle \cos[\omega_1 t + (\varphi_1 + \varphi_2)] \rangle_t + \langle \cos[\omega_1 t - (\varphi_1 - \varphi_2)] \rangle_t \right\}. \]

Since usually the averaging times are long compared to \( t = \frac{1}{\nu_1 + \nu_2} \), the cosine terms vanish and we have

\[ \langle E^2 \rangle_t = \frac{1}{2} A_1^2 + \frac{1}{2} A_2^2. \]

If \( I \) is an intensity and if we let \( I_d = k \langle E^2 \rangle_t \) and \( I_1 = k \langle E_1^2 \rangle_t, I_2 = k \langle E_2^2 \rangle_t \), we can write

\[ I_d = I_1 + I_2 \]

where \( I_d \) is the detected intensity and \( I_1, I_2 \) are the
individual field intensities. $k$ is a constant of proportionality. We may generalize this result to a continuum of frequencies obtaining

$$I_d = \int I(\nu) d\nu.$$

Next we wish to obtain a quantitative measure of the temporal coherence between interfering signals in terms of their spectral compositions.

First, consider figure 1. Suppose that $S$ emits a single frequency $\nu$, and that we observe the interference pattern formed at $P$. Assuming that the detector records only field intensities, we have

$$I_d = I_1 + I_2 + 2 \sqrt{I_1 I_2} \cos 2\pi \nu \theta.$$

Since $\langle \cos 2\pi \nu (2\nu \theta) \rangle = 0$, we have

$$I_d = I_1 + I_2 + \frac{2}{c} \frac{\nu}{c} \cos 2\pi \nu \theta,$$

where $\nu(\tau)$ is a measure of the correlation between the two fields as recorded by the detector at $P$.

Now, suppose $S$ emits two different frequencies. By a procedure similar to the above we obtain

$$I_d = I_1 (1 + \cos 2\pi \nu_1 \tau) + I_2 (1 + \cos 2\pi \nu_2 \tau),$$

which, if $I_o = I_1 + I_2$, $P_1 = I_1 / I_o$, and $P_2 = I_2 / I_o$, may be written

$$I_d = I_o [1 + P_1 \cos 2\pi \nu_1 \tau + P_2 \cos 2\pi \nu_2 \tau],$$

or

$$I_d = I_o [1 + \gamma(\tau)],$$

with $\gamma(\tau) = P_1 \cos 2\pi \nu_1 \tau + P_2 \cos 2\pi \nu_2 \tau$.

These results are readily generalized to a continuum.
of frequencies emitted by $S$, and we have
$$I_d = \int_0^\infty I(\nu) \left[ 1 + \cos 2\pi \nu \tau \right] d\nu,$$
or
\begin{equation}
I_d = I_0 \left[ 1 + \nu(\tau) \right]
\end{equation}
where
\begin{equation}
I_0 = \int_0^\infty I(\nu) d\nu
\end{equation}
and
\begin{equation}
\nu(\tau) = \int P(\nu) \cos 2\pi \nu \tau d\nu
\end{equation}
with
\begin{equation}
P(\nu) = \frac{I(\nu)}{I_0}.
\end{equation}

We may use the equation for $\nu(\tau)$ to get a quantitative measure of coherence by relating it to the interference fringes produced by the interferometer, if any. First, we consider the case where $P(\nu)$ is non-zero only for the approximate frequency range, $\nu = \nu_0 \pm \frac{\Delta \nu}{2}$. We may then write
$$\nu(\tau) = \int_{\nu_0 - \frac{\Delta \nu}{2}}^{\nu_0 + \frac{\Delta \nu}{2}} P(\nu) \cos 2\pi \nu \tau d\nu
\begin{equation}
= \int_{\nu_0 - \frac{\Delta \nu}{2}}^{\nu_0 + \frac{\Delta \nu}{2}} P(\nu) \cos 2\pi \nu \tau d\nu,
\end{equation}
or if
$$P(\nu) = D(\nu - \nu_0) = D(\mu)
\begin{equation}
then,
\begin{align*}
\nu(\tau) &= \int_{-\nu_0}^{\nu_0} D(\mu) \cos 2\pi (\nu_0 + \mu) \tau d\mu \\
&= \int_{-\infty}^{\infty} D(\mu) \cos 2\pi (\nu_0 + \mu) \tau d\mu \\
&= \cos 2\pi \nu_0 \tau \int_{-\infty}^{\infty} D(\mu) \cos 2\pi \mu \tau d\mu
\end{align*}
\begin{equation}
= \sin 2\pi \nu_0 \tau \int_{-\infty}^{\infty} D(\mu) \sin 2\pi \mu \tau d\mu.
\end{equation}
If \( C(v) = \int D(\mu) \cos 2\pi \mu v d\mu \) and \( S(v) = \int D(\mu) \sin 2\pi \mu v d\mu \), then

(5) \( Y(\tau) = C(v) \cos 2\pi \nu \tau + S(v) \sin 2\pi \nu \tau \)

or

(6) \( Y(\tau) = U(\tau) \cos [2\pi \nu \tau + \phi(\tau)] \)

with

\[
U(\tau) = \left[ C^2(v) + S^2(v) \right]^{1/2}
\]

and

\[
\phi(\tau) = \tan^{-1} \left( \frac{S(v)}{C(v)} \right).
\]

Since \( |Y(\tau)| \leq 1 \), it follows from the above that \( 0 \leq U(\tau) \leq 1 \).

We now observe that \( Y(\tau)_{\max} = U(\tau) \) and \( Y(\tau)_{\min} = -U(\tau) \).

If \( I_d = I_0 \left[ 1 + Y(\tau) \right] \) and if we define the Visibility of fringes by

(7) \( V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \),

we then obtain,

\[
V = \frac{I_0 [1 + U(\tau)] - I_0 [1 - U(\tau)]}{I_0 [1 + U(\tau)] + I_0 [1 - U(\tau)]}
\]

or

(8) \( V = U(\tau) \).

This result is of practical importance because for the band-limited signals in which we are interested, we have a quantitative relationship between coherence and fringe visibility.

A2.2b Wiener-Khintchine Theorem.\(^4,5\) The Wiener-Khintchine theorem states that

\[
\frac{\langle E(\tau) E(\tau^*) \rangle}{\langle E(\tau)^2 \rangle} = \int P(\omega) \cos 2\pi \nu \tau d\nu = Y(\tau)
\]

where

\[
\frac{\langle E(\tau) E(\tau^*) \rangle}{\langle E(\tau)^2 \rangle}
\]

is called the autocorrelation function of \( E(\tau) \). The theorem expresses the fact that the normalized autocorrelation function of a signal is the cosine Fourier transform of its spectral distribution function. In other words the Wiener-Khintchine theorem establishes the connection between the
statistical properties of a signal and its spectral properties. Furthermore, it provides an operational method for obtaining the power spectrum of a signal and/or obtaining a quantitative measure of its temporal coherence. The latter is given practical meaning through the fringe visibility measure.

To prove the theorem we begin with the normalized autocorrelation function and express the fields in terms of their Fourier transforms. In order to ensure convergence of the integrals we time-truncate the fields. Thus the normalized autocorrelation function, N.A.F., is given by

\[
N.A.F. = \lim_{T \to \infty} \left\{ \frac{\int_{t_0}^{t_0+T} E(t') \int_{t_0}^{t_0+T} E(t') \, dt'}{\int_{t_0}^{t_0+T} E(t')^2 \, dt'} \right\}
\]

where and \( \hat{E}_S(v) \) and \( \hat{E}_S(v') \) are the Fourier transforms of the time-truncated fields. Rewriting the above, we get

\[
N.A.F. = \lim_{T \to \infty} \left\{ \frac{\int_{t_0}^{t_0+T} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{E}_S(v) \hat{E}_S(v') e^{2\pi i (v-v') t'} \, dv \, dv' \, dt'}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{E}_S(v) \hat{E}_S(v') e^{2\pi i (v-v') t} \, dv \, dv'} \right\}
\]

\[
= \lim_{T \to \infty} \left\{ \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{E}_S(v) \hat{E}_S(v') e^{2\pi i (v-v') t} \, dv \, dv'}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{E}_S(v) \hat{E}_S(v') e^{2\pi i (v-v') t} \, dv \, dv'} \right\}
\]
Letting \( u = u + u' \) and using the Dirichlet formula that
\[
\lim_{T \to \infty} \left[ \sin \pi u T / \pi u \right] = \delta(u),
\]
we obtain
\[
\text{N.A.F.} = \lim_{T \to \infty} \left\{ \frac{\int_{-\infty}^{\infty} du' e^{-2\pi i u' \nu} \hat{E}_s(u') \left[ \int_{-\infty}^{\infty} du \hat{E}_s(u-u') e^{2\pi i u' T / \pi u} \right]}{\int_{-\infty}^{\infty} du' \hat{E}_s(u') \left[ \int_{-\infty}^{\infty} du \hat{E}_s(u-u') e^{2\pi i u' T / \pi u} \right]} \right\}
\]
\[
= \frac{\int_{-\infty}^{\infty} du' e^{-2\pi i u' \nu} \hat{E}_s(u') \hat{E}_s(-u')}{\int_{-\infty}^{\infty} du' \hat{E}_s(u') \hat{E}_s(-u')}.
\]
Now \( \hat{E}_s(u')^* = \hat{E}_s(-u') \), so this last expression can be written
\[
\text{N.A.F.} = \frac{2 \int_{0}^{\infty} du' \hat{E}_s(u') \hat{E}_s(u')^* \cos 2\pi u' \nu}{2 \int_{0}^{\infty} du' \hat{E}_s(u') \hat{E}_s(u')^*}.
\]
Letting \( I(u) = k \lim_{S \to \infty} \frac{1}{S} \hat{E}_s(u') \) we can write
\[
\text{N.A.F.} = \int_{0}^{\infty} I(u) \cos 2\pi u' \nu \frac{d\nu'}{\int_{0}^{\infty} I(u) d\nu'} = \int_{0}^{\infty} P(u) \cos 2\pi u' \nu d\nu'.
\]
We have thus shown that
\[
(1) \quad \gamma(t) = \int_{0}^{\infty} P(u) \cos 2\pi u' \nu d\nu' = \frac{\langle E(t') E(t'-t) \rangle_t}{\langle E(t')^2 \rangle_t}.
\]

A2.2c Classical Theory of Spatial Coherence. 4

In order to discuss spatial coherence, consider the Young's double-slit experiment shown in figure 1. The source \( S' \) emits monochromatic light of frequency \( \nu \). If \( E_a(p) \) and \( E_b(p) \) are the fields at \( P \) from slits \( a \) and \( b \)
respectively, then the intensity at \( P \) is given by

\[
I = I_a + I_b + 2\sqrt{I_a I_b} \cos 2\pi n L \frac{(\theta' - \theta)}{c}
\]

\[
= I_a \left[ 1 + \cos 2\pi n L \frac{\Delta \theta_a}{c} \right] + I_b \left[ 1 + \cos 2\pi n L \frac{\Delta \theta_b}{c} \right]
\]

where \( \Delta \theta = \theta' - \theta \), \( I_1 = I_a + I_b \), and where we have assumed \( I_a = I_b \).

Next suppose we have two independent sources \( S \) and \( S' \), each of which emits signals of the same frequency \( \nu \), but with a random phase relationship. The intensity of the radiation at \( P \) can be derived by similar arguments to the above and is found to be

\[
I = I_a \left[ 1 + \cos 2\pi n L \frac{\Delta \theta_a}{c} \right] + I_b \left[ 1 + \cos 2\pi n L \frac{\Delta \theta_b}{c} \right]
\]

where \( \Delta \theta_a = \theta' - \theta_1 \) and \( \Delta \theta_b = \theta' - \theta_2 \).

As in the case of temporal coherence we may generalize this result for a continuum of sources. In this case each element of source contributes

\[
dI = I(\theta) \left[ 1 + \cos 2\pi n L \frac{(\theta' - \theta)}{c} \right] d\theta
\]

so that the total intensity received at \( P \) is given by

\[
I(\theta) = \int_{\theta_a}^{\theta_b} I(\theta) \left[ 1 + \cos 2\pi n L \frac{(\theta' - \theta)}{c} \right] d\theta
\]

This equation may be rewritten as

\[
I(\theta') = I_{tot} (1 + \gamma z)
\]
where

\[ I_{\text{tot}} = \int_{\Theta_a}^{\Theta_b} I(\Theta) \, d\Theta \]

and

\[ Y_{12}'(\theta') = \int_{\Theta_a}^{\Theta_b} \frac{I(\Theta) \cos \left[ 2\pi \nu \frac{(\theta' - \theta)}{c} \right] \, d\Theta}{I_{\text{tot}}}. \]

In order to extend these results to the case where the sources emit signals over a finite frequency band, we multiply equation (3) by the spectral distribution function, \( P(\nu) \), and integrate over all frequencies. Thus

\[
Y_{12}'(\theta') = \int_{\nu=0}^{\infty} P(\nu) \left[ \int_{\Theta_a}^{\Theta_b} I(\Theta) \cos \left[ 2\pi \nu \frac{(\theta' - \theta)}{c} \right] \, d\Theta \right] \, d\nu / I_{\text{tot}}
\]

\[ = \int_{\Theta_a}^{\Theta_b} I(\Theta) \, d\Theta \int_{\nu=0}^{\infty} P(\nu) \cos \left[ 2\pi \nu \frac{(\theta' - \theta)}{c} \right] \, d\nu / I_{\text{tot}} \]

Recalling equation (A2.2a-3), we may write this last result as

\[
Y_{12}'(\theta') = \frac{\int_{\Theta_a}^{\Theta_b} I(\Theta) \, d\Theta}{I_{\text{tot}}} \cdot Y(\frac{t'(\theta' - \theta)}{c})
\]

We next wish to relate the preceding formulation of \( Y_{12}'(\theta') \) to the statistical properties of the signals producing the fringes. To do this consider the electric field at point \( P \) in figure (2). This will be the sum of the contributions from \( S_i \) and \( S_2 \) which are, respectively, \( E_i(t' - \frac{1S_iP_i}{c}) \) and \( E_2(t' - \frac{1S_2P}{c}) \). Defining \( \tau \) by

\[ \tau = \frac{1S_iP_i - 1S_2P}{c} \approx \frac{\theta'}{c} \]

and introducing the time variable \( t = t' - \frac{1S_iP_i}{c} \) which represents the arrival time at \( S_i \), we may express the field at \( P \) by
Squaring this and taking the time average we obtain

\[ \langle E_\rho^2 \rangle_t = \langle E_1(t)^2 \rangle_t + \langle E_2(t-t)^2 \rangle_t + 2 \langle E_1(t)E_2(t-t) \rangle_t. \]

Since \( \langle E_\rho(t)^2 \rangle_t = \langle E_\rho(t-t)^2 \rangle_t \), this can be written

\[ \langle E_\rho^2 \rangle_t = \langle E_1^2 \rangle_t + \langle E_2^2 \rangle_t + 2 \sqrt{\langle E_1^2 \rangle_t \langle E_2^2 \rangle_t} \cdot \frac{\langle E_1(t)E_2(t-t) \rangle_t}{\sqrt{\langle E_1^2 \rangle_t \langle E_2^2 \rangle_t}}. \]

The intensities are proportional to the time averages of the square of the fields so we may write

\[ I = I_1 + I_2 + 2 \sqrt{I_1I_2} \gamma_{12}, \]

where

\[ \gamma_{12} = \frac{\langle E_1(t)E_2(t-t) \rangle_t}{\sqrt{\langle E_1^2 \rangle_t \langle E_2^2 \rangle_t}}. \]

Equation (3) defines the normalized correlation function or cross-correlation function of \( E_1 \) and \( E_2 \).

Now suppose the fields at \( S_1 \) and \( S_2 \) are given by

\[ E_1(t) = \sum_j E_{1j}(t) \]

and

\[ E_2(t-t) = \sum_j E_{2j}(t-t) \]

where the index \( j \) represents the \( j \)th source. If the averaging time is long compared to the coherence time of the signals, it follows that

\[ \langle E_{1j}(t)E_{2k}(t-t) \rangle_t = \delta_{jk} \]

so that

\[ \gamma_{12} = \frac{\langle E_{1j}(t) \rangle \langle E_{2j}(t-t) \rangle_t}{\langle E_{1j}(t)^2 \rangle_t}. \]

For a distant source \( j \), it is evident from figure 1 that

\[ E_{1j}(t) = E_{2j}(t-L \frac{\theta_j}{c}) \]

so that

\[ \gamma_{12} = \frac{\langle E_{2j}(t-L \frac{\theta_j}{c}) \rangle \langle E_{2j}(t-t+L \frac{\theta_j}{c}) \rangle_t}{\langle E_{2j}(t)^2 \rangle_t}. \]

Changing variables we obtain

\[ \gamma_{12} = \frac{\langle E_{2j}(t) \rangle \langle E_{2j}(t-t-L \frac{\theta_j}{c}) \rangle_t}{\langle E_{2j}(t)^2 \rangle_t}. \]

By equation (A2.2b-1) we have

\[ \gamma_j(t-L \frac{\theta_j}{c}) = \frac{\langle E_{2j}(t) \rangle \langle E_{2j}(t-(t-L \frac{\theta_j}{c})) \rangle_t}{\langle E_{2j}(t)^2 \rangle_t}. \]
If the source distribution is continuous, the summation is replaced by an integration over the angle variable \( \theta \), and we then have
\[
\gamma_{\text{12}} = \int_{\theta_a}^{\theta_b} \int \frac{v^2}{k} \left< \left[ E_{\text{j}}(\hat{e}) \right]^2 \right> \frac{I(\theta) d\theta}{I_{\text{tot}}}.
\]
Writing
\[
I_{\text{tot}} = \int I(\theta) d\theta
\]
and noting that
\[
\tau = L \frac{c}{v},
\]
we obtain
\[
\gamma_{\text{12}} = \int \frac{I(\theta) v^2}{I_{\text{tot}}} (\frac{1}{\tau^2}) d\theta.
\]
This last equation is identical to equation (2) and demonstrates the equivalence of the spectral and statistical formulations of the normalized correlation function.

**A2.3a Helmholtz-Kirchoff Theorem.** The fundamental theorem upon which the whole of scalar diffraction theory is based is the Helmholtz-Kirchoff Theorem. In words, this theorem states that the electric field at some observation point is given by an integration of the electric field and its normal derivatives over a closed boundary surface surrounding the observation point but excluding all sources.

To derive the theorem we begin with the wave equation
\[
\nabla^2 \Psi(\xi, t) - \frac{1}{c^2} \frac{\partial^2 \Psi(\xi, t)}{\partial t^2} = 0,
\]
where \( \Psi(\xi, t) \) is some field, and \( c \) is the velocity of propagation. If we now assume a time-dependence of the form,
\[
\Psi(\xi, t) = \Psi(\xi) e^{i\omega t},
\]
then we may separate the wave equation and obtain the Helmholtz equation,
\[
\nabla^2 \Psi(\xi) + k^2 \Psi(\xi) = 0
\]
provided that \( k = \frac{\omega}{c} \).

Next let us introduce a second field, \( \phi(\xi) \), and write
the Helmholtz equation for it,

\[ \nabla^2 \phi(\mathbf{r}) + k^2 \phi(\mathbf{r}) = 0, \]

where the same time-dependence as before is assumed.

Multiplying equation (1) on the left with \( \phi(\mathbf{r}) \) and equation (2) on the left with \( \Psi(\mathbf{r}) \), we obtain after subtraction,

\[ \Psi(\mathbf{r}) \nabla^2 \phi(\mathbf{r}) - \phi(\mathbf{r}) \nabla^2 \Psi(\mathbf{r}) = 0. \]

Integrating and applying the divergence theorem we get,

\[ \iint_{\Sigma} \left[ \Psi(\mathbf{r}') \nabla \phi(\mathbf{r}') - \phi(\mathbf{r}') \nabla \Psi(\mathbf{r}') \right] \cdot d\mathbf{s}', \]

where \( \Sigma \) is a surface that surrounds the observation point and excludes any sources.

Now let us choose \( \phi(\mathbf{r}) = E(\mathbf{r}) \) where \( E(\mathbf{r}) \) is one component of the electric field. Also let \( \Psi(\mathbf{r}) = e^{-ik|\mathbf{r}'-\mathbf{r}_0|}/|\mathbf{r}'-\mathbf{r}_0| \).

Both functions satisfy the wave equation for all points within the volume bounded by \( \Sigma \) except at \( \mathbf{r} = \mathbf{r}_0 \), where \( \Psi(\mathbf{r}) \) becomes singular. To handle the singularity, write \( \Sigma = \Sigma + \Sigma_o \), where \( \Sigma_o \) is a small spherical surface of radius \( \epsilon \) surrounding the point \( \mathbf{r}_0 \) and \( \Sigma \) is the remainder of \( \Sigma \).

Substituting into equation (3) we obtain

\[ 0 = \iint_{\Sigma} \left[ \Psi(\mathbf{r}') \nabla \phi(\mathbf{r}') - \phi(\mathbf{r}') \nabla \Psi(\mathbf{r}') \right] \cdot d\mathbf{s}', \]

As \( \epsilon \to 0 \), the second and third terms vanish and the last becomes \(-4\pi E(\mathbf{r}_0)\). We then get

\[ \iint_{\Sigma} \left[ \frac{e^{-ik|\mathbf{r}'-\mathbf{r}_0|}}{|\mathbf{r}'-\mathbf{r}_0|} \nabla \phi(\mathbf{r}') - \phi(\mathbf{r}') \frac{e^{-ik|\mathbf{r}'-\mathbf{r}_0|}}{|\mathbf{r}'-\mathbf{r}_0|} \nabla \psi(\mathbf{r}') \right] \cdot d\mathbf{s}' = 4\pi E(\mathbf{r}_0), \]

which embodies the Helmholtz-Kirchoff Theorem stated in the first paragraph.

A2.3b Fresnel-Kirchoff Integral. In order to apply the Helmholtz-Kirchoff theorem, we must relate the surface
more specifically to actual geometries and make certain assumptions about the fields on the boundaries. We shall be concerned with a single component \( \mathbf{E}(\xi) \) of the electric field. Suppose that \( \mathbf{E} = \mathbf{E}_a + \mathbf{E}_b + \mathbf{E}_c \) as shown in figure 1. \( \Sigma_a \) is that part of the surface corresponding to the diffracting aperture, \( \Sigma_b \) that part corresponding to the mask, and \( \Sigma_b \) that part corresponding to the remainder of the bounding surface. As \( R \) becomes very large, we assume that \( \mathbf{E}(\xi) \) and \( \frac{\partial \mathbf{E}(\xi)}{\partial n} \) on \( \Sigma_b \) approach zero. Further we assume that \( \mathbf{E}(\xi) = \frac{\partial \mathbf{E}(\xi)}{\partial n} = 0 \) on \( \Sigma_a \), and that \( \frac{\partial \mathbf{E}(\xi)}{\partial n} = \frac{\partial \mathbf{E}_{inc}(\xi)}{\partial n} \) and \( \mathbf{E}(\xi) = \mathbf{E}_{inc}(\xi) \) on \( \Sigma_a \).

Substituting into equation (A2.3a-4), we obtain

\[
E(\xi) = \frac{1}{4\pi} \int \int_{\Sigma_a} \left[ \frac{e^{ik|\xi'-\xi|}}{|\xi'-\xi|} \frac{\partial \mathbf{E}_{inc}(\xi')}{\partial n} - \mathbf{E}_{inc}(\xi') \frac{\partial e^{-ik|\xi'-\xi|}}{\partial n} \right] d\sigma',
\]

In order to put this expression into a more wieldy form, consider figure 2. We can write

\[
\frac{\partial \mathbf{E}_{inc}(\xi')}{\partial n} = \mathbf{n} \cdot \frac{\partial \mathbf{E}_{inc}(\xi')}{\partial \xi'} = -\cos \theta \frac{\partial \mathbf{E}_{inc}(\xi')}{\partial \rho}.
\]

and

\[
\frac{\partial}{\partial n} \left[ \frac{e^{ik|\xi'-\xi|}}{|\xi'-\xi|} \right] = \frac{\rho}{\rho'} \frac{\partial e^{ik\rho}}{\rho} = -\cos \theta' \left( \frac{k}{\rho} \right) e^{ik\rho}.
\]

Assuming that \( \mathbf{E}_{inc}(\xi') \) has the local dependence \( \mathbf{E}_{inc}(\xi') = \text{Const.} e^{-ik\rho} \), we obtain

\[
\frac{\partial \mathbf{E}_{inc}(\xi')}{\partial n} = \frac{\partial \mathbf{E}_{inc}(\xi')}{\partial \rho} = ik \mathbf{E}_{inc}(\xi').
\]

Combining the above results, neglecting the term involving \( \frac{\cos \theta'}{\rho} \), and substituting into equation (1), we obtain

\[
E(\xi) = \frac{i}{\lambda} \int \int_{\Sigma_a} \frac{e^{ik\rho}}{\rho} \mathbf{E}_{inc}(\xi') \left[ \frac{\cos \theta + \cos \theta'}{\lambda} \right] d\sigma'.
\]
which is the Fresnel-Kirchoff Integral. It states that the scalar field $E(r)$ can be found by integrating the incident field over the aperture, taking due consideration of the geometries involved. If the source and observation points are both close to the optical axis, equation (2) may be approximated by

$$E(r) = \frac{i}{\lambda} \int_{\mathcal{S}} \frac{e^{-ik|\mathbf{r}' - \mathbf{r}|}}{|\mathbf{r}' - \mathbf{r}|} E_{\text{inc}}(\mathbf{r}') d\sigma'$$

where $|\mathbf{r}' - \mathbf{r}|$ is the distance between the point $\mathbf{r}'$ in the entrance aperture and $\mathbf{r}$, the observation point.

A2.3c Fresnel Transformation. By repeated application of the Fresnel-Kirchoff diffraction integral, we can relate the electric field in one transverse plane to that in another such plane. In applying the integral it is convenient to make the so-called parabolic approximation where spherical waves are replaced by parabolic ones. This limits the rays to relatively small angles from the axis of the optical system. The approximation is equivalent to the Gaussian approximation of geometrical optics, and the resultant diffraction integral is called the Fresnel transformation.

Consider figure 1 and, using the Fresnel-Kirchoff integral, express the electric field at $P_0$ in terms of the field in the plane $\mathcal{P}$. Thus,

$$E(P_0) = \frac{i}{\lambda} \int_{\mathcal{S}} \frac{e^{ikR}}{R} E(x_s, y_s, z_s) dx_s dy_s$$

where the factor $\cos \theta + \cos \theta'$ has been replaced by unity since the angles are small by hypothesis.

Next we make the parabolic approximation which consists of expanding $R$ in terms of the coordinates and then neglecting higher order terms. Thus
\[ R = \left[ (x_0 - x_s)^2 + (y_0 - y_s)^2 + (z_0 - z_s)^2 \right]^{1/2} \]
\[ = (z_0 - z_s) \left[ 1 + \frac{(x_0 - x_s)^2}{(z_0 - z_s)^2} + \frac{(y_0 - y_s)^2}{(z_0 - z_s)^2} \right]^{1/2} \]
\[ = (z_0 - z_s) \left[ 1 + \frac{(x_0 - x_s)^2}{2(z_0 - z_s)^2} + \frac{(y_0 - y_s)^2}{2(z_0 - z_s)^2} \right] \]

This result is valid provided \(|z_0 - z_s| \gg |x_0 - x_s|\) and \(|z_0 - z_s| \gg |y_0 - y_s|\)

It may be slightly simplified if we let \(L = z_0 - z_s\), so

\[ R \approx L + \frac{(x_0 - x_s)^2}{2L} + \frac{(y_0 - y_s)^2}{2L}. \]

Substituting this into equation (1), we get

\[ (2) \quad E(r) = \frac{i}{\lambda} \frac{e^{ikr}}{L} \int \int_{S} E(x, y, z) \frac{e^{-ikr}}{r} \frac{e^{ikr}}{r} E(x', y', z') dx'dy' \]

where we have replaced \(R\) by \(L\) in the denominator since \(E(r)\) is not sensitive to the error introduced under the conditions of the approximation. Equation (2) is the Fresnel transformation, and it relates the scalar field in plane \(S\) to that in the observation plane \(O\).

**A2.3d Alternate Form for Fresnel Transformation.**

The Fresnel-Kirchoff Integral, equation (A2.3b-2), is

\[ E(x) = \frac{i}{\lambda} \int \int_{S} \frac{e^{-ikr}}{r} E_{inc}(x', z') \left[ \frac{\cos \theta + \cos \theta'}{2} \right] d\sigma' \]

If \(\theta\) and \(\theta'\) are small we may approximate this by

\[ (1) \quad E(x) = \frac{i}{\lambda} \int \int_{S} \frac{e^{-ikr}}{r} E_{inc}(x', z') d\sigma' \]

---

Figure 1
Now consider figure 1 which for simplicity shows a two dimensional system. Let us identify $P'$ with $\rho$, $\eta$ with $x'$, $z'$ with $r$, and $d\eta$ with $d\sigma'$. Equation (1) then becomes

$$E(x') = \frac{i}{\lambda} \int_{-\eta_{\text{max}}}^{\eta_{\text{max}}} \left[ \frac{e^{-ik\rho}}{\rho} \right] E_{\text{inc}}(\eta) \, d\eta.$$  

If $\rho_0$ corresponds to $\rho$ when $P$ is at 0, then we can write approximately

$$E(x') \approx \frac{i}{\lambda} \int_{-\eta_{\text{max}}}^{\eta_{\text{max}}} \left[ \frac{e^{-ik\rho}}{\rho_0} \right] E_{\text{inc}}(\eta) \, d\eta,$$

or

$$E(x') = \frac{ie^{-ik\rho_0}}{\lambda \rho_0} \int_{-\eta_{\text{max}}}^{\eta_{\text{max}}} e^{ik\eta} E_{\text{inc}}(\eta) \, d\eta.$$  

However,

$$\eta \cos \alpha = \eta u = \text{on}$$

so

$$E(x') \approx \frac{ie^{-ik\rho_0}}{\lambda \rho_0} \int_{-\eta_{\text{max}}}^{\eta_{\text{max}}} e^{ik\eta} E_{\text{inc}}(\eta) \, d\eta.$$  

Generalizing to the three dimensional case we can write

$$E(P') = \frac{ie^{-ik\rho_0}}{\lambda \rho_0} \iiint_{\mathbf{z}_0} E_{\text{inc}}(\eta, \xi) e^{ik(\eta \eta + \xi \varsigma)} \, d\xi \, d\eta \, d\varsigma,$$

where $\nu = \cos \beta$ direction cosine with respect to $\xi$. Replacing $ie^{-ik\rho_0}/\lambda \rho_0$ by $A$, we have finally

$$E(P') = A \iiint_{\mathbf{z}_0} E_{\text{inc}}(\eta, \xi) e^{ik(\eta \eta + \xi \varsigma)} \, d\xi \, d\eta \, d\varsigma.$$  

which is the desired alternate form for the Fresnel transformation.

**A2.3e Basic Principles of Lensless Holography**

The feasibility of holography derives from the fact that if we know the electric field on some plane in space, we can by suitable transformation determine the electric field on another plane; for example, the image plane in an
optical system. The transformation is just the Fresnel transformation or an adaptation of it for the particular circumstances.

The basic problem is to find a way of recording on a given plane both the amplitudes and phases of the incident electric field. It is not sufficient to place an array of detectors or a photographic plate in the plane because both are sensitive only to time averages of the electric field where the averages extend over long enough time intervals to destroy the phase information.

The way the phase problem is solved is by recording an interference pattern between a reference beam and the object beam. For such an interference pattern to exist, we know from the theory of coherence, that there must be a definite phase relationship between every point in the reference beam and a corresponding point in the object beam; a relationship which must be maintained over times at least as great as the averaging time of the detectors.

To show that we can indeed reconstruct the object beam from the above described two dimensional interference pattern, first consider figure 1. For simplicity we shall consider only a one dimensional field at the recording plane. Assume the incident beam is a plane wave and that the object is a pinhole in an otherwise opaque mask. Assume further that the reference beam is off-axis and plane as shown in the figure. The amplitude of the reference wave on the photographic plate is given by $E_r(\kappa) = \text{Re} [A_r e^{-i\alpha \kappa}]$ where $\alpha$ is a constant. The amplitude of the object wave is approximately $E_o(\kappa) = \text{Re} [A_o e^{ikz}]$ where the Gaussian
approximation has been used. In both representations we have dropped the time dependence, since in the detection process, it averages out. We may now write for the recorded field information

\[ \langle|E(t)|^2 \rangle = \frac{1}{2} A_r^2 + \frac{1}{2} A_0^2 + \text{Re} [A_r A_0 e^{i(ux + k \frac{L}{2})}] + \text{Re} [A_r A_0 e^{-i(ux + k \frac{L}{2})}] . \]

Next we assume that the transmittance of the developed photographic plate is proportional to some power of the field intensity; that is

\[ T(x) \propto |I(x)|^p \propto \langle |E(t)|^2 \rangle . \]

Assuming that \( A_r \gg A_0 \) and expanding equation (1) in terms of \( \chi \), we get approximately

\[ T(x) \propto \left[ 1 + \chi |\frac{A_0}{A_r}| \text{Re} e^{i(ux + k \frac{L}{2})} + \chi |\frac{A_r}{A_0}| \text{Re} e^{-i(ux + k \frac{L}{2})} \right] . \]

To reconstruct the original wave we illuminate the hologram with a new incident beam. The transmitted beam can then be described by

\[ E_{\text{trans}} = E_{\text{inc}} \cdot T(x) . \]

Without specifying further the form of \( E_{\text{inc}} \), we can write

\[ E_{\text{trans}} \propto \left[ 1 + \chi |\frac{A_0}{A_r}| \text{Re} e^{i(ux + k \frac{L}{2})} + \chi |\frac{A_r}{A_0}| \text{Re} e^{-i(ux + k \frac{L}{2})} \right] E_{\text{inc}} . \]

This equation can be interpreted as follows. Part of \( E_{\text{inc}} \) passes through the hologram unaltered except possibly in absolute amplitude. If we recognize that \( e^{\pm iux} \) describes the phase transforming action of an upward(+) or downward(-) deflecting prism and that \( e^{\pm i k \frac{L}{2}} \) describes the phase transforming action of a concave(+)/convex(-) thin lens, then we see that \( E_{\text{inc}} \) is split into two additional beams; one that is deflected upward and that forms a virtual image to the left of the hologram, and the other that is deflected downward and that forms a real image to the right of the

---

Figure 2
hologram. The various beams are shown in figure 2.

**A2.3f Abbé Sine Condition.** One of the aberrations or departures from simple first order optical theory is called coma. An off-axis object point will not ordinarily image to a point but rather will produce a comet-shaped image, the size of which varies linearly with the off-axis image distance. In microscopes this can be corrected for, and when it is, the instrument is said to obey the Abbé sine condition.

To derive the sine condition, consider figure 1. For simplicity assume the indices of refraction are unity on either side of the optical system. Assume further that the system has been so designed that each set of parallel rays forms a point image in the focal plane $F'$; that is, the system has been corrected for coma, at least to first order. If such is the case, then the path lengths corresponding to rays 1 and 2 must be equal to first order.

First compare rays 1 and 3. They are equal to $R'$ and hence the difference is given by

$$|1-3| = \sqrt{x'^2 + h'^2} - x' \sim \frac{h'^2}{2x'} = O(h'^2).$$

Now, compare rays 3 and 4. These correspond to imaging of the on-axis object point $P'$, so they are equal. Thus,

$$|3-4| = 0.$$

[Diagram of optical system with rays labeled 1, 2, 3, 4, $P$, $h$, $F'$, $R'$, $Q'$, $P'$, $X'$, $P_1$, $P_2$, $R_1$, $R_2$.]

---

Figure 1
Next, compare rays 2 and 4. \( P_2Q' \) and \( MQ' \) are equal by hypothesis. Furthermore, \( QP' - QN' = (P_2N')^2/2Q'N' = O(h^4) \). Therefore,

\[
(3): \quad |2 - 4| = \frac{1}{2}P_2N + O(h^4).
\]

Combining (1), (2), and (3), we obtain

\[
|2 - 1| = \frac{1}{2}P_2N + O(h^4).
\]

Since \( N'P_2' = h'S\sin\Theta' \) and \( P_1N = hS\sin\Theta \) we can write to first order

\[
|2 - 1| = h'S\sin\Theta' - h\sin\Theta.
\]

Finally, since rays 1 and 2 are equal to first order, it follows that

\[
(4): \quad h\sin\Theta = h'S\sin\Theta'.
\]

This is the Abbé sine condition which holds if the optical system is corrected to first order for coma.

**A2.3g Rayleigh Resolution Criterion.** Fundamentally, the Rayleigh resolution criterion expresses the resolution limit due to the combination of a non-zero wavelength and a non-infinite optical aperture. We can derive this criterion from the alternate form of the Fresnel transformation.

Consider figure 1. From equation (A2.3d-2), we have
for the amplitude at \( P' \)

\[
E(P') = A \iint_{\mathbb{R}^2} e^{i k (u \eta + v \xi)} \, d\eta d\xi ,
\]

when \( E_{\text{inc}}(\eta, \xi) = 1 \). Letting \( \eta = a \cos \Theta, \xi = a \sin \Theta, u = \alpha \cos \Theta', v = \alpha \sin \Theta' \), we may write

\[
u \eta + v \xi = a \alpha \cos (\Theta - \Theta') .
\]

Redefining the origin of \( \Theta' \) this reduces to

\[
u \eta + v \xi = a \alpha \cos \Theta .
\]

Equation (1) can now be written

\[
E(P') = A \iint_{\mathbb{R}^2} e^{i k a \cos \Theta} \, a \, da d\Theta
\]

\[
= 2 \pi A \int_{a_o}^{a_0} \frac{2 \pi}{a_o^2} \int_{0}^{2 \pi} J_{0}(ka) \, a \, da d\Theta
\]

where \( J_{0} \) is the zero\(^{\text{th}} \) order Bessel Function. Since

\[
\int_{0}^{ka } J_{0}(ka \alpha) \, ka \alpha \, d(ka \alpha) = ka \alpha J_{1}(ka \alpha)
\]

we obtain,

\[
E(P') = 2 \pi a_o^2 A J_{1}(ka \alpha)/ka \alpha .
\]

The first zero corresponds to \( ka \alpha = 3.83 \), so

\[
(2) \quad \alpha = \frac{3.83}{2 \pi a_o} \lambda = \frac{1.22 \lambda}{2 a_o}
\]

where \( \alpha \) is the angular radius of the central disk. Now \( \alpha = \rho R \) so the radius of central disk in the image plane is

\[
\rho = \frac{\lambda R}{2 a_o} .
\]

Consider now figure 2. In a diffraction limited system the minimum resolvable distance in the object corresponds to an image plane dis-

\[Figure 2\]
tance equal to \( y'_{\text{min}} \). Thus

\[
y'_{\text{min}} = \frac{\lambda}{\tan \alpha_0'}
\]

\( \simeq \frac{\lambda}{\sin \alpha_0'} \)

However, from the Abbé sine condition which any microscope can be expected to satisfy we have

\[
y' \sin \alpha_0 = y' \sin \alpha_0'
\]

Letting \( y' = y'_{\text{min}} \), \( \alpha_0' \) will be small so we may write

\[
y' \sin \alpha_0 \simeq \frac{\lambda}{\sin \alpha_0'}
\]

so that

\[
y \simeq \frac{\lambda}{\sin \alpha_0}.
\]

It follows from this that the absolute limit of microscope resolution for two object points is on the order of \( \lambda \).

A2.4a Absorption and Emission Probability Rates. Assume we have an electromagnetic field in a cavity containing a gas of one kind of atom. Let the x-component of its electric field be given by

\[
E_x(t) = 2 E_x^0 \cos 2\pi vt = 2 E_x^0 \cos \omega t
\]

\[
= E_x^0 (e^{i\omega t} + e^{-i\omega t})
\]

Since the radiation is in the visible region, we have wavelengths on the order of \( 10^3 \text{Å} \) or greater. We may therefore take \( E_x(t) \) as constant over the dimensions of an atom at any given instant.

The interaction of the electric field with the atom causes the atom to have an additional transition probability to change states which it did not have before. We may treat the interaction potential as a perturbation to be added to the original undisturbed Hamiltonian, since ordinarily the interaction fields are very small compared to the atomic fields. For convenience, let us choose the atom's center as the origin for the potential energy and apply \( E_x(t) \). Electrons in the atom will feel a force of magnitude \( |eE_x(t)| \) and the work done by the field in inducing a polarization...
of the atom will be given by \( \sum_{j} \epsilon_j \langle X_j \rangle \), where \( j \) is an index corresponding to the \( j \)th electron in the atom. We may therefore write the perturbation due to the external field as,
\[
\psi(x,t) = \epsilon \left( e^{i\omega t} + e^{-i\omega t} \right) \sum_{j} \epsilon_j X_j
\]

Next, we wish to calculate the transition probability and the induced transition rate. We write Schroedinger's equation for the perturbed Hamiltonian \( \mathcal{H}' \),
\[
\mathcal{H}' \psi(x,t) = i\hbar \frac{d}{dt} \psi(x,t)
\]
where \( \mathcal{H}' = \mathcal{H} + V(x,t) \)

and
\[
\psi(x,t) = \sum_{n} a_n(t) \psi_n(x,t).
\]

Substituting, and using the fact that
\[
\mathcal{H} \psi_n(x,t) = i\hbar \frac{d}{dt} \psi_n(x,t)
\]
we get
\[
\sum_{n} a_n(t) V(x,t) \psi_n(x,t) = i\hbar \sum_{n} \frac{d a_n(t)}{dt} \psi_n(x,t)
\]

Multiplying by \( \psi_m(\tau,t) \) and integrating over all coordinate \( \chi \),
\[
\sum_{n} a_n(t) \int \psi_m(\tau,t) V(\tau,t) \psi_n(\tau,t) d\tau = i\hbar \sum_{n} \frac{d a_n(t)}{dt} \int \psi_m(\tau,t) \psi_n(\tau,t) d\tau
\]

Now,
\[
\psi_m(x,t) = e^{-i\chi \omega t} \psi_m(\chi) \text{ and } \psi_n(\chi,t) = e^{-i\chi \omega t} \psi_n(\chi)
\]
so,
\[
\sum_{n} a_n(t) e^{-i\chi \omega t} \int \psi_m(\chi) V(\chi,t) \psi_n(\chi) d\chi = i\hbar \sum_{n} \frac{d a_n(t)}{dt} \psi_n(\tau,t) d\tau
\]

where \( \int \psi_m(\chi) V(\chi,t) \psi_n(\chi) d\chi = \langle \chi | V(\tau,t) | n \rangle \).

Let us now assume the following initial conditions,
\[
a_n(0) = \begin{cases} 0 & n \neq k \\ 1 & n = k \end{cases}
\]
and the perturbation conditions,
\[
a_n(t) = \begin{cases} 0 & n \neq k \\ 1 & n = k \end{cases}
\]

Consider now a state \( l \) such that \( E_l > E_k \) and a time \( t \) greater than \( t = 0 \). Using the above perturbation conditions,
\[
\frac{d a_q(t)}{dt} = \frac{1}{i\hbar} \left[ a_{\kappa}(t) e^{i\hbar(E_{\kappa}-E_{\ell})t} \langle \ell | \Gamma_{\kappa}(t) | K \rangle \right]
\]

or
\[
\frac{d a_q(t)}{dt} = \frac{1}{i\hbar} e^{i\hbar(E_{\ell}-E_{\kappa})t} \langle \ell | \Gamma_{\kappa}(t) | K \rangle .
\]

Since \( V(x,t) = E_x^0 (e^{i\omega t} + e^{-i\omega t}) \hat{X} \hat{X}_j \), we can write
\[
\langle \ell | \Gamma_{\kappa}(t) | K \rangle = \int \psi_\ell^*(x) E_x^0 (e^{i\omega t} + e^{-i\omega t}) \hat{X} \hat{X}_j \psi_\kappa(x) \, dx 
= E_x^0 (e^{i\omega t} + e^{-i\omega t}) \int \psi_\ell^*(x) \hat{X} \hat{X}_j \psi_\kappa(x) \, dx
\]

or
\[
\langle \ell | \Gamma_{\kappa}(t) | K \rangle = E_x^0 (e^{i\omega t} + e^{-i\omega t}) \psi_{\ell k}
\]

where
\[
\psi_{\ell k} = \int \psi_\ell^*(x) \hat{X}_j \psi_\kappa(x) \, dx .
\]

Substituting equation (2) into (1), and integrating from \( t=0 \) to \( t=t' \), we get
\[
a_q(t) = E_x^0 U_{\ell k} \int_0^{t'} (e^{i\omega t} + e^{-i\omega t}) e^{i\hbar(E_{\ell}-E_{\kappa})t'} \, dt'
\]
\[
= E_x^0 U_{\ell k} \left[ \frac{1-e^{i\hbar(E_{\ell}-E_{\kappa}+i\omega)t}}{(E_{\ell}-E_{\kappa}+i\omega)} + \frac{1-e^{i\hbar(E_{\ell}-E_{\kappa}-i\omega)t}}{(E_{\ell}-E_{\kappa}-i\omega)} \right] .
\]

From this result it is clear that \( a_q(t) \) has appreciable magnitude only when \( E_{\ell} - E_{\kappa} \approx \pm \hbar \omega \). Since \( E_{\ell} > E_{\kappa} \), this implies we must use the positive sign. Thus
\[
a_q(t) \approx E_x^0 U_{\ell k} \left[ \frac{1-e^{i\hbar(E_{\ell}-E_{\kappa}+i\omega)t}}{(E_{\ell}-E_{\kappa}+i\omega)} \right] .
\]

This result is the probability amplitude for a transition from state \( \ell \) to state \( k \) in a time \( t \) from the initiation of the perturbation. To find the transition probability, first multiply equation (3) by its complex conjugate,
\[
a_{\ell}^*(t) a_{q}(t) = |E_x^0|^2 |U_{\ell k}|^2 \left[ \frac{2-e^{i\hbar(E_{\ell}-E_{\kappa}+i\omega)t} - e^{i\hbar(E_{\ell}-E_{\kappa}-i\omega)t}}{(E_{\ell}-E_{\kappa}+i\omega)^2} \right]
\]
Next integrate over all frequencies for which the term in square brackets has a finite value. Since the range is small and since the value of the term is effectively zero outside of it, remove $|E_x^0|^2$ from under the integral sign and integrate from $-\infty$ to $\infty$. This gives

$$\mathcal{P}_{sk} = \int_{-\infty}^{\infty} a_x^*(t) a_y(t) \, dt = \frac{1}{\hbar} \frac{|E_x^0|^2 |u_{x,sk}|^2}{4\pi}.$$

Differentiating with respect to time we obtain for the transition rate for absorption,

$$\frac{d\mathcal{P}_{sk}}{dt} = \frac{1}{\hbar} |E_x^0|^2 |u_{x,sk}|^2.$$

Similar results are obtained for the other components so that if $|E_x^0|^2 = |E_y^0|^2 = |E_z^0|^2$, we find

$$(4) \quad \frac{d}{dt} \left[ \mathcal{P}_{sk} + \mathcal{P}_{y,k} + \mathcal{P}_{x,k} \right] = R_{sk} = \frac{|E_x^0|^2}{\hbar^2} \left[ |u_{x,sk}|^2 + |u_{y,sk}|^2 + |u_{z,sk}|^2 \right].$$

Now the energy density in an electromagnetic field in a vacuum is given by

$$U(t) = \frac{1}{8\pi} \left( |E|^2 + |B|^2 \right) = \frac{1}{4\pi} |E|^2 = \frac{1}{4\pi} |E_0|^2 \cos^2 \omega t.$$

Taking the average over time gives

$$\langle U \rangle = \frac{1}{8\pi} |E_0|^2 \equiv \rho(v).$$

Since

$$|E_0|^2 = |E_x|^2 + |E_y|^2 + |E_z|^2 = 3|E_x|^2,$$

and

$$E_x = 2E_x^0,$$

we have

$$|E_x^0|^2 = \frac{2}{3\pi} \rho(v).$$
Substituting into equation (4) and letting

\[ u_{z_{k}}^{2} = |u_{z_{k}}|^{2} = \left[ |u_{x_{z_{k}}}|^{2} + |u_{y_{z_{k}}}|^{2} + |u_{z_{z_{k}}}|^{2} \right], \]

we have finally

\[ R_{z_{k}} = \frac{2\pi u_{z_{k}}^{2}}{3\hbar^{2}} \rho(\nu). \]

Since \( u_{x_{z_{k}}} = \int \psi^{*}_{z_{k}}(\kappa) \psi_{e_{x_{z_{k}}}}(\kappa) d\kappa \), it is evident that \( u_{x_{z_{k}}} = u_{x_{z_{k}} \ell} \), so it follows that

\[ R_{z_{k}} = R_{z_{k} \ell} \]

and the rates of absorption and induced emission are equal.

A2.4b Transition Probability Rates. Suppose we have a blackbody radiation field interacting with an atom. The energy density of the field is given by equation (A2.1b-4),

\[ \rho(\nu) = \frac{8\pi\hbar\nu^{3}}{c^{3}} \frac{1}{e^{\frac{\nu}{kT}} - 1}. \]

Assuming that the transition probability rate from a lower energy level \( k \) to a higher level \( \ell \) is proportional to this energy density, we write

\[ R_{z_{k} \ell} = B_{z_{k} \ell} \rho(\nu). \]

However the transition probability rate from energy level \( \ell \) to level \( k \) must have the additional contribution from spontaneous emission. In this case we write

\[ R_{k \ell} = B_{k \ell} \rho(\nu) + A_{k \ell}. \]

At thermal equilibrium, the total transition rate up must equal that down. Since the transition rate is equal to the transition probability rate times the number of atoms per unit volume at the particular energy level, we have

\[ n_{k} B_{z_{k} \ell} \rho(\nu) = n_{\ell} B_{k \ell} \rho(\nu) + n_{\ell} A_{k \ell}. \]

Solving for \( \rho(\nu) \) we get,

\[ \rho(\nu) = \frac{A_{k \ell} / B_{z_{k} \ell}}{n_{k} / n_{\ell} - B_{k \ell} / B_{z_{k} \ell}}. \]

Using the Boltzmann equation and allowing for the possible
degeneracies of the levels, we have

\[ \rho(\nu) = \frac{A_{kl}}{B_{lk}} \left( \frac{g_k}{g_l} e^{\theta_k \nu - B_k / \nu} - \frac{g_l}{g_k} e^{\theta_l \nu - B_l / \nu} \right) \]

where

\[ \frac{n_{kl}}{n_l} \equiv \frac{g_k}{g_l} e^{\theta_k \nu - B_k / \nu} \quad \omega_{kl} = \omega_k - \omega_l = \nu . \]

Comparing this last result with equation (1), we see that

\[ B_{kl} = \frac{g_k}{g_l} B_{lk} \]

and

\[ A_{kl} / B_{kl} = \frac{g_k}{g_l} \frac{8 \pi h \nu^3 / c^3}{\nu} . \]

From equation (2) the induced-transition probability rate from level \( l \) to level \( k \) due only to the interaction with the field is,

\[ R_{kl} = \frac{g_k}{g_l} \frac{8 \pi h \nu^3 / c^3}{\nu} \rho(\nu) . \]

If we define a spontaneous transition time

\[ t_{\text{spon}} = A_{kl}^{-1} \]

then,

\[ R_{kl} = \frac{g_k}{g_l} \frac{c^3}{8 \pi h \nu^3 t_{\text{spon}}} \rho(\nu) . \]

This result has the general form that we are looking for; however, it assumes a blackbody radiation field whereas, in fact, the radiation field of a laser cavity is sharply peaked with a linewidth distribution which may be written \( g(\nu - \nu_0) \). For some particular frequency, we can write

\[ \rho'(\nu') = \rho_\nu \delta(\nu - \nu) \]

so that the energy density at frequency \( \nu' \) is given by

\[ \rho(\nu') g(\nu' - \nu_0) = \rho_\nu \delta(\nu - \nu) g(\nu - \nu_0) . \]

To find the total induced-transition probability rate over the radiation linewidth, we integrate and obtain

\[ W_{kl} = \frac{g_k}{g_l} \int_{-\infty}^{\infty} \frac{c^3 \rho_\nu}{8 \pi h \nu^3 t_{\text{spon}}} \cdot g(\nu - \nu_0) d\nu' \]

or

\[ W_{kl} = \frac{g_k}{g_l} \frac{c^3 \rho_\nu}{8 \pi h \nu^3 t_{\text{spon}}} g(\nu - \nu_0) . \]
Since $I_\nu$, the energy flux, is given by $c \rho_\nu$, we can write finally,

$$W_{kl} = \frac{g_k}{g_\nu} \frac{c^2 I_\nu}{8 \pi h \nu^3 t_{\text{span}} \cdot g(\nu-\nu_0)}.$$  

In similar fashion, but noting that $A_{kl}/B_{kl} = \frac{8 \pi^3}{c^3}$, we get for the induced-transition probability rate from level $k$ to level $l$,

$$W_{lk} = \frac{c^2 I_\nu}{8 \pi h \nu^3 t_{\text{span}} \cdot g(\nu-\nu_0)}.$$  

If we know $t_{\text{span}}$ experimentally, we can use it in equation (4) or (5) to calculate the desired induced-transition probability rate. However, we can also estimate it by solving equation (3), namely,

$$t_{\text{span}} = \frac{c^2 \rho(\nu)}{8 \pi h \nu^3 R_{kl}}$$

and then using equation (A2.4a-5) for $R_{kl}$. Thus,

$$t_{\text{span}} = \frac{c^2 \rho(\nu)}{8 \pi h \nu^3 \left( \frac{2 \pi}{3 \hbar^2 \omega_{kl}^2} \rho(\nu) \right)}$$

or

$$t_{\text{span}} = \frac{3 hc^3}{8 \pi \omega^3 \omega_{kl}^2}.$$  

A2.4c Amplification. Consider a quasimonochromatic wave propagating in a medium with level population densities $n_k$ and $n_k$. Further, consider a unit area normal to the direction of propagation. As the wave crosses this unit area and travels a unit distance, it will increase the net transition rate to the lower energy level by

$$n_k W_{kl} - n_k W_{lk}.$$  

The net increase in energy flux per unit distance can then be written

$$\frac{dI_\nu}{d\nu} = \left[ n_k W_{kl} - n_k W_{lk} \right] h \nu.$$  

One can think of this as the wave moving along and in doing
so adding volume which then contributes to the energy flux already in the wave. Substituting from equations (A2.4b-4) and (A2.4b-5) we have,

\[
\begin{align*}
\frac{dI_v}{dz} &= -\left[n_k \frac{g_i}{g_k} - n_l\right] \frac{c^2 g(u-v_0)}{8\pi v^2 t_{spom}} I_v.
\end{align*}
\]

Solving for \( I_v \),

\[
\int_{I_v(0)}^{I_v(z)} \frac{dI_v}{I_v} = -\left[n_k \frac{g_i}{g_k} - n_l\right] \frac{c^2 g(u-v_0)}{8\pi v^2 t_{spom}} \int dz,
\]

we get

\[
I_v(z) = I_v(0) e^{-\alpha z}
\]

where

\[
\alpha = \left[n_k \frac{g_i}{g_k} - n_l\right] \frac{c^2 g(u-v_0)}{8\pi v^2 t_{spom}}.
\]

We see immediately that the condition for exponential growth of flux, assuming no losses, is a population inversion given by

\[
n_l > n_k \frac{g_i}{g_k}.
\]

If the growth is sufficient to compensate for whatever loss mechanisms there are, then the energy density in the cavity will increase, and oscillation near the resonant frequency of the cavity will be established.

A2.4d Cavities. The archetype resonant cavity is the Fabry-Perot etalon or interferometer, figure 1. There are three properties of such a cavity that are of particular interest here; the longitudinal modes, the transmission coefficient, and the decay time constant. The details of losses and other factors which affect bandwidth are discussed in (A2.4f).
Longitudinal Modes: Assume that both plates have identical reflection and transmission properties. Let $\theta$ be the phase shift that takes place in the cavity wave at each reflection, and let $\omega$ be an angular frequency that satisfies the condition for standing waves. Then, for the field pattern to reproduce itself, the phase shift $\delta$ must be an integral multiple of $2\pi$ for each round trip. Thus

$$\delta = \omega \frac{2d}{c} - 2\theta = m \cdot 2\pi$$

where $m$ is an integer and $c$ is the velocity of light. The angular frequency interval between modes that satisfies this condition is

$$\Delta \omega = \frac{\pi c}{d}$$

or

$$\Delta \nu = \frac{c}{2d}$$

for the frequency spacing of the modes.

Transmission Properties: The Fabry-Perot etalon can be related to the laser cavity by focusing attention on the energy already in the cavity, say at point $0$. How the energy got there is of no concern, nor is the energy that escapes to the left pertinent. This latter would contribute to the reflected beam if the etalon were used conventionally.

If we start at $0$ and follow the various rays that contribute to the transmitted beam, we can write for the transmitted electric field,

$$E^T = E_0 \tau + E_0 \rho^2 e^{-i\delta} + E_0 \tau \rho^4 e^{-2i\delta} + \ldots$$

where $E_0$ is the initial electric field at $0$, and $\delta$, the total phase shift per transit. Summing the series we find,

$$E^T = \frac{E_0 \tau}{(1 - \rho^2 e^{-i\delta})} .$$

Now the transmitted intensity is proportional to $E^T \cdot E^T$, so we can write for the intensity transmission coefficient $T$, 
\[ T' = \frac{|E'|^2}{|E_0|^2} = \frac{T^2}{1 - \rho^2 e^{i\delta}} \]

Clearly \( T' \) is a maximum when \( \cos \delta = 1 \), so

\[ T_{\text{max}}' = \frac{T^2}{(1 - \rho^2)^2} \]

Replacing \( \rho \) by \( R \), the intensity reflection coefficient, and substituting into (1) we get

\[ T' = T_{\text{max}}' \frac{(1-R)^2}{1 + R^2 - 2R \cos \delta} \]

A useful form of this equation is found by defining

\[ F = \frac{4R}{(1-R)^2} \]

which is called the finesse and is actually a measure of the contrast between the maximum and minimum transmission coefficients as \( R \) is varied. Using the trigonometric relation, \( \cos \delta = 2 \sin^2 \frac{\delta}{2} - 1 \), and writing equation (2) in terms of \( F \), we obtain for the intensity transmission coefficient,

\[ T' = T_{\text{max}}' \left[ 1 + F \sin^2 \frac{\delta}{2} \right]^{-1} \]

Decay Time Constant: A useful index of the characteristics of a cavity is the quality factor \( Q \). There are a number of equivalent ways in which \( Q \) may be defined. These are:

\[ Q = \left[ \frac{\nu_0}{\delta \nu} \right] \]

where \( \nu_0 \) is the resonance frequency and \( \delta \nu \) is the half-power bandwidth;

\[ Q = 2\pi \left[ \frac{\text{Stored Oscillator Energy}}{\text{Energy lost per period}} \right] \]

or

\[ Q = 2\pi \nu_0 \left[ \frac{\text{Stored Oscillator Energy}}{\text{Energy lost per second}} \right] \]

Now, suppose that \( \nu_c \) is the fractional loss of power
per one way transit in the cavity due to all causes. For simplicity consider the case of a single mode only, and let \( U \) be its total cavity energy. Using equation (6) we find for the \( Q \) of the cavity

\[
Q = \frac{2\pi \nu_0 \nu}{U \nu_c c/d} = \frac{2\pi \nu_0 d}{c\nu_c}.
\]

Using equation (4) we find for the cavity bandwidth

\[
\Delta \nu_c = \frac{c\nu_c}{2\pi d}
\]

and an approximate value for the decay time constant

\[
\tau_c \approx 2\pi \frac{d}{c \nu_c}
\]

Alternatively, we may derive \( \tau_c \) as follows. Let \( \rho(\nu) \) be the energy density at frequency \( \nu \) and consider the time rate of change of density due to cavity losses.

\[
\frac{\Delta \rho(\nu)}{\rho(\nu)} = -\gamma_c \frac{\Delta t}{\pi} \quad \Rightarrow \quad \hat{\rho}(\nu) = -\frac{c\nu_c \rho(\nu)}{d}
\]

Solving, we get

\[
\rho(\nu) = \rho(\nu) e^{-\frac{c\nu_c t}{d}}.
\]

We use this last expression to define \( \tau_c \), the time in which the cavity energy drops to \( e' \) of its initial value. Thus,

\[
\tau_c \equiv \frac{d}{c \nu_c}, \quad (7)
\]

From equation (5) we can obtain an expression for \( \tau_c \) in terms of the cavity \( Q \). Let \( \tau_0 \) be the period of one oscillation and noting that it is very small compared to \( \tau_c \), we have

\[
Q = \left( \frac{2\pi \rho_0}{\rho_0 e^{-\tau_0 / \tau_c}} \right) = 2\pi \frac{\tau_c}{\tau_0} = \omega_0 \tau_c.
\]

So,

\[
\tau_c = \frac{Q}{\omega_0} \quad , \quad (8)
\]

A2.4e Threshold Condition. Let us assume that we have a single mode which interacts with the atoms in the cavity. The decay time of the mode due to losses can be expressed by equation (A2.4d-8),
where \( Q \) is the quality factor of the system. Therefore, the rate of flux decay can be written as

\[
\frac{dI_v}{dt} = -I_v \frac{dt}{\tau_c}
\]

or

\[
(1) \quad \frac{dI_v}{dt} = -I_v/\tau_c.
\]

From equations (A2.4c-1) and (A2.4c-2), we have

\[
\frac{dI_v}{dt} = \frac{dI_v}{dt} \cdot \frac{dt}{\tau_c} = \frac{1}{\tau_c} \frac{dI_v}{dt} = -\alpha I_v,
\]

or

\[
(2) \quad \frac{dI_v}{dt} = -\alpha c I_v.
\]

Equating equations (1) and (2), we obtain the threshold condition for oscillation. Thus,

\[-\alpha c I_v = I_v/\tau_c\]

or

\[
\left[ n_2 - \frac{g_e}{g_\lambda} n_k \right] \frac{c^2 g(v-v_o)}{8\pi v^2 t_{\text{span}}} = \frac{1}{\tau_c}.
\]

Solving for the required population inversion we get

\[
\left[ n_2 - \frac{g_e}{g_\lambda} n_k \right] = \frac{8\pi v^2}{c^2 g(\alpha)} \cdot \frac{t_{\text{span}}}{\tau_c}.
\]

But at the threshold only the peak frequency \( v_o \) is involved so

\[
\left[ n_2 - \frac{g_e}{g_\lambda} n_k \right] = \frac{8\pi v_o^2}{c^2 g(\alpha)} \cdot \frac{t_{\text{span}}}{\tau_c}.
\]

Therefore oscillation will take place if

\[
(3) \quad \left[ n_2 - \frac{g_e}{g_\lambda} n_k \right] \geq \frac{8\pi v_o^2}{c^2 g(\alpha)} \cdot \frac{t_{\text{span}}}{\tau_c}.
\]

\[A2.4f\] **Output Bandwidth.**\(^{20-24}\) The factors that affect the output bandwidth of a laser can be grouped into two principle categories; those that are related to the emissions or emitting atoms, and those that are related to the cavity. Let us first consider the emission factors.

Natural Linewidth: \(^{20}\) The linewidth of an emission is related through the uncertainty principle to the lifetime
of a state and this, in turn, is related to the total probability rate for spontaneous transition to all lower energy states. Defining \( \rho_L = \sum_k A_{kL} \) as the spontaneous transition probability rate, we can write for the total number of transitions in a period \( dt \),

\[
N(t) \rho_L dt
\]

where \( N(t) \) is the number of atoms in a given state at time \( t \). This, however, is equal to the decrease in the number during \( dt \), so

\[
dN(t) = -N(t) \rho_L dt.
\]

Solving for \( N(t) \) we find

\[
N(t) = N(\omega) e^{-\rho_L t}.
\]

Now let us define the lifetime of a state as the average time an atom spends in the state:

\[
\tau_L = \frac{\int_0^\infty tN(t)dt}{\int_0^\infty N(t)dt} = \frac{\rho}{\rho_L}
\]

or

\[
\tau_L = \frac{1}{\sum_k A_{kL}}^{-1},
\]

where \( A_{kL} \) is the spontaneous transition probability rate from state \( l \) to some lower state \( k \).

From the uncertainty principle,

\[
\Delta t \Delta E \sim \hbar
\]

or

\[
\tau_L \hbar \Delta \nu_L \sim \hbar
\]

so

\[
\tau_L \sim \frac{1}{2\pi \Delta \nu_L},
\]

and the bandwidth is given by

\[
\Delta \nu_L \sim \frac{1}{2\pi \tau_L} = \frac{\rho_L}{2\pi \tau_L}.
\]

Doppler Broadening: Doppler broadening is an important factor in laser bandwidth considerations and leads to a Gaussian linewidth distribution.

Assume that the emitting atoms are in a gaseous state and consider only the components of velocity along
the axis of the laser. Due to Doppler shifting, the emission frequency will be given by

$$\nu = \nu_0 \left(1 \pm \frac{V_x}{c}\right)$$

where $\nu_0$ is the central frequency and $V_x/c$ is the ratio of an atom's x-component of velocity and the speed of light.

Solving for $V_x$ we get

$$V_x = (\nu - \nu_0) c / \nu_0$$

and

$$d\nu_x = \left(\frac{c}{\nu_0}\right) d\nu.$$

To find the probability that at temperature $T$ the x-component of the velocity lies between $V_x$ and $V_x + d\nu_x$, we use the Maxwell velocity distribution:

$$P(\nu_x) = \frac{\theta(\nu_x)}{d\nu_x} = \left[\frac{\beta m}{2\pi}\right]^{\frac{1}{2}} e^{-\frac{\beta m}{2} \frac{\nu_x^2}{\nu_0^2}} d\nu_x$$

Changing the variable back to frequency we get

$$P(\nu - \nu_0) = \frac{\theta(\nu - \nu_0)}{d\nu} = \frac{\theta}{\nu_0} \left[\frac{\beta m}{2\pi}\right]^{\frac{1}{2}} e^{-\frac{\beta m}{2} \frac{\nu_x^2}{\nu_0^2}} d\nu_x$$

We can therefore write the intensity of emission as

$$\mathcal{I}(\nu) = \mathcal{I}(\nu_0) e^{-\frac{\beta m}{2} \frac{\nu_x^2}{\nu_0^2} (\nu - \nu_0)^2}.$$ 

To find the bandwidth at the half-power points, let $\mathcal{I}(\nu) = 1/2$ and then take the natural logarithm. Thus,

$$\Delta\nu_{1/2} = -\frac{\beta m}{2} \frac{\nu_x^2}{\nu_0^2} (\nu - \nu_0)^2.$$ 

Rearranging,

$$\Delta(\nu - \nu_0) = 2\nu_0 \left[\frac{2\ln 2}{\beta mc^2}\right]^{1/2}$$

and substituting for $\Delta(\nu - \nu_0)$, we get

$$\Delta\nu_{1/2} = 2\nu_0 \left[\frac{2\ln 2}{\beta mc^2}\right]^{1/2}.$$

Next consider the cavity-related factors which affect the output bandwidth.

Transmission Losses: From the earlier discussion of cavities, an expression for the intensity transmission coefficient was derived. We now write an analogous expression for the transmitted intensity,

$$I(\nu) = I_{\text{max}} \left[1 + F \sin^2 \frac{\nu - \nu_0}{\Delta \nu_{1/2}}\right]^{-1}.$$
where

\[ F = \frac{4R}{(1-R)^2} \quad \text{and} \quad S = \frac{2\pi n}{c} \cdot \frac{2d}{\lambda} - 2\theta. \]

Assume that \( V \) is close to one of the modal frequencies so that \( S \ll 1 \) and \( \sin \frac{S}{2} \sim \frac{S}{2} \). We may then write

\[ I(\omega) \simeq I_{\text{max}} \left( \frac{1}{F} + \left( \frac{S}{2} \right)^2 \right)^{-1}. \]

But this is just a Lorentzian distribution function whose half-power points are given by

\[ \frac{1}{2} = \frac{1}{F} + \left( \frac{S}{2} \right)^2 \]

or

\[ \left( \frac{S}{2} \right)^2 = \frac{1}{F} \quad \Rightarrow \quad S = \pm 2/\sqrt{F}. \]

This gives for the full half-power bandwidth,

\[ \Delta \nu_T = \frac{c}{2\pi d\sqrt{F}}. \]

Finally, substituting for \( F \) we get

\[ \Delta \nu_T = \frac{c(1-R)}{2\pi d\sqrt{F}}. \]

\( \Delta \nu_T \) is the bandwidth resulting from losses in transmission due to having an intensity reflection coefficient less than unity.

**Diffraction Losses:** In any real laser, the cavity is not a Fabry-Perot etalon with infinite plane plates. Usually the cavity has spherical mirrors with diameters which are small compared to the separation distance. While the sphericity of the mirrors will not be discussed in this simplified analysis, the effects of diffraction will be.

Consider the mirror configuration shown in figure 1, and suppose a more or less paraxial light beam is reflected back and forth between the mirrors. If the beam is to make \( n \) transits, only those rays which subtend angles less than \( \theta /n \)
will make all trips without "walking out" of the system. Let us now analyze the "walkout" loss which is really a diffraction loss.

After each reflection most of the energy is diffracted into an angle given by the Airy disc; namely, $\frac{\lambda}{a^2}$. The condition for minimum loss for an $n$-reflection process is then

$$\frac{a}{n^2} > \frac{\lambda}{a^2}.$$ 

Solving for $n$ and noting that $\frac{a^2}{\lambda^2}$ is the Fresnel number $N$, (that is, the number of zone rings seen by, say, mirror A) we have

$$n < \frac{a^2}{\lambda^2},$$

or

$$N > n,$$

for the minimum loss condition.

Now consider the situation during a single pass. Light energy is reflected by mirror B back toward A. That part of the energy which subtends an angle less than $\frac{\lambda}{a}$ will remain in the system but that part which subtends angles between $\frac{a}{\lambda}$ and $\frac{a+\delta}{\lambda}$, where $\frac{\lambda}{a+\delta}$ corresponds to the Airy disc, will be diffracted out of it. The ratio of energy impinging on the annulus to that on the mirror is a measure of the fractional energy loss per pass. Therefore,

$$\gamma_F = \frac{\text{Energy on Annulus}}{\text{Energy on Mirror}} \approx \frac{2\pi a \delta}{\lambda a^2}.$$

Assuming a diffraction angle $\theta = \frac{\lambda}{a}$, and that $\delta = \frac{\theta d}{2}$ we obtain

$$\gamma_F = \frac{\lambda d}{a^2} = \frac{1}{N}.$$

Next, using equation (A2.4d-6) for $Q$ we may write

$$Q = 2\pi \nu \frac{\rho(\nu)}{\gamma_F \rho(\nu) a/d} = \frac{2\pi \nu d}{\gamma_F}.$$

where $\rho(\nu)$ is the energy density of the field and $\gamma_F \rho(\nu) a/d$ is the energy density lost per second. We may find the effect on the bandwidth due to diffraction losses by using equation (A2.4d-4). Thus,
Absorption and Other Losses: It is clear from the foregoing analysis that if absorption can be expressed in terms of fractional loss per pass, then the effect on the bandwidth is given by an expression

\[
\Delta \nu = \frac{c \gamma_A}{2 \pi d} = \frac{c \lambda}{2 \pi a^2}
\]

where \( \gamma_A \) is the fractional loss of radiation intensity per pass due to absorption within the cavity.

We can extend these results to other losses also and furthermore, we can combine them to obtain

\[
\Delta \nu_{\text{cav}} = \frac{c}{2 \pi d} \left( \gamma_T + \gamma_F + \gamma_A + \gamma_o \right) = \frac{c}{2 \pi d} \gamma_{\text{cav}}
\]

where \( \Delta \nu_{\text{cav}} \) is the full frequency width of the cavity due to transmission, diffraction, absorption, and other losses, respectively, and \( \gamma_{\text{cav}} \) is the fractional intensity loss from all sources.

Beam Bandwidth: We need now to relate the emission and cavity factors to the actual bandwidth of the output beam.

The emission factors essentially determine both the number of cavity modes into which energy is pumped and the gains of the respective modes. The cavity factors affect the losses of the modes. By equating gains to losses in deriving the threshold condition, it would appear that the \( Q \) of a laser would be infinite at threshold. However, this is not the case. There is a finite output bandwidth due to the fact that noise from spontaneous emission has been introduced in the process of generating the beam itself. To understand this, consider a particular mode which is both receiving coherent energy due to stimulated emission and losing it through various dissipation processes.
We can define the $Q$ of the mode by

$$Q = 2\pi \nu \left( \frac{\text{Coherent Energy Stored in Mode}}{\text{Net Loss of Coherent Energy per second}} \right).$$

Now the net loss of coherent energy per second is given by

$$P_{\text{coh}}^{(\text{out})} - P_{\text{coh}}^{(\text{in})}.$$

However, for steady state operation,

$$P_{\text{coh}}^{(\text{in})} + P_{\text{inc}}^{(\text{in})} = P^{(\omega)} = P_{\text{coh}}^{(\text{out})}.$$

So,

$$P_{\text{coh}}^{(\text{out})} = P_{\text{coh}}^{(\text{in})} + P_{\text{inc}}^{(\text{in})}.$$

The expression for $Q$ now becomes

$$Q = 2\pi \nu \left( \frac{\text{Coherent Energy Stored in Mode}}{P_{\text{inc}}^{(\text{in})}} \right).$$

We want to evaluate this expression at threshold. First the coherent energy in the cavity in the mode of interest is given by

$$Nh\nu = I_{\nu}a d/c = P^{(\omega)} \frac{d}{c\gamma_{\text{CAV}}}.$$

where $N$ is the number of photons in the cavity mode, $I_{\nu}$ is the mode intensity, $a$ is the mode cross-section, and $\gamma_{\text{CAV}}$ is the total fractional loss for the mode per cavity transit. The rate of stimulated emission into the mode for the whole cavity is then

$$B_{kl} n_{k} I_{\nu} g(\nu - \nu_{o}) a \frac{d}{c\gamma_{\text{CAV}}}.$$

Since the ratio of the stimulated to spontaneous emission rates is equal to the number of photons in the mode, and since $h\nu$ times the spontaneous emission rate is $P_{\text{inc}}^{(\text{in})}$, we have

$$P_{\text{inc}}^{(\text{in})} = h\nu \frac{B_{kl} n_{k} g(\nu - \nu_{o}) a \frac{d}{c}}{I_{\nu}a d/c\gamma_{\text{CAV}}}.$$

or

$$P_{\text{inc}}^{(\text{in})} = B_{kl} n_{k} g(\nu - \nu_{o}) (h\nu)^{2}.$$

Using equation (7) for the $Q$ of the cavity, we obtain

$$Q = 2\pi \nu \frac{P^{(\omega)} d/c\gamma_{\text{CAV}}}{B_{kl} n_{k} g(\nu - \nu_{o}) (h\nu)^{2}}.$$. 
Next, make use of the threshold condition, equation (A2.43-3), to get

\[ \left[ n_e - \frac{g_e}{g_e} n_k \right] = \frac{8 \pi \nu^2}{c^3 g(o)} \cdot \frac{\tau_{\text{pump}}}{\tau} \]

or

\[ A_k \left[ n_e - \frac{g_e}{g_e} n_k \right] = \frac{g_e \nu^2}{c^3 g(o)} \cdot \frac{c \gamma_{\text{cav}}}{d} \cdot \frac{1}{\tau} \]

Since \( \frac{8 \pi \nu^2}{c^3} = \frac{A_{\text{cav}}}{B_{\text{cav}}} \), we can write this last result as

\[ B_{\text{cav}} n_e g(o) = \frac{c \gamma_{\text{cav}}}{hd} \left[ 1 - \frac{g_e n_k}{g_e n_k} \right]^{-1} \]

Substituting into equation (8) we obtain

\[ Q = \frac{2 \pi}{\nu} \frac{p^0}{c^2 \gamma_{\text{cav}}} \left[ 1 - \frac{g_e n_k}{g_e n_k} \right] \]

Since \( Q = \frac{\gamma_{\text{cav}}}{\nu} \), and \( \Delta \nu_{\text{cav}} = \frac{\epsilon}{2 \pi} \gamma_{\text{cav}} \) we have finally

(9) \[ \Delta \nu = \frac{2 \pi (\Delta \nu_{\text{cav}})^2 h \nu}{p^0} \left[ 1 - \frac{g_e n_k}{g_e n_k} \right] \]

or for large population inversions,

(10) \[ \Delta \nu \approx \frac{2 \pi (\Delta \nu_{\text{cav}})^2 h \nu}{p^0} \]

These two last equations relate the bandwidth of the cavity to the bandwidth of the output laser beam.
Appendix A

References

2. Feynman et al, p. 4-8.
8. Klein, Sect. 8.3.
24. Maitland and Dunn, p. 102.
Appendix B: Properties of Helium-4
B3.2a Bose-Einstein Distribution. We wish to determine the mean occupancy distribution for He-4; that is, the mean number of atoms that occupy any given "single particle" state. As with photons, helium atoms are assumed to be essentially non-interacting Bose particles; however, their total number is fixed.

One approach is to assume that the system of interest is some particular orbital (single particle state). Viewed in this way, the grand canonical distribution applies, since the system can exchange both atoms and energy with the reservoir (all other orbitals). We have, therefore, for the probability of finding \( n_i \) atoms in a given orbital,

\[
P_{n_i} = \frac{e^{-\beta n_i (E-\mu)}}{\sum_{n_r} e^{-\beta n_r (E-\mu)}} \quad ; \quad n_r = 0,1,2, \ldots ,
\]

where \( n_i \) is the number of atoms in the orbital, \( E \) is the orbital energy, and \( \mu \) is the chemical potential. The restrictions that the total energy and the total number of atoms are fixed are accounted for by \( \beta \) and \( \mu \).

We find the average orbital occupancy from

\[
\bar{n} = \frac{\sum n_r e^{-\beta n_r (E-\mu)}}{\sum_{n_r} e^{-\beta n_r (E-\mu)}}
\]

\[
= \frac{\sum n_r [e^{-\beta (E-\mu)}]^{n_r}}{\sum_{n_r} e^{-\beta n_r (E-\mu)}}
\]

If, momentarily, we let \( x = e^{-\beta (E-\mu)} \), we can write

\[
\bar{n} = \frac{\sum n_r x^{n_r}}{\sum_{n_r} x^{n_r}} = \frac{\int x^{n_r} \frac{d}{dx} \sum_{n_r} x^{n_r}}{\sum_{n_r} x^{n_r}}
\]

But \( \sum_{n_r} x^{n_r} \approx 1/(1-x) \) if we assume the total number of atoms to be very large so
\[ \bar{n} = \frac{x \cdot 1/(1-x)^2}{1/(1-x)} = \frac{x}{1-x} = \frac{1}{x-1} \].

Substituting back we obtain, finally, the Bose-Einstein distribution function,

\[ (2) \quad \bar{n} = \frac{1}{(e^{\beta(\varepsilon - \mu)} - 1)} \]

B3.2b Chemical Potential.\(^2,3\) The grand canonical distribution function can be derived by considering the change in entropy of a system due to an exchange of energy and/or an exchange of particles with a juxtaposed reservoir. To first order, this change is given by

\[ \delta S \approx \left( \frac{\partial S}{\partial E} \right)_{res} \delta E + \left( \frac{\partial S}{\partial n} \right)_{res} \delta n \]

\[ \approx \left( \frac{\partial S}{\partial E} \right)_{res} \delta E + \left( \frac{\partial E}{\partial n} \right)_{res} \delta n \]

We define \( \bar{T} \equiv \left( \frac{\partial S}{\partial E} \right)_{res} \) and \( \bar{\mu} \equiv \left( \frac{\partial E}{\partial n} \right)_{res} \), where \( \bar{T} \) is the absolute temperature and \( \bar{\mu} \) is the chemical potential. Therefore,

\[ (1) \quad \delta S = \bar{T} \left\{ \delta E - \bar{\mu} \delta n \right\} \]

For the case in which the system is a single particle orbital, \( \delta E = \varepsilon \delta n \), where \( \varepsilon \) is the single particle orbital energy. We have therefore

\[ (2) \quad \delta S = \bar{T} \left\{ \varepsilon - \bar{\mu} \right\} \delta n \]

To show that the grand canonical distribution follows from this equation, recall that

\[ S \equiv k_B \ln \Omega(E,n) = k_B \ln \Omega(n) \]

where \( \Omega(n) \) is the number of states corresponding to particle number \( n \). Using this equation, we may form \( S - S_o \). Then letting \( \delta S = S - S_o \) and \( \delta n = n - n_o \), we can write
\[ \Omega(n) = \Omega_0(n_0) e^{-\frac{1}{kT}[\varepsilon - \mu](n-n_0)} = \Omega_0(n_0) e^{-\frac{1}{\beta}[\varepsilon - \mu]} n. \]

Since \( P_n = c_m e^{\frac{1}{\beta}[\varepsilon - \mu]n} \), we find upon normalization

\[ P_n = \frac{e^{-\beta(\varepsilon - \mu)n}}{\sum e^{-\beta(\varepsilon - \mu)n}}. \]

Equation (2) is just a special case of the fundamental thermodynamic identity and can be written

\[ T dS = (\varepsilon - \mu) d\theta \]

where \( T dS \) is the heat absorbed by the orbital when \( \delta n \) atoms are added to it. Since \( (\varepsilon - \mu) \) is clearly a generalized force conjugate to the generalized coordinate \( n \), we may think of it as tending to move atoms in or out of the orbital depending on the direction to equilibrium.

To evaluate \( \mu \), consider the Bose-Einstein distribution function, equation (B3.2a-2). Assume the system corresponds to the ground orbital with energy \( \varepsilon_0 \). When \( T \) approaches absolute zero, \( n \rightarrow N \) where \( N \) is the total number of atoms. This means that \( (\varepsilon_0 - \mu) \) must become a very small positive number. Expanding the exponential, we obtain the approximate expression

\[ N \approx \frac{1}{1 + \beta(\varepsilon - \mu) - 1} = \frac{1}{\beta(\varepsilon_0 - \mu)}. \]

Solving for \( \mu \) we get

\[ \mu = \varepsilon_0 - \frac{1}{\beta N}, \]

which says that in the limit as \( T \rightarrow 0 \), \( \mu \rightarrow \varepsilon_0 \). If \( U = N\varepsilon_0 \), then

\[ N\mu = U - \frac{1}{\beta} \]

or

\[ (3) \quad N\mu = U - kT. \]
B3.2c Ground Orbital Occupancy. Because of the connection of the ground orbital to superfluidity, it is useful to obtain a practical expression relating the occupancy to temperature. To do this we shall need an expression for the density of orbitals with respect to energy.

Assume we have a single helium atom contained in a rectangular box. Solving Schrödinger's equation for these boundary conditions yields the energy expression

\[
\varepsilon = \frac{\hbar^2 k^2}{2m} = \frac{\pi^2 \hbar}{2m} \left[ \left( \frac{n_x}{L_x} \right)^2 + \left( \frac{n_y}{L_y} \right)^2 + \left( \frac{n_z}{L_z} \right)^2 \right],
\]

where \( L_x, L_y \) and \( L_z \) are the box dimensions and \( n_x, n_y, n_z \) and \( n_z = 1, 2, \ldots \). A single state is specified by the number triplet \( (n_x, n_y, n_z) \) and the \( k \)-space volume per state is readily seen to be \( \pi^3 / L_x L_y L_z = \pi^3 / V \). The total \( k \)-space volume in the first quadrant is given by \( \frac{1}{6} \pi k^3 \) and the differential volume from \( k \) to \( k + \delta k \) by \( \frac{1}{6} \pi k^2 \delta k \). The number of states in the differential volume is therefore

\[
\delta n_k = \frac{\pi^2 \varepsilon \delta k}{\pi^3 / V} = \frac{V}{2\pi^2} k^2 \delta k.
\]

Now, from equation (1)

\[
k^2 = \frac{2m \varepsilon}{\hbar^2}
\]

and

\[
\delta k = \frac{1}{2} \left[ \frac{2m}{\hbar^2} \right]^{1/2} \varepsilon^{-1/2} \delta \varepsilon.
\]

Therefore substituting into equation (2) we get for the number of states between \( \varepsilon \) and \( \varepsilon + \delta \varepsilon \)

\[
\delta n_\varepsilon = \frac{V}{4\pi^2} \left[ \frac{2m}{\hbar^2} \right]^{3/2} \varepsilon^{1/2} \delta \varepsilon
\]
or

\[
\sigma(\varepsilon) = \frac{\delta n_\varepsilon}{\delta \varepsilon} = \frac{V}{4\pi^2} \left[ \frac{2m}{\hbar^2} \right]^{3/2} \varepsilon^{1/2}.
\]

If \( N \) is the total number of helium atoms, \( N_0 \) the number in the ground orbital, and \( N_\varepsilon \) the number in excited
orbitals, we may write

\[ N = N_o(T) + N_e(T) \]

\[ = N_o(T) + \sum_{e=1}^{\infty} n_e \]

where \( n_e \) is the number of atoms in a particular excited orbital. We may express the summation in terms of energies by using equation (1) for the density of orbitals. Thus we may write, approximately,

\[ N \approx N_o(T) + \int_{\epsilon_0 - \mu}^{\infty} \sigma(\epsilon) n(\epsilon, T) \, d\epsilon, \]

where \( n(\epsilon, T) = \frac{1}{e^{\frac{\epsilon}{kT}} - 1} \) and \( N_o(T) = \frac{1}{e^{\frac{\epsilon_0 - \mu}{kT}} - 1} \).

Using the approximation that the lower limit of the integral is zero and that \( \mu \approx 0 \) in comparison to all \( \epsilon \) except \( \epsilon_0 \), we may write

\[ N \approx N_o(T) + \int_{0}^{\infty} \frac{\sqrt{T}}{2\pi^{2}} \left[ \frac{2m}{\pi^{2}} \right]^{\frac{3}{2}} e^{\frac{T}{kT} \left[ \frac{2e}{kT} - 1 \right]} \, d\epsilon, \]

which after integration becomes

\[ N \approx N_o(T) + 2.612 V \left[ \frac{mkT}{2\pi \hbar^2} \right]^{\frac{3}{2}}. \]

Solving for \( N_o(T) \) we obtain

\[ (4) \quad N_o(T) = N - 2.612 V \left[ \frac{mkT}{2\pi \hbar^2} \right]^{\frac{3}{2}}. \]

If we define a temperature \( T_c \), called the Einstein condensation temperature, as that temperature where the total number of atoms equals the number of excited atoms, then

\[ T_c = \frac{2\pi \hbar^2}{mk} \left[ \frac{N}{2.612 V} \right]^{\frac{2}{3}}. \]

Note that if we write \( \rho = m \frac{N}{V} \), \( \rho_s = m \frac{N_o}{V} \), and \( \rho_n = m \frac{2.612}{(2\pi \hbar^2)^{\frac{3}{2}}} (mkT)^{\frac{3}{2}} \), where \( m \) is the mass of a helium atom, then \( \rho = \rho_s + \rho_n \). Here \( \rho_s \) is the density of the combined fluid, \( \rho_n \) the density of the ground state helium or "superfluid" and \( \rho_n \) the density of the excited or "normal" fluid. Thus the ideal Bose gas theory of liquid helium suggests a two fluid model which will be discussed later.
We may now write
\[ N_e = N \left( \frac{T}{T_c} \right)^{3/2}, \]
so that equation (4) becomes
\[ N_e(\tau) = N \left( 1 - \frac{T}{T_c} \right)^{3/2}. \]

This is the desired relation for the ground state occupancy as a function of temperature.

B3.2d *Feynman Theory of the Ground State.* If one were to solve completely the liquid helium problem, then it would be necessary to solve the corresponding Schrodinger equation: namely
\[ \frac{-\hbar^2}{2m} \sum_i \nabla^2 \psi + \sum_{ij} V(x_{ij}) \psi = E \psi. \]

Clearly this equation cannot be solved directly but a number of the important properties of the ground state can be deduced by indirect and general arguments. The following discussion is an outline of Feynman's arguments for the properties of the ground state and the form of the corresponding wave function.

Basically the wave function is a complicated function of the coordinate positions of all the atoms in the system. For any given positional configuration the wave function will yield a number, the square of whose modulus is proportional to the probability of the configuration.

The ground state wave function will have no nodes for Bose particles and will everywhere be positive. The pairwise potential between atoms has a Lennard-Jones form but the minimum is shifted outward from what one would expect due to the zero point energy. The greatest amplitudes will be found when the density distribution over the liquid is reasonably constant. Macroscopic fluctuations in the density correspond to soundwaves which can be expressed in terms of a set of simple harmonic normal modes. Translated into
quantum mechanical terms, these modes correspond to the zero-point motions of quantum oscillators each of whose wave function is a Gaussian. Therefore the amplitude for a density fluctuation falls off exponentially with the square of the fluctuation.

Because of the strong repulsion between atoms for distances less than the "hard sphere" diameter of about \(2.7\ \bar{A}\) the wave function must fall rapidly toward zero. To get a qualitative idea of what the ground state wave function \(\Phi\) must look like, consider two helium atoms in a one dimensional system of atoms having some average spacing. If we choose one coordinate axis for the position of one atom relative to its equilibrium position and the second coordinate axis for the position of the second atom relative to the first one's origin, and the third coordinate for the amplitude of the combined wave function, we obtain the ground state characterization shown in figure 1.

B3.3a **Lattice Vibrations.** Normal Modes: When looking at the ground state of He-4 below the \(\lambda\) -point, the first model considered was that of a non-interacting Bose gas. Comparison of predictions by this model with the results of specific heat experiments, for example, shows only a qualitative resemblance. Obviously a better model is needed and it is clear that somehow interactions between atoms must be included.
One way to include interactions is to take an extreme model in the opposite direction; namely, a solid in which all atoms have well defined equilibrium positions and in which only small departures from these equilibrium positions are allowed.

First let the potential energy of the solid be given by

\[ U = U(q_1, q_2, \ldots, q_N) \quad \text{where} \quad q_i = q_{0i} + \eta_i, \quad q\_ = q_{0\_} + \eta_\text{a}, \ldots \]

Next, expand \( U \) about the equilibrium coordinates to get

\[ U(q_1, q_2, \ldots, q_N) = U(q_{01}, q_{02}, \ldots, q_{0N}) + \sum \frac{\partial^2 U}{\partial q_i \partial q_k} \eta_i \eta_k + \cdots \]

Since the expansion is about the equilibrium position

\[ \frac{\partial U}{\partial q_i} \bigg|_{0} = 0 \]

and we may arbitrarily choose \( U(q_{01}, q_{02}, \ldots, q_{0N}) = 0 \).

Then,

\[ U(q_1, q_2, \ldots, q_N) \approx \frac{1}{2} \sum_{jk} \frac{\partial^2 U}{\partial q_j \partial q_k} \eta_j \eta_k \]

if we neglect higher order terms. Letting \( U_{jk} = \frac{\partial^2 U}{\partial q_j \partial q_k} \)

the last result becomes

(1) \[ U(q_1, q_2, \ldots, q_N) \approx \sum_{jk} U_{jk} \eta_j \eta_k. \]

Now, consider the total kinetic energy of the solid. This can be written

\[ T(q_1, \dot{q}_1, \ldots, q_N, \dot{q}_N) = \sum_{i} \frac{1}{2} m_i \dot{q}_i^2. \]

Next make a coordinate transformation to a new set of coordinates defined by

\[ q'_i = q'_i(q_1, q_2, \ldots, q_N, t), \quad i = 1, 2, \ldots, N. \]

We have then

\[ dq'_i = \frac{\partial q'_i}{\partial q_1} dq_1 + \frac{\partial q'_i}{\partial q_2} dq_2 + \cdots + \frac{\partial q'_i}{\partial q_N} dq_N + \frac{\partial q'_i}{\partial t} dt. \]

Furthermore,

\[ \dot{q}'_i = \frac{\partial q'_i}{\partial q_1} \dot{q}_1 + \frac{\partial q'_i}{\partial q_2} \dot{q}_2 + \cdots + \frac{\partial q'_i}{\partial q_N} \dot{q}_N + \frac{\partial q'_i}{\partial t}. \]
Squaring we get
\[
(\dddot{q}_i)^2 = \sum_j \frac{\partial q_j}{\partial \xi_j} \frac{\partial q_j}{\partial \xi_j} \dddot{q}_j \dddot{q}_k + 2 \sum_j \frac{\partial q_j}{\partial \xi_j} \dddot{q}_j \frac{\partial q_j}{\partial \xi_k} + (\frac{\partial q_j}{\partial \xi_k})^2
\]

If we define
\[
T_{jk} = \sum_i m_i \frac{\partial q_i}{\partial \xi_k} \frac{\partial q_i}{\partial \xi_j}, \quad T_j = \sum_i m_i \frac{\partial q_i}{\partial \xi_j} \dddot{q}_i, \quad T_0 = \sum_i m_i \left(\frac{\partial q_i}{\partial \xi_k}\right)^2
\]
and if \( q_j = q_0 + \xi_j, \quad \dot{q}_j = \dddot{q}_j, \quad \dddot{q}_j \neq \dddot{q}_i (t) \) then \( T_j = T_0 = 0 \), and we can write
\[
(2) \quad T = \frac{1}{2} \sum_{jk} T_{jk} \dddot{q}_j \dddot{q}_k = \frac{1}{2} \sum_{jk} T_{jk} \dddot{q}_j \dddot{q}_k.
\]

Subtracting equation (1) from (2) we obtain for the Lagrangian,
\[
L = T - V = \frac{1}{2} \sum_{jk} \left( T_{jk} \dddot{q}_j \dddot{q}_k - U_{jk} \eta_j \eta_k \right).
\]
The equations of motion are
\[
\frac{d}{dt} \frac{\partial L}{\partial \eta_j} - \frac{\partial L}{\partial \dddot{q}_j} = 0, \quad \text{for } j = 1, 2, \ldots, 3N,
\]
or,
\[
\frac{d}{dt} \sum_k \left( T_{jk} \dddot{q}_k \right) + \sum_k U_{jk} \eta_k = 0,
\]
or,
\[
\sum_k \left[ T_{jk} \dddot{q}_k + U_{jk} \eta_k \right] = 0.
\]
Assume solutions of the form
\[
\eta_k = C a_\omega e^{i \omega t}
\]
and substitute. This gives
\[
\sum_k U_{jk} a_\omega T_{jk} a_\omega \omega^2 = 0,
\]
or
\[
(3) \quad \sum_k [U_{jk} - \omega^2 T_{jk}] a_\omega = 0, \quad j = 1, 2, \ldots, 3N.
\]
By requiring that the characteristic determinant vanish, a set of \( 3N \) frequencies can be found. Assuming non-degeneracy
the \(3N\) equations (3) can then be solved for the \(3N\) eigenvectors, and the modal matrix formed. Using the modal matrix in a congruence transformation, \([U_{jk}]\) is transformed to diagonal form where the diagonal elements are the squares of the normal frequencies. \([T_{jk}]\) is transformed to the unit matrix by the same transformation. We obtain as a consequence of these operations:

\[
\eta_i = \frac{1}{\sqrt{2}} a_{ij} \xi_j \\
V = \frac{1}{2} \sum_k \omega_k^2 \xi_k^2 \\
T = \frac{1}{2} \sum_k \xi_k^2
\]

The Hamiltonian is then

\[
H' = T + V = \frac{1}{2} \sum_{k=1}^{3N} \left[ \xi_k^2 + \omega_k^2 \xi_k^2 \right].
\]

If we had not set \(U(q_0, q_1, ..., q_{3N}) = 0\), the Hamiltonian would have been written,

\[
(4) \quad H = V_0 + \frac{1}{2} \sum_{k=1}^{3N} \left[ \xi_k^2 + \omega_k^2 \xi_k^2 \right] = V_0 + \frac{1}{2} \sum_k \xi_k H_k.
\]

where \(V_0\) is the equilibrium potential energy and \(H_k\) is the one-dimensional simple harmonic oscillator Hamiltonian.

**B3.3b Thermodynamic Relations.** In the discussion of the phonon modes of He–4 we found the expression for the Hamiltonian to be

\[
H = V_0 + \frac{1}{2} \sum_k \xi_k H_k.
\]

As was pointed out, this corresponds to the Hamiltonian of a set of \(3N\) simple harmonic oscillators. It follows that the energy is given by

\[
E_{ph} = V_0 + \frac{3N}{\hbar \omega_r} (\frac{1}{2} + n_r) \hbar \omega_r
\]

\[
= V_0 + \frac{1}{2} \sum_r \xi \hbar \omega_r + \sum_r n_r \hbar \omega_r
\]

or

\[
E_{ph} = -N \eta + \sum_r n_r \hbar \omega_r,
\]

(1)
where \(-N\eta = V_0 + \frac{1}{2} \sum_{r=1}^{N} \hbar \omega_r\). \(N\) is the total number of atoms and \(\eta\) is called the affinity. It corresponds to the effective binding energy of an atom at \(T=0^\circ\text{K}\).

Using equation (1) we may now calculate the partition function \(Z\),

\[
Z = \sum_{\text{all modes}} e^{-\beta \left[-N\eta + \sum_{r=1}^{N} \hbar \omega_r \right]}
\]

\[
= e^{\beta N\eta} \sum_{\text{all occupancies}} e^{-\beta \left[\sum_{r=1}^{N} \hbar \omega_r \right]}
\]

\[
= e^{\beta N\eta} \left( \frac{1}{1-e^{-\beta \hbar \omega_1}} \right) \left( \frac{1}{1-e^{-\beta \hbar \omega_2}} \right) \cdots
\]

Taking the natural logarithm we obtain the working expression

\[
\ln Z = \beta N\eta - \sum_{r=1}^{N} \beta \hbar \omega_r
\]

We may find the average energy by making use of the formula \(\bar{E} = -\frac{\partial}{\partial \beta} \ln Z\). Thus

\[
\bar{E}_{ph} = -N\eta + \sum_{r=1}^{N} \frac{\hbar \omega_r}{e^{\beta \hbar \omega_r} - 1}
\]

Because the number of modes is very large and because they are closely spaced, we may replace the summation by an integration and write,

\[
\bar{E}_{ph} = -N\eta + \int_0^\infty \frac{e^{-\beta \hbar \omega} \hbar \omega \sigma(\omega) d\omega}{1 - e^{-\beta \hbar \omega}}
\]

where \(\sigma(\omega)\) is the mode density in the angular frequency interval from \(\omega\) to \(\omega + d\omega\). In the Debye approximation, the approach is to assume that \(\sigma(\omega)\) has the same form as that for a non-interacting gas. Furthermore, the number of modes is constrained by the number of degrees of freedom of the substance. Thus,

\[
\sigma(\omega) d\omega = \frac{V}{2\pi^2 c^2} \omega^2 d\omega
\]
subject to the condition,
\[ \int \sigma(\omega) d\omega = \int_0^{\omega_d} \frac{V}{2\pi^2c^2} \omega^2 d\omega = 3N, \]
where \( N \) is the number of atoms. To find the maximum frequency \( \omega_d \), we integrate this last expression and obtain

(4) \[ \omega_d = C \left( \frac{18\pi^2 N}{V} \right)^{1/3}. \]

Substituting these results into equation (3) yields

\[ \bar{E}_p = -N\gamma + \frac{V}{2\pi^2c^2} \int_0^{\omega_d} \frac{\omega^3 d\omega}{(e^{\beta\omega} - 1)} \]
\[ = -N\gamma + \frac{V}{2\pi^2c^2} \int_0^{\omega_d} \frac{(\beta\omega)^3}{(e^{\beta\omega} - 1)} d(\beta\omega). \]

For low temperatures \( \beta\hbar \omega_d \) becomes very large and the contribution from the integral at the high frequencies negligible. Therefore we may replace the upper limit by \( \infty \) and write

\[ \bar{E}_p = -N\gamma + \frac{V}{2\pi^2\hbar^3c^2\beta^4} \int_0^{\infty} \frac{\xi^3 d\xi}{(e^\xi - 1)}. \]

where \( \xi = \beta\hbar\omega \). If we solve equation (4) for \( V \), we may substitute the result into the last equation and obtain

\[ \bar{E}_p = -N\gamma + \frac{18\pi^2c^3N}{\omega_d^3} \frac{1}{2\pi^2\hbar^3c^2\beta^4} \int_0^{\infty} \frac{\xi^3 d\xi}{(e^\xi - 1)} \]
\[ = -N\gamma + \frac{9N}{\hbar^3\omega_d^3\beta^4} \int_0^{\infty} \frac{\xi^3 d\xi}{(e^\xi - 1)}. \]

Defining \( \frac{\hbar^3\omega_d^3}{\beta^4} = k_B^3 \Theta_D^3 \), where \( \Theta_D \) is called the Debye temperature, we may write

(5) \[ \bar{E}_p = -N\gamma + 9Nk_B \frac{\gamma^4}{\Theta_D^3} \int_0^{\infty} \frac{\xi^3 d\xi}{(e^\xi - 1)}. \]
Now consider the integral itself:

\[
\int_{0}^{\infty} \frac{\xi^3 d\xi}{(e^\xi - 1)} = \int_{0}^{\infty} \frac{e^{-\xi}}{(1 - e^{-\xi})} d\xi
\]

\[
= \int_{0}^{\xi} e^{-\xi} (1 + e^{-\xi} + e^{-2\xi} + \ldots) d\xi
\]

\[
= \sum_{n=1}^{\infty} \int_{0}^{\xi} e^{-n\xi} d\xi
\]

\[
= \sum_{n=1}^{\infty} -\frac{d^3}{dn^3} \int_{0}^{\xi} e^{-n\xi} d\xi
\]

\[
= \sum_{n=1}^{\infty} -\frac{d^3}{dn^3} \left[ -\frac{e^{-n\xi}}{n} \right]_{0}^{\infty} = \sum_{n=1}^{\infty} -\frac{d^3}{dn^3} \left( \frac{1}{n} \right)
\]

\[
= \sum_{n=1}^{\infty} \frac{3!}{n^4} = \frac{\pi^4}{15}.
\]

Substituting into equation (5) we get

\[
\overline{E}_{ph} = -N\eta + 9Nk_{B} \frac{\pi^4}{15} \left( \frac{T^4}{\Theta^3} \right)
\]

or

\[
(6) \quad \overline{E}_{ph} = -N\eta + \frac{3\pi^4 Nk_{B}}{5} \left( \frac{T^4}{\Theta^3} \right).
\]

The specific heat is found from the formula \( C_{v} = \left( \frac{\partial \overline{E}}{\partial T} \right)_{v} \).

Therefore,

\[
(7) \quad C_{vph} = \frac{12\pi^4 Nk_{B}}{5} \left( \frac{T^3}{\Theta^3} \right).
\]

We may find the entropy from the relation

\[
S_{ph} = \int_{0}^{T} C_{v} \frac{dT}{T}.
\]

Thus

\[
S_{ph} = \int_{0}^{T} \frac{12\pi^4 Nk_{B}}{5} \left( \frac{T^2}{\Theta^3} \right) dT
\]

or

\[
(8) \quad S_{ph} = \frac{4\pi^4 Nk_{B}}{5} \left( \frac{T^3}{\Theta^3} \right).
\]
B3.3c Feynman Theory of Phonon States. One possible type of excitation in liquid helium is that of longitudinal sound waves. We have already discussed this somewhat in the section on lattice vibrations. Each possible state of excitation corresponds to one of the modes of the system. Each mode can be thought of as a set of one or more oscillators, the number of which is determined by the phonon occupancy.

Now the density \( \rho(\vec{R}) \) of the liquid at any coordinate position \( \vec{R} \) can be expressed as a superposition of waves in k-space; that is, as a superposition of normal modes with appropriate amplitudes. Explicitly,

\[
\rho(\vec{R}) = \frac{1}{(2\pi)^3} \int q_{\vec{k}} e^{-i \vec{k} \cdot \vec{R}} d^3k,
\]

where

\[
q_{\vec{k}} = \int \rho(\vec{R}) e^{i \vec{k} \cdot \vec{R}} d^3R.
\]

In this representation \( q_{\vec{k}} \) is the classical normal coordinate corresponding to the mode \( \vec{k} \).

In quantum mechanics the wave functions for a harmonic oscillator can be expressed as the product of a normalizing factor, a Hermite polynomial, and a Gaussian exponential. The Hermite polynomial for the ground state is unity and for the first excited state, simply a coordinate. Thus, if \( \Phi \) is the ground state wave function of liquid helium, the first excited phonon state must have a wave function of the form,

\[
\psi_{\text{ph}} = q_{\vec{k}} \Phi.
\]

Noting that the density function must have the form

\[
\rho(\vec{R}) = \frac{1}{i} \delta(\vec{R} - \vec{R}'),
\]

we can write the first excited state as, using equations (2), (3), and (4),

\[
\psi_{\text{ph}} = \left[ \frac{1}{i} e^{i \vec{k} \cdot \vec{R}} \right] \Phi.
\]
where normalizing factors are omitted. This equation is valid provided the phonon wave length is much larger than the atomic spacing.

Feynman next argues that phonon states are the lowest possible excited states. Since the energy of the first excited phonon states is \( \hbar \omega = \frac{hc}{\lambda} \) (omitting the zero-point energy) and since \( \lambda \) can be quite large, it is evident that the energies can be quite low. These excited states are the result of long range density variations; in fact any pure long range density variation can be described in terms of phonon modes using equation (1). Therefore, if there are other types of excitations, they must involve changes in the atomic configuration that leave the density unaffected; for example, stirring changes. Furthermore, for a given state, the change in configuration required to go from maximum positive to maximum negative probability amplitude must take place over the largest possible distances; for example, by moving atoms all the way across a macroscopic container. If this is not required then the energy of the state will be too high to equal the energy of the low lying phonon states. But, and this is the crux of Feynman's argument, any long range exchange can be replaced by an equivalent short range exchange through appropriate permutations of the atoms. Such short range exchanges are indistinguishable from the corresponding long range ones due to the Bose properties of the He-4 atoms; namely, symmetry under exchange and indistinguishability. The net result is that the change in configuration from maximum positive to maximum negative can be accomplished by moving no atom on the average more than half the local atomic spacing. Thus there cannot be other types of excitations having comparable energies to the phonon excitations.

B3.4a **Thermodynamic Relations for Rotons.** Starting with Landau's postulate for the energy spectrum of rotons,
\[ E_{\text{rot}} = \Delta + \frac{(p-p_0)^2}{2\mu} , \]

we may derive the basic thermodynamic relations. The theory is not as complete as that for phonons since we don't have the equivalent of the harmonic oscillator model from which to start. It is necessary, therefore, to assume that the number of rotons is not bounded although, most probably, the number of possible states is. Because of the energy gap, \( \Delta \), it is not necessary to assume that rotons are Bose particles since the statistics are effectively Boltzmann. Let us begin by evaluating the partition function.

Assuming either Bose-Einstein or Fermi statistics we may write for the logarithm of the partition function,

\[ \ln Z = \mp \sum \ln \left[ \frac{1}{2} e^{-\beta \left( \frac{(p-p_0)^2}{2\mu} \right) } \right] , \]

where the (-) sign corresponds to Bose-Einstein statistics and the (+) sign to Fermi statistics. By an argument similar to that used in the discussion of the Debye approximation, we may convert the summation to an integration, the upper limit of which is \( \infty \). Thus,

\[ \ln Z \approx \mp \frac{4\pi V}{(2\pi \hbar)^3} \int_0^\infty \ln \left[ \frac{1}{2} e^{-\beta \left( \frac{(p-p_0)^2}{2\mu} \right) } \right] p^2 dp . \]

If the temperatures are low, then \( e^{-\beta \left( p-p_0 \right)^2} \ll 1 \) and the logarithm may be expanded to yield

\[ \ln Z \approx \frac{4\pi V}{(2\pi \hbar)^3} \int_0^\infty e^{-\beta \left( p-p_0 \right)^2} \frac{p^2 dp}{\mu} . \]

It is convenient now to make a change of variable, \( p' = p - p_0 \). We then have

\[ \ln Z \approx \frac{4\pi V}{(2\pi \hbar)^3} \int_{-p_0}^\infty e^{-\beta \left( p' \right)^2} \left( p' + p_0 \right)^2 dp' . \]

However since the density of rotons is known experimentally to be a maximum at \( p = p_0 \) and to fall off rapidly on either side, we may replace the lower limit by \( -\infty \). Thus,
\[
\ln Z = \frac{4\pi V}{(2\pi \hbar)^3} \int_{-\infty}^{\infty} e^{-\beta (\frac{p^2}{2\mu} + \frac{p_0^2}{2})} (p' + p_0)^2 dp'
\]

\[
= \frac{4\pi V}{(2\pi \hbar)^3} e^{-\beta \Delta} \int_{-\infty}^{\infty} e^{-\beta \frac{p'^2}{2\mu}} [p'_2 + p_0^2] dp'
\]
since the odd term vanishes. Proceeding with the integration we obtain,

\[
\ln Z \approx \frac{4\pi V}{(2\pi \hbar)^3} e^{-\beta \Delta} \left[ (-2\mu) \frac{d}{d\beta} + \frac{p_0^2}{\beta} \right] \int_{-\infty}^{\infty} e^{-\beta \frac{p^2}{2\mu}} dp'
\]

\[
= \frac{4\pi V}{(2\pi \hbar)^3} e^{-\beta \Delta} \left[ (-2\mu) \frac{d}{d\beta} + \frac{p_0^2}{\beta} \right] \frac{2\mu}{\beta} \frac{1}{\sqrt{\pi}}
\]

\[
= \frac{4\pi V}{(2\pi \hbar)^3} e^{-\beta \Delta} \left[ -2\mu \left[ 2\mu \right] \frac{1}{\beta^\frac{3}{2}} + p_0^2 \left[ \frac{2\mu}{\beta} \right]^\frac{1}{2} \right]
\]

\[
= \frac{4\pi V}{(2\pi \hbar)^3} e^{-\beta \Delta} \left[ \frac{2\mu}{\beta} \right]^\frac{1}{2} \left[ \frac{1}{8} + \frac{p_0^2}{2\mu} \right].
\]

For temperatures less than \( \sim 2^\circ K \), \( \frac{p_0^2}{2\mu} \gg 15 \) so that we may write approximately:

(1) \[
\ln Z \approx \frac{k_B^2 p_0^2 V}{(2\pi \hbar)^3 \hbar^3} \left( \frac{2\mu}{\beta} \right)^{\frac{1}{2}} e^{-\beta \Delta} = KV \left[ \frac{e^{-\beta \Delta}}{\beta^{\frac{3}{2}}} \right].
\]

The average energy is given by the formula \( -\frac{\beta}{\partial \beta} \ln Z \), so that

\[
E_{rot} = -KV \frac{\partial}{\partial \beta} \left[ \frac{e^{-\beta \Delta}}{\beta^{\frac{3}{2}}} \right] = KV \left[ \frac{4}{\beta^{\frac{3}{2}}} e^{-\beta \Delta} + \frac{1}{2} \frac{e^{-\beta \Delta}}{\beta^{3\frac{1}{2}}} \right]
\]

\[
= KV \frac{e^{-\beta \Delta}}{\beta^{\frac{3}{2}}} \left[ \Delta + \frac{1}{2} \beta \right]
\]

(2) \[
= k_B T \ln Z \left[ \frac{4}{k_B^2} + \frac{1}{2} \right].
\]

The specific heat is given by the formula \(-k_B^2 (\frac{\partial E}{\partial \beta})\), so that

\[
C_{\text{rot}} = -k_B \beta^2 \frac{\partial}{\partial \beta} \left[ \left( KV \frac{e^{-\beta \Delta}}{\beta^{\frac{3}{2}}} \right) \left[ \Delta + \frac{1}{2} \beta \right] \right],
\]
This last expression can be written also as

\[ C_{\text{rot}} = k_B \ln \mathbb{Z} \left[ \frac{\Delta^2}{(\beta T)^2} + \frac{\Delta}{\beta T} + \frac{3}{4} \right]. \tag{3} \]

Next, we solve for the entropy. Since \( S = -\frac{\partial F}{\partial T} = k_B \beta^2 \frac{\partial F}{\partial \beta} \), we can write,

\[ S_{\text{rot}} = k_B \beta^2 \frac{\partial F}{\partial \beta} = k_B \beta^2 \frac{\partial}{\partial \beta} \left[ -kV \frac{e^{-\beta \Delta}}{\beta^{2/3}} \right] \]

\[ = k_B \beta^2 KV \frac{e^{-\beta \Delta}}{\beta^{2/3}} \left[ \frac{1}{2} + \frac{3}{2\beta^2} \right] \]

or

\[ S_{\text{rot}} = k_B \ln \mathbb{Z} \left[ \frac{A}{\beta T} + \frac{3}{2} \right]. \tag{4} \]

B3.5a Superfluidity.\(^5,9\) Suppose we have liquid helium at \( T=0^\circ \text{K} \) flowing through a capillary with velocity \( V \). Also, suppose that there is an interaction between the fluid and the walls such that an elementary excitation is formed in the liquid. The question is: what is the energy of such an excitation when viewed by an observer in the laboratory frame?

At first consider a coordinate system moving with the superfluid. In this frame the energy of an excitation is given by

\[ \varepsilon_o = \varepsilon(p), \]
and the momentum by
\[ p_0 = p. \]

If \( v_s \) is the velocity of the moving frame relative to the laboratory frame then, as a consequence of the excitation, the change in energy when determined by a laboratory observer is

\[ \epsilon_L = \epsilon(p) + p \cdot v_s. \]

(1)

To the laboratory observer the average momentum density is given by
\[ j = p_0 v_s + \langle p \rangle. \]

Now if we consider the equilibrium case where the excitations are in equilibrium with the walls, that is with the laboratory frame, then \( \epsilon_L \) is a minimum. This implies that \( \langle p \cdot v_s \rangle \) is negative; in fact, if \( p = -v_s \), \( \epsilon_L \) is maximum negative.

Using this observation we may define \( p_n \) by
\[ \langle p \rangle = -p_n v_s. \]

Further, defining \( p = p_0 - p_n \), we find for the total current when the excitations are in equilibrium with the walls,
\[ j = (p_n + p_n) v_s - p_n v_s = p_n v_s. \]

If the excitations are not in equilibrium with the walls but are drifting relative to them with a velocity \( v_n \), then the current becomes
\[ j = p_n v_s - p_n (v_s - v_n) = (p_n + p_n) v_s - p_n v_s + p_n v_n \]

or
\[ (2) \quad j = p_n v_s + p_n v_n. \]

This equation is one of the cornerstones of the two fluid theory.

Returning to equation (1) it is evident that for an excitation to be favorable
\[ \epsilon(p) + p \cdot v_s < 0. \]
The left hand side is minimum when the excitation momentum, \( \mathbf{p} \), and the superfluid velocity, \( \mathbf{V}_s \), are oppositely directed. In this case we find that the minimum velocity for excitation is

\[
|\mathbf{V}_s| \approx \frac{\epsilon(p)}{\rho}.
\]

This result can be derived from a different and somewhat more physical point of view. Consider flow through a capillary again, and choose the liquid as the fixed frame of reference. Then the capillary walls are moving. Suppose that there is a roughness as shown in figure 1, which gives rise to an excitation in the liquid. Conservation of energy gives:

\[
\frac{1}{2} M_v V_i^2 = \frac{1}{2} M_v V_o^2 + \epsilon(p)
\]

or

\[
\frac{1}{2} M_v V_o^2 = \frac{1}{2} M_v V_i^2 - \epsilon(p)
\]

where

- \( M_v \) is the wall mass,
- \( V_i \) is the initial wall velocity,
- \( V_o \) is the final wall velocity, and
- \( \epsilon(p) \) is the excitation energy.

Conservation of momentum requires

\[
M_v V_i = M_v V_o + \mathbf{p}.
\]

Solving for \( M_v V_o \), squaring and dividing through by \( 2M_v \) we obtain,

\[
\frac{1}{2} M_v V_o^2 = \frac{1}{2} M_v V_i^2 + \frac{|\mathbf{p}|^2}{2M_v} - \mathbf{p} \cdot \mathbf{V}_i.
\]

Subtracting equation (6) from (4) we get

\[
\epsilon(p) = -\frac{|\mathbf{p}|^2}{2M_v} + \mathbf{p} \cdot \mathbf{V}_i.
\]

Rearranging we have

\[
\mathbf{p} \cdot \mathbf{V}_i = \epsilon(p) + \frac{|\mathbf{p}|^2}{2M_v}.
\]
The minimum velocity $V_i$, for which this equation can be satisfied, is obtained when $\mathbf{p}$ and $\mathbf{v}$ are parallel. Thus,

$$|V_i|_{\text{min}} = \left[ \frac{\epsilon(p) + \frac{1}{2 \mathcal{M}_0}}{|p|^2} \right] / |p|.$$

Because $M_0$ is very large this expression becomes to a good approximation

$$|V_i|_{\text{min}} \approx \frac{\epsilon(p)}{|p|^2}$$

which is the same result obtained in equation (3) since

$$|V_i|_{\text{min}} = |V_s|.$$

Suppose the excitation is a phonon. Then the energy is given by

$$\epsilon(p) = p V_{\text{ph}}$$

so that

$$V_{i \text{min}} = \frac{p V_{\text{ph}}}{p} = V_{\text{ph}}.$$

If the excitation is a roton, then

$$V_{i \text{min}} = \left( \Delta + \frac{(p-p_0)^2}{2 \mu} \right) / p.$$

To find the absolute minimum, differentiate with respect to $p$, and set the result equal to zero.

$$\frac{dV_{i \text{min}}}{dp} = -\frac{\Delta + (p-p_0)^2}{p^2} + \frac{2(p-p_0)}{2 \mu p} = 0$$

$$= -2 \mu \Delta + p^2 - p_0^2 = 0$$

so

$$p = \frac{(p_0^2 + 2 \mu \Delta)}{p_0^2}.$$

However since $2 \mu \Delta \sim 5 \times 10^{-4}$ and $p_0^2 \sim 4 \times 10^{-38}$, we have that $p \sim p_0$ for a minimum. Substituting we get

$$V_{\text{min}} \sim \frac{\Delta}{p_0}.$$

It appears from the above results that unless the velocity of relative motion exceeds a threshold value of about $\epsilon(p)/p$, there are no excitations in helium and the fluid remains superfluid. In actuality, another mechanism contributes to the generation of excitations; namely, the formation of vortices, so that the critical velocities turn out
to be much lower than those predicted by $E^{(p)}/\rho$. This will be discussed in a later section.

B3.5b Equation of Continuity.\(^\text{10}\) Consider a control volume $V$ of fluid. If elements of the fluid are in motion, then the flow out of the volume can be written

$$\int \rho \mathbf{v} \cdot d\mathbf{A}$$

This flow leads to a loss of mass per unit time given by

$$-\frac{d}{dt} \int_{\text{vol}} \rho \, dv$$

Equating these two results we get

$$-\frac{d}{dt} \int_{\text{vol}} \rho \, dv = \int_{\text{surface}} \rho \mathbf{v} \cdot d\mathbf{A}.$$  

Using Gauss's theorem we can transform the right hand integral to a volume integral and obtain

$$-\frac{d}{dt} \int_{\text{vol}} \rho \, dv = \int_{\text{vol}} \mathbf{V} \cdot \rho \mathbf{v} \, dv$$

or,

$$\int_{\text{vol}} \left[ \frac{\partial \rho}{\partial t} + \mathbf{V} \cdot \rho \mathbf{v} \right] \, dv = 0.$$  

Since the result must hold for any volume, it follows that

$$\frac{\partial \rho}{\partial t} + \mathbf{V} \cdot \rho \mathbf{v} = 0.$$  

This equation is called the Equation of Continuity and, if we define the mass flux density by $\mathbf{j} = \rho \mathbf{v}$, we can rewrite it as

$$\frac{\partial \rho}{\partial t} + \mathbf{V} \cdot \mathbf{j} = 0.$$  

B3.5c Adiabatic Equation.\(^\text{10}\) In an ideal fluid it is assumed that the motion is adiabatic; that is, there is no heat exchange between parts. Therefore, the entropy for any part is considered constant (isentropic).

If $s$ is the entropy per unit mass, then the isentropic
condition can be expressed by
\[ \frac{ds}{dt} = 0. \]
Now, \( S = S(\mathbf{r}, \mathbf{v}, t) \) in general so
\[ \frac{ds}{dt} = \frac{\partial S}{\partial t} + \nabla \cdot \mathbf{v} \]
or
\[ (1) \quad \frac{ds}{dt} = \frac{\partial S}{\partial t} + (\mathbf{v} \cdot \nabla) S = 0. \]

By an exactly analogous argument to that used in the derivation of mass continuity, we may derive the equation of entropy continuity. We get
\[ (2) \quad \frac{\partial s}{\partial t} + \nabla \cdot \mathbf{\rho} = 0. \]

B3.5d Euler's Equation. Consider some volume element of a fluid and the external pressure acting on it. The total force is obtained by integrating over the surface. Thus
\[ F = - \int \mathbf{p} d\mathbf{A}. \]
The negative sign is a convention and expresses the fact that the pressure on the volume element is inwardly directed whereas the surface normal is outward.

By the usual methods, this integral may be transformed to a volume integral
\[ -\int \mathbf{p} d\mathbf{A} = -\int \mathbf{v} d\mathbf{V} \]
We may interpret this equation by saying that the pressure exerts a force on a unit volume of fluid of \(-\mathbf{v}\mathbf{p}\). Equating the inertial response of this unit volume to the force yields the equation of motion,
\[ (1) \quad \rho \frac{dv}{dt} = -\nabla \mathbf{p} \]
where \( \rho \) is the mass of the unit volume; that is, the density.

However, \( \mathbf{v} = \mathbf{v}(\mathbf{r}, \mathbf{v}, t) \) so that
\[ \frac{dv}{dt} = \frac{\partial \mathbf{v}}{\partial t} + \frac{\partial \mathbf{v}}{\partial x} v_x + \frac{\partial \mathbf{v}}{\partial y} v_y + \frac{\partial \mathbf{v}}{\partial z} v_z \]
\[ = \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} . \]
Substituting this result into equation (1) we obtain Euler's equation for an incompressible fluid

\[ \frac{\partial \mathbf{V}}{\partial t} + (\mathbf{V} \cdot \nabla) \mathbf{V} = -\frac{1}{\rho} \nabla P. \]

B3.5e Momentum Flux. The \( i \)th component of the mass flux density or momentum per unit volume is given by

\[ \rho \mathbf{v}_i. \]

The time rate of change of this is

\[ \frac{\partial \rho \mathbf{v}_i}{\partial t} = \frac{\partial \rho}{\partial t} \mathbf{v}_i + \rho \frac{\partial \mathbf{v}_i}{\partial t}. \]

From the continuity equation (B3.5b-1), and using the summation convention

\[ \frac{\partial \rho}{\partial t} = -\frac{\partial \rho}{\partial x_k} \mathbf{v}_k. \]

From Euler's equation (B3.5d-2),

\[ \frac{\partial \mathbf{v}_i}{\partial t} = -\mathbf{v}_k \frac{\partial \mathbf{v}_i}{\partial x_k} - \frac{1}{\rho} \frac{\partial P}{\partial x_i}, \]

so that substituting into equation (1) we get,

\[ \frac{\partial \rho \mathbf{v}_i}{\partial t} = -\mathbf{v}_i \cdot \frac{\partial \rho}{\partial x_k} \mathbf{v}_k - \rho \mathbf{v}_k \frac{\partial \mathbf{v}_i}{\partial x_k} - \frac{\partial P}{\partial x_i}. \]

If we define the momentum flux tensor by

\[ \mathbf{T}_{ik} = \rho \delta_{ik} + \mathbf{v}_i \rho \mathbf{v}_k, \]

the last equation can be written

\[ \frac{\partial \rho \mathbf{v}_i}{\partial t} = \frac{\partial \mathbf{T}_{ik}}{\partial x_k}. \]

The interpretation is that momentum flux is conserved.

Thus equation (3) is a continuity equation for momentum flux.

B3.5f Navier-Stokes Equation. The starting point for the derivation of the equation of motion for a viscous
fluid is the momentum flux tensor, \( \Pi_{ik} = \rho \delta_{ik} + V_i \rho V_k - \sigma_{ik}' \).

This tensor represents the reversible transfer of momentum. To take into account transfers of momentum that are irreversible such as phonon or roton generation, etc., a term \(-\sigma_{ik}'\) is added to it. The momentum flux tensor for a viscous fluid then becomes

\[
(1) \quad \Pi_{ik} = \rho \delta_{ik} + V_i \rho V_k - \sigma_{ik}'
\]

where \( \sigma_{ik}' \) is called the viscosity stress tensor. We wish now to deduce the form of \( \sigma_{ik}' \).

Consider first figure 1. An incompressible fluid is contained between two plates, the lower of which is fixed and the upper, moving. The shear force per unit area turns out experimentally to be proportional to the velocity \( V_o \) in the direction of plate motion and inversely proportional to the separation distance between plates. That is,

\[
\frac{\Delta F}{\Delta A} \propto \frac{V_o}{d}.
\]

Using this experimental observation, we define the coefficient of viscosity, \( \eta \), by

\[
\frac{\Delta F}{\Delta A} = \eta \frac{\partial V_x}{\partial y}
\]

In the more complicated three dimensional case, we consider an elemental cube of fluid. Across each face of the cube there is a similar force per unit area. We relate the viscous shear stress, \( \sigma_{ij}' \), to the viscous forces by the definition

\[
(2) \quad \sigma_{ik}' = \eta \left( \frac{\partial V_i}{\partial x_k} + \frac{\partial V_k}{\partial x_i} \right).
\]

This definition has the form such that if

\[
\frac{\partial V_i}{\partial x_k} = -\frac{\partial V_k}{\partial x_i},
\]
then $\sigma_{ik}'$ vanishes, which simply means that the viscous stress vanishes for pure rotation.

Substituting (2) into (1), we obtain for the momentum flux tensor

$$\Pi_{ik} = \rho \delta_{ik} + v_i \rho \frac{\partial v_k}{\partial x_i} - \eta \left( \frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right).$$

Inserting this into the momentum flux conservation equation (B3.5e-3), we get

$$\frac{\partial \rho v_i}{\partial t} = -\frac{\partial}{\partial x_k} \Pi_{ik}$$

$$= -\frac{\partial}{\partial x_k} \left[ \rho \delta_{ik} + v_i \rho v_k - \eta \left( \frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right) \right]$$

$$= -\frac{\partial \rho}{\partial x_i} - \rho \frac{\partial v_k}{\partial x_k} - v_i \frac{\partial \rho v_k}{\partial x_k} + \eta \frac{\partial^2 v_i}{\partial x_k^2} + \eta \frac{\partial}{\partial x_i} \frac{\partial v_k}{\partial x_k}.$$

This may be rearranged to yield

$$\frac{\partial \rho v_i}{\partial t} + \rho \frac{\partial v_k}{\partial x_k} \frac{\partial v_i}{\partial x_k} + v_i \frac{\partial \rho v_k}{\partial x_k} = -\frac{\partial \rho}{\partial x_i} + \eta \frac{\partial^2 v_i}{\partial x_k^2} + \eta \frac{\partial}{\partial x_i} \frac{\partial v_k}{\partial x_k}.$$

We may also write this last result as a vector equation,

$$\frac{\partial \rho \vec{v}}{\partial t} + (\rho \vec{v} \cdot \nabla) \vec{v} + \nabla (\nabla \cdot \rho \vec{v}) = -\nabla p + \eta \nabla^2 \vec{v} + \eta (\vec{v} \cdot \nabla \vec{v}).$$

However, since

$$\nabla \cdot \rho \vec{v} = -\frac{\partial \rho}{\partial t},$$

we have

$$\rho \frac{\partial \vec{v}}{\partial t} + \vec{v} \frac{\partial \rho}{\partial t} + (\rho \vec{v} \cdot \nabla) \vec{v} - \vec{v} \frac{\partial \rho}{\partial t} = -\nabla p + \eta \nabla^2 \vec{v} + \eta (\vec{v} \cdot \nabla \vec{v}),$$

or

$$\rho \frac{\partial \vec{v}}{\partial t} + \rho (\vec{v} \cdot \nabla) \vec{v} = -\nabla p + \eta \nabla^2 \vec{v} + \eta (\vec{v} \cdot \nabla \vec{v}).$$

Since the fluid is assumed incompressible,

$$\nabla \cdot \vec{v} = 0,$$

and we get the Navier-Stokes equation in vector form,

$$\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla) \vec{v} = -\nabla p + \frac{\eta}{\rho} \nabla^2 \vec{v}.$$ (3)
There is a still more general form for the Navier-Stokes equation when the fluid is compressible and we derive this next.

Viscosity arises from the relative motion of one part of a fluid with another. Provided the velocity gradients are small, we expect only the first derivatives of the velocities to be involved. Furthermore, we expect all first derivatives to be included, so that not only do we expect terms of the form

\[ \left( \frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right) \]

but also terms of the form

\[ \left( \frac{\partial v_i'}{\partial x_i} \right). \]

Assuming a linear combination of such terms, we write

\[ \sigma_{ik}' = \eta \left( \frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} \right) + \eta' \frac{\partial v_i'}{\partial x_i}, \]

where \( \eta \) is called the first coefficient of viscosity and \( \eta' \), the second.

Carrying through a similar derivation to that used before, we obtain for the compressible fluid Navier-Stokes equation,

\[ \frac{\partial v}{\partial t} + (v \cdot \nabla)v = -\frac{1}{\rho} \nabla p + \frac{\eta}{\rho} \nabla^2 v + \frac{(\eta + \eta')}{\rho} \nabla (v \cdot v). \]

B3.5g Equations of Motion. The two fluid model is essentially a phenomenological description of liquid helium below the \( \lambda \)-point. The basic postulate is that helium II consists of two components, a normal component and a superfluid component, such that the total density may be written

\[ \rho = \rho_n + \rho_s \]

where \( \rho_n \) is the normal density and \( \rho_s \) is the superfluid density. It is assumed that any viscosity or entropy carried by the helium is carried by the normal component.
If $j$ is the momentum per unit volume and $y_n$, $y_s$ are the respective velocities of the normal and superfluid components, then we can write

$$j = \rho_n y_n + \rho_s y_s.$$  \hspace{1cm} (2)

The flow vector $j$ is connected to $\rho$ through the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot j = 0.$$ \hspace{1cm} (3)

Next, we can relate the pressure $P$ to the velocity through the Euler equation,

$$\frac{\partial v}{\partial t} = -\frac{1}{\rho} \nabla P$$

where $\frac{\partial v}{\partial t}$ is the so-called substantive derivative and is given by

$$\frac{\partial v}{\partial t} = \frac{\partial v}{\partial t} + (v \cdot \nabla) v.$$  \hspace{1cm} (4)

If the velocity is sufficiently small, the quadratic terms in the velocity can be neglected and the second term on the right vanishes. We then have,

$$\frac{\partial \rho v}{\partial t} + \nabla P = \frac{\partial \rho}{\partial t} + \nabla P = 0.$$ \hspace{1cm} (5)

When viscosity and other sources for irreversible energy transformations are neglected, entropy is conserved, so we have from equation (B3.5c-2)

$$\frac{\partial \rho_s s}{\partial t} + \nabla \cdot \rho_s s y_n = 0,$$

where $y_n$ is used instead of $v$ since entropy is carried only by the normal component.

Consider now the forces acting on the superfluid component only when we increase the number of ground state atoms at constant volume. By Newton's second law we may write

$$\rho_s \frac{d v_s}{d t} = -\nabla \phi,$$
where $\Phi$ is the potential energy per unit volume of the superfluid component only. We may rewrite this as

$$\frac{dV_s}{dt} = -\nabla \Phi$$

where $\Phi$ is the potential energy per unit mass of the superfluid. But, the available energy to accelerate the superfluid component is the free energy, so we can write

$$\frac{dV_s}{dt} = -\nabla G,$$

$G$ being the Gibbs free energy per unit mass. However, $\nabla G$ is given by

$$\nabla G = v \nabla P - s \nabla T$$

$$= \frac{1}{\rho} \nabla P - s \nabla T$$

where $v$ is the volume per unit mass and $s$ is the specific entropy. Substituting, we get

$$\frac{dV_s}{dt} = s \nabla T - \frac{1}{\rho} \nabla P$$

or

$$\frac{dV_s}{dt} + (V_s \cdot \nabla) V_s = s \nabla T - \frac{1}{\rho} \nabla P.$$

Multiplying this last equation through by $\rho$, we obtain

$$\frac{d\rho V_s}{dt} + \rho (V_s \cdot \nabla) V_s = \rho s \nabla T - \nabla P.$$

Using equations (1), (2), and (4) this becomes

$$\frac{d\rho V_s}{dt} + \frac{d\rho}{dt} \frac{V_s}{\rho} + \rho (V_s \cdot \nabla) V_s = \rho s \nabla T + \frac{d\rho}{dt} + \frac{\partial \rho V_s}{\partial t}$$

or

$$\rho \frac{d(V_n-V_s)}{dt} + \rho s \nabla T = -\rho (V_s \cdot \nabla) V_s.$$

Finally, assuming small $|V_s|$, the right hand term can be neglected and we get

$$\rho \frac{d(V_n-V_s)}{dt} + \rho s \nabla T = 0.$$
The interesting conclusion is that a temperature gradient produces a separation in the velocities of the normal and superfluid components.

**3.5h First and Second Sound Propagation.** In the previous section we developed a set of equations which defined the two fluid model of helium. They were

\begin{align*}
(1) & \quad \rho = \rho_n + \rho_s \\
(2) & \quad j = \rho_n \nu_n + \rho_s \nu_s \\
(3) & \quad \frac{\partial \rho}{\partial t} + \nabla \cdot j = 0 \\
(4) & \quad \frac{\partial \rho}{\partial t} + \nabla \cdot P = 0 \\
(5) & \quad \frac{\partial \rho_s}{\partial t} + \nabla \cdot \rho_s \nu_n = 0 \\
(6) & \quad \rho_n \frac{\partial}{\partial t} (\nu_n - \nu_s) + \rho_s \nabla T = 0.
\end{align*}

These equations will now be used to derive the equations governing the propagation of first and second sound in helium.

First, differentiate equation (3) with respect to time and combine with equation (4) to obtain

\begin{equation}
\frac{\partial^2 \rho}{\partial t^2} - \nabla^2 P = 0.
\end{equation}

Next, take the divergence of equation (6) to get

\begin{equation}
\rho_n \frac{\partial}{\partial t} \nabla \cdot (\nu_n - \nu_s) + \rho_s \nabla^2 T = 0.
\end{equation}

In order to eliminate \( \nu_n \) and \( \nu_s \), which are not observable, from this last expression, we begin by combining equations (1), (2), and (3) as follows:

\[ j = \rho \nu_n - \rho_s \nu_n + \rho_s \nu_s. \]
\[ \nabla \cdot j = \rho \nabla \cdot V_n - \rho_s [ \nabla \cdot (V_n - V_s)] , \]

(9) \[ \nabla \cdot (V_n - V_s) = \frac{1}{\rho_s} \left[ \frac{\partial P}{\partial t} + \rho \nabla \cdot V_n \right] . \]

Next consider the term \( \nabla \cdot \rho_S V_n \) in equation (5). Written out this is

\[ \nabla \cdot \rho_S V_n = \frac{\partial \rho_S V_{nx}}{\partial x} + \frac{\partial \rho_S V_{ny}}{\partial y} + \frac{\partial \rho_S V_{nz}}{\partial z} , \]

where, for example, \( \frac{\partial \rho_S V_{nx}}{\partial x} = \frac{\partial \rho_S}{\partial x} V_{nx} + \rho_S \frac{\partial V_{nx}}{\partial x} . \)

Assuming that \( \frac{\partial \rho_S}{\partial x} V_{nx} \) is very small in comparison to \( \rho_S \frac{\partial V_{nx}}{\partial x} \), we may neglect the first term on the right. Therefore, generalizing, \( \nabla \cdot \rho_S V_n \approx \rho_S \nabla \cdot V_n \), and equation (5) becomes

or

\[ \frac{\partial \rho_S}{\partial t} + \rho_S \nabla \cdot V_n = 0 \]

\[ -\frac{\partial \rho}{\partial t} - \frac{\rho}{\rho_S} \frac{\partial \rho_S}{\partial t} \approx \rho \nabla \cdot V_n . \]

Substituting this result into equation (9), we get

\[ \nabla \cdot (V_n - V_s) = -\frac{\rho}{\rho_S} \frac{\partial \rho_S}{\partial t} \]

and inserting into equation (8),

\[ \rho \frac{\partial}{\partial t} \left[ -\frac{\rho}{\rho_S} \frac{\partial \rho_S}{\partial t} \right] + \rho_S \nabla^2 T = 0 , \]

or approximately,

(10) \[ \frac{\partial^2 \rho_S}{\partial t^2} \approx \frac{\rho_S S}{\rho n} \nabla^2 T . \]

This equation describes the propagation of second sound in liquid helium; whereas equation (7),

\[ \frac{\partial^2 P}{\partial t^2} = \nabla^2 P \]

describes the propagation of first sound.

The first step in the solution of equations (7) and (10) is to express them in terms of the variables \( S \) and \( \rho \) only. Noting that we may write \( T = T(\rho, S) \) and \( P = P(\rho, T) \), we
the respective solutions are
\[-\omega^2 S \approx - \frac{\rho_0 S^2}{\rho_n} \left( \frac{\partial T}{\partial S} \right)_\rho k^2 S - \frac{\rho S^2}{\rho_n} \left( \frac{\partial T}{\partial \rho} \right)_S k^2 \rho\]
and
\[-\omega^2 \rho \approx - \left( \frac{\partial P}{\partial \rho} \right)_S k^2 S - \left( \frac{\partial P}{\partial \rho} \right)_S k^2 \rho\]

Writing \( \omega/k = c \), \( a\alpha = \frac{\rho_0 S^2}{\rho_n} \left( \frac{\partial T}{\partial S} \right)_\rho \), \( a\beta = \frac{\rho S^2}{\rho_n} \left( \frac{\partial T}{\partial \rho} \right)_S \), \( \gamma = \left( \frac{\partial P}{\partial S} \right)_\rho \) and \( \delta = \left( \frac{\partial P}{\partial \rho} \right)_S \), we have
\[(a\alpha-c^2)S + a\beta \rho = 0\]
and
\[\delta S + (\gamma-c^2) \rho = 0.\]

The condition on the coefficients for these two equations to be compatible is that the determinant of the coefficients vanish. In this case we get
\[(a\alpha-c^2)(\gamma-c^2) = a\beta S.\]

Multiplying out,
\[c^4 - (\gamma+a\alpha)c^2 + a(a\gamma-a\beta \delta) = 0.\]

Solving, we obtain two basically different solutions,
\[c^2 = \frac{\gamma + a\alpha}{2} \pm \frac{1}{2} \left[ (\gamma + a\alpha)^2 - 4a(a\gamma - a\beta \delta) \right]^{1/2}\]

\[= \frac{\gamma + a\alpha}{2} \pm \frac{1}{2} \left[ (\gamma - a\alpha)^2 - 4a\beta \delta \right]^{1/2}.\]

Assuming that \( 4a\beta \delta \ll (\gamma-a\alpha)^2 \), we obtain
\[c^2 \approx \frac{\gamma + a\alpha}{2} \pm \frac{\gamma - a\alpha}{2}\]
so that,
\[c^2 \approx \begin{cases} \gamma \\ a\alpha \end{cases}
or
\[(12) \quad c_1 \approx \left( \frac{\rho_0 S^2}{\rho_n} \right)^{1/2} \rho,\]
and
\[(13) \quad c_2 \approx \left( \frac{\rho S^2}{\rho_n} \left( \frac{\partial T}{\partial S} \right)_\rho \right)^{1/2}.\]
Now \( \left( \frac{\partial p}{\partial \rho} \right) \) is the square of the speed of ordinary sound so \( c \) is identified. Equation (13) gives the speed of second sound which corresponds to a thermal wave which does not affect the density. Since \( \beta \to 0 \) as \( T \to T_2 \), this wave mode can only propagate in helium II. It may be interpreted as an ordinary sound wave which propagates in the medium of the excitation gas.

**B3.6a Classical Rotational Flow**

Suppose we have a long cylinder in slow steady rotation containing an incompressible viscous fluid. We wish to determine the velocity distribution in the cylinder. Since the cylinder is long we can approximate the three dimensional problem by a two dimensional one and use polar coordinates. The first step, however, is to write down the Navier-Stokes equation.

\[
\frac{\partial u}{\partial t} + (u \cdot \nabla) u = -\frac{1}{\rho} \nabla p + \frac{\eta}{\rho} \nabla^2 u.
\]

Because of the symmetries and boundary conditions

\[
\begin{align*}

v_r &= 0, \quad \frac{\partial p}{\partial \theta} = 0, \quad \frac{\partial v_r}{\partial t} = 0, \\

\frac{\partial v_r}{\partial r} &= 0, \quad \frac{\partial v_r}{\partial \theta} = 0, \quad \frac{\partial v_\theta}{\partial \theta} = 0, \quad \frac{\partial v_\theta}{\partial t} = 0.
\end{align*}
\]

In polar coordinates the two components of the Navier-Stokes equation are

\begin{align*}
(1) \quad & \frac{\partial v_r}{\partial t} + v_r \frac{\partial v_r}{\partial r} + \frac{v_\theta}{r} \frac{\partial v_r}{\partial \theta} - \frac{v_\theta^2}{r} \\
& = -\frac{1}{\rho} \frac{\partial p}{\partial r} + \frac{\eta}{\rho} \left[ \frac{\partial^2 v_r}{\partial r^2} + \frac{1}{r} \frac{\partial v_r}{\partial r} - \frac{2}{r^2} \frac{\partial v_\theta}{\partial \theta} - \frac{v_r}{r} \right]
\end{align*}

\begin{align*}
(2) \quad & \frac{\partial v_\theta}{\partial t} + v_r \frac{\partial v_\theta}{\partial r} + \frac{v_\theta}{r} \frac{\partial v_\theta}{\partial \theta} + \frac{v_r v_\theta}{r} \\
& = -\frac{1}{\rho r} \frac{\partial p}{\partial \theta} + \frac{\eta}{\rho} \left[ \frac{\partial^2 v_\theta}{\partial r^2} + \frac{1}{r} \frac{\partial v_\theta}{\partial r} + \frac{2}{r^2} \frac{\partial v_\theta}{\partial \theta} - \frac{v_\theta}{r} \right]
\end{align*}

Applying the given conditions, equation (1) becomes

\[
\frac{v_\theta^2}{r} = -\frac{1}{\rho} \frac{\partial p}{\partial r},
\]

\[
- \frac{v_\theta^2}{r} = -\frac{1}{\rho} \frac{\partial p}{\partial r}.
\]
and equation (2) becomes

\[
\begin{align*}
0 &= \frac{\rho}{r^2} \left[ \frac{\partial^2 v_\theta}{\partial r^2} + \frac{1}{r} \frac{\partial v_\theta}{\partial r} - \frac{v_\theta}{r^2} \right].
\end{align*}
\]

It is obvious that equation (3) is just the usual centrifugal force expression; namely,

\[ F = -\nabla P = \frac{\rho v_\theta^2}{r} = \rho \omega^2 r. \]

The second equation (4) can be written

\[ \frac{\partial^2 v_\theta}{\partial r^2} + \frac{1}{r} \frac{\partial v_\theta}{\partial r} - \frac{v_\theta}{r^2} = 0. \]

and has two solutions,

\[ v_\theta = k_1 r \]

and

\[ v_\theta = k_2 r^{-1}. \]

If we let \( k_1 = \omega \), the first solution clearly corresponds to rigid body motion. The second solution corresponds to vortex motion.

**B3.6b Minimal Energy of Classical Irrotational Flow**

If we have irrotational flow, a necessary and sufficient condition is that the velocity field be derivable from a potential. Thus we may write the expression for the kinetic energy of a two dimensional flow system as

\[ T = \frac{1}{2} \rho \int v_i^2 \, d^2 x = \frac{1}{2} \rho \int (\frac{\partial \phi}{\partial x_i})^2 d^2 x \]

where \( \phi \) is the potential field.

Let us now assume that the kinetic energy involves velocities not derivable from a potential field. That is, assume

\[ T' = \frac{1}{2} \rho \int u_i^2 \, d^2 x \]

where \( u_i = \frac{\partial \phi'}{\partial x_i} \). Taking the difference between these two expressions we get

\[ T' - T = \frac{1}{2} \rho \int \left[ u_i^2 - (\frac{\partial \phi}{\partial x_i})^2 \right] d^2 x \]
In an incompressible fluid, so the last line may be rewritten to give

\[ T' - T = \frac{1}{2} \rho \int \left[ u_i - \left( \frac{\partial v_i}{\partial x_i} \right) \right]^2 d^3x + \rho \int \left[ \left( \frac{\partial u_i}{\partial x_i} \right)^2 - u_i \frac{\partial \varphi}{\partial x_i} \right] d^3x. \]

which by Gauss's theorem becomes,

\[ T' - T = \frac{1}{2} \rho \int \left[ u_i - \left( \frac{\partial v_i}{\partial x_i} \right) \right]^2 d^3x + \rho \int \frac{\partial \varphi}{\partial x_i} \left[ \frac{\partial \varphi}{\partial x_i} - u_i \right] d^3x. \]

We assume that the two velocity distributions are compatible on the outside boundary so that the second integral vanishes. This leaves for the difference between the two kinetic energies

\[ T' - T = \frac{1}{2} \rho \int \left[ u_i - \left( \frac{\partial v_i}{\partial x_i} \right) \right]^2 d^3x. \]

The integral is positive definite and so is \( \rho \). Therefore, \( T' \geq T \). It follows then that the minimum occurs when

\[ u_i = \frac{\partial \varphi}{\partial x_i}; \]

that is, when the velocity field is irrotational.

Note that the above proof depends upon equating the two velocity fields at the same boundary surface. When they are equal there, irrotational motion is the minimum energy motion.

**B3.6c Helmholtz Vortex Theorem.** Essentially Helmholtz's theorem states that vorticity is conserved. This is equivalent to the statement that the vortex lines "adhere" to the same fluid particles and move with them.

Suppose that \( \vec{A} \) is a vector associated with a moving fluid and that \( \delta \sigma \) is a small segment of surface that also
moves with the fluid. In a time $\delta t$, the area $\delta \sigma$ sweeps out the small volume $\delta \tau$ shown in figure 1. By Gauss's theorem, \[
\text{Total Lines} = \int_{\text{Vol}} A \cdot d\sigma = \int_{\text{Surface}} \nabla \cdot A \, d\tau,
\]
which for a small enough volume element can be written
\[
\text{Total Lines} = \int_{\text{Surface}} A \cdot d\sigma = \int_{\text{Surface}} \nabla \cdot A \, d\tau.
\]
If at first we consider $A$ a constant vector, then the surface integral can be written
\[
\int_{\text{Surface}} A \cdot d\sigma = \int_{\text{Surface}} A \cdot d\sigma_1 + \int_{\text{Surface}} A \cdot d\sigma_2 + \int_{\text{Surface}} A \cdot d\sigma_3
\]
where
\[
da = d\sigma \times \nabla \delta t.
\]
Now,
\[
A \cdot d\sigma = A \cdot d\sigma \times \nabla \delta t = d\sigma \times \nabla \delta t \times A,
\]
and by Stokes theorem
\[
\int_{\text{Surface}} A \cdot d\sigma = \int_{\text{Surface}} \nabla \times (A \cdot d\sigma) \cdot d\sigma = \int_{\text{Surface}} (\nabla \times A) \cdot d\sigma,
\]
For a small enough end area, we obtain
\[
\int_{\text{Surface}} A \cdot d\sigma \approx \delta t \left[ \nabla \times (A \cdot d\sigma) \right] \cdot \delta \sigma,
\]
and we may now write for the change of lines per unit time across $\delta \sigma$,
\[
\delta (A \cdot d\sigma) = (A \cdot d\sigma)_2 - (A \cdot d\sigma)_1 = \int_{\text{Total Surface}} A \cdot d\sigma - \int_{\text{Surface}} A \cdot d\sigma.
\]
where $(\nabla \cdot \mathbf{A})\mathbf{n}-(\nabla \times (\nabla \times \mathbf{A}))\mathbf{n}$ is the normal component.

Thus

\[
\delta (A_n \delta \sigma) = \frac{1}{2} (\nabla \cdot \mathbf{A}) \mathbf{n} - [(\nabla \times (\nabla \times \mathbf{A}))\mathbf{n}] \delta \sigma.
\]

Now the change in the number of lines due to a change in the vector $\mathbf{A}$ per unit time is

\[
\delta \sigma \frac{\partial A_n}{\partial t},
\]

so that the total line change per unit time is given by

\[
\frac{d}{dt} (A_n \delta \sigma) = \int \frac{\partial A_n}{\partial t} + v_n (\nabla \cdot \mathbf{A}) - (\nabla \times (\nabla \times \mathbf{A}))\mathbf{n} \delta \sigma.
\]

If we now identify $\mathbf{A}$ with $\Omega$, the vorticity, we have

\[
\frac{d}{dt} (\Omega \cdot \delta \sigma) = \frac{d}{dt} (A_n \delta \sigma) = \int \frac{\partial \Omega_n}{\partial t} - [(\nabla \times (\nabla \times \mathbf{A}))\mathbf{n}] \delta \sigma = 0,
\]

since $\nabla \cdot \mathbf{A} = \nabla \cdot (\nabla \times \mathbf{v}) = 0$ and $\frac{\partial \Omega_n}{\partial t} - \nabla \times (\nabla \times \mathbf{v}) = 0$.

Therefore,

\[
\frac{d}{dt} [\Omega_n \delta \sigma] = 0
\]

and

\[
(1) \quad \Omega_n \delta \sigma = \Omega \cdot \delta \sigma = \text{constant}.
\]

This result expresses the fact that for an inviscid, incompressible fluid, vorticity is conserved.

We may relate this to the circulation by Stokes theorem,

\[
\text{Circulation} = \oint_{\text{Perimeter}} \mathbf{v} \cdot ds = \int_{\text{Surface}} (\nabla \times \mathbf{v}) \cdot d\mathbf{a} = \int_{\text{Surface}} \Omega \cdot d\mathbf{a}.
\]

It follows from this and equation (1) that circulation is conserved also.

A vortex tube is a tube of fluid that includes one or
more vortex lines and whose sides are parallel to those lines. If \( \delta \sigma \) is the cross sectional area of a tube which includes a single vortex line, we may define vortex strength by

\[
\mu = \Omega_s \cdot \delta \sigma
\]

where \( \Omega_s \) is the vorticity due to a single line.

B3.6d **Classical Velocity Field.** Any continuous vector field \( V \) defined everywhere in space and vanishing at infinity together with its first derivatives, can be represented as the sum of an irrotational field \( V_i \) and a solenoidal field \( V_s \). Thus,

\[
V = V_i + V_s
\]

where

\[
\nabla \times V_i = 0, \quad \nabla \cdot V_s = 0.
\]

It follows then that

\[
\nabla \times V = \nabla \times V_s
\]

and

\[
\nabla \cdot V = \nabla \cdot V_i,
\]

so that a vector field \( V \) as described above may be defined uniquely to within a vector constant by giving its divergence and its curl. We shall discuss this theorem but not prove it.

An irrotational vector can be derived from the gradient of a scalar potential \( \Phi \). Thus

\[
V_i = -\nabla \Phi + \text{const}.
\]

Taking the divergence we get

\[
\nabla \cdot V_i = \nabla \cdot V = -\nabla^2 \Phi.
\]

The solution is well known and is given by

\[
4\pi \Phi = \int \frac{\nabla \cdot V}{r} \, d\tau.
\]
A solenoidal vector can be derived from the curl of a vector potential. Thus

\[ \mathbf{A} = \nabla \times \mathbf{A} + \text{Const vector} , \]

with the arbitrary gauge condition \( \nabla \cdot \mathbf{A} = 0 \). Taking the curl of this expression and applying the gauge condition yields

\[ \nabla \times \mathbf{A} = \nabla \times \mathbf{A} = -\nabla^2 \mathbf{A} . \]

The solution again is well known and is given by

(2) \[ 4\pi \mathbf{A} = \int \frac{\mathbf{A} \cdot \mathbf{d} \ell}{r'} d\tau' . \]

We now wish to apply the above results to the velocity field of a fluid. Noting that \( \mathbf{\Omega} = \nabla \times \mathbf{A} \), we have from equation (2)

\[ 4\pi \mathbf{A} = \int \frac{\mathbf{\Omega}}{r} d\tau' . \]

For a volume element that includes a single line we may write for the incremental contribution to \( \mathbf{A} \),

\[ 4\pi \delta \mathbf{A} = \frac{\mathbf{\Omega} \delta \tau}{r} \]
\[ = \frac{\mathbf{\Omega} \delta \sigma \delta \ell}{r} , \]

where \( \delta \sigma \delta \ell = \delta \tau \) and \( \mathbf{\ell} = \mathbf{\hat{A}} \) by hypothesis. Since

\[ \mu = \mathbf{\omega} \delta \sigma , \]

we have

\[ 4\pi \delta \mathbf{A} = \frac{\mu}{r} \delta \ell . \]

Taking the curl we obtain the contribution of the elemental vortex to the velocity field. Thus

\[ 4\pi \delta \mathbf{v} = 4\pi \nabla \times \delta \mathbf{A} = \nabla \times \frac{\mu}{r} \delta \ell . \]

Choosing \( \mathbf{\ell} = \mathbf{\hat{A}} \) for convenience we get

\[ 4\pi \delta \mathbf{v} = \begin{vmatrix} \hat{z} & \hat{j} & \hat{k} \\ \partial x & \partial y & \partial z \\ 0 & 0 & \frac{\mu}{r} \delta \ell \end{vmatrix} , \]
so that

\[ 4\pi \delta v_x = -\frac{\mu}{r^2} \delta l \frac{y}{r} \]
\[ 4\pi \delta v_y = \frac{\mu}{r^2} \delta l \frac{x}{r} \]
\[ 4\pi \delta v_z = 0. \]

These last results can be compactly expressed by

\[ 4\pi \delta v = \frac{\mu}{r^2} \delta l \times \hat{r} \]

or

\[ \delta v = \frac{\mu}{4\pi r^3} \delta \delta \times \hat{r} . \] (3)

This final expression is the equivalent of the law of Biot and Savart in electromagnetism and gives the velocity field at some point in space as a function of the vorticity in a given tube and the geometry of the situation. Clearly the correspondences between the electromagnetic and fluid cases are

\[ \mathbf{B} \leftrightarrow \mathbf{v} \]
\[ I/c \leftrightarrow \frac{\mu}{4\pi} \].

The integral form is

\[ \mathbf{v} = \frac{\mu}{4\pi} \int \frac{d\delta \times \hat{r}}{r^2} . \] (4)

**B3.6e Velocity Field of a Classical Ring Vortex.** From equation (B3.6d-3), the equivalent of the Biot-Savart law for fluids, we have

\[ \delta v = \frac{\mu}{4\pi} \frac{\delta \delta \times \hat{r}}{r^2} . \] (1)

Suppose we have a vortex ring and wish to determine the velocity field at the center of the ring. Each element of length \( \delta l \) contributes

\[ \delta v = \frac{\mu}{4\pi a^2} \delta l \hat{z} , \]

where \( a \) is the ring radius and \( \hat{z} \) is the direction normal to the plane of the ring. The total velocity field at the
center is then

\[ V = \frac{\mu}{2\pi \hat{z}}. \]

The field at each point on the ring due to the contributions of all other elements of the ring is found as follows. Consider figure 1 and use equation (1) to write

\[ \delta V(p) = \frac{\mu}{4\pi} \frac{\delta S \times \phi}{r^3} = \frac{\mu}{4\pi} \frac{\delta l \sin \phi \hat{z}}{r^2}. \]

But \( \delta l = a \delta \theta, \phi = \Theta / \pi \) and \( r^2 = 4a^2 \sin^2 \Theta / 2 \).

Substituting we get

\[ \delta V(p) = \frac{\mu}{4\pi} \frac{a \delta \theta \sin \Theta / 2}{2a^2 \sin^2 \Theta / 2} \hat{z} \]

\[ = \frac{\mu}{8\pi a} \csc \Theta / 2 \delta \Theta / 2 \hat{z}. \]

Figure 1

Therefore,

\[ V(p) = \frac{\mu}{8\pi a} 2 \int_{\Theta_{\min}}^{\Theta_{\max}} \csc \frac{\Theta}{2} d \frac{\Theta}{2} \hat{z}, \]

where \( \Theta_{\max} \approx \pi \). This last result needs some discussion. Physically there is some limiting inner dimension of a vortex line since the velocity cannot become infinite. We therefore require the observation point to be displaced from the theoretical vortex line by an amount \( \epsilon \) as shown in figure 2. This requires that the minimum \( \Theta \) be given by

\[ \Theta_{\min} \approx \epsilon / a. \]

Therefore we may substitute into the last equation to obtain

\[ V(p) = \frac{\mu}{4\pi a} \ln \frac{\epsilon a}{\epsilon}, \]

Figure 2

Since the size of the inner
radius is unknown, this result can be written

\[ V(P) = \frac{\mu}{4\pi a} \ln \frac{a}{\epsilon} \]

where \( \epsilon = 4\epsilon' \). *

**B3.6h Energy Dissipation in a Classical Fluid.** The total kinetic energy of a fluid is given by

\[ T = \frac{1}{2} \int \rho v^2 d^3x \]

What we wish to find out is first, how this kinetic energy changes in time and second, how much of the change is due to irreversible losses related to the viscosity. In order to do this we consider the kinetic energy in a volume that is fixed in space, partially differentiate equation (1) with respect to time, and make use of the Navier-Stokes equation to bring in the viscous loss terms. Thus,

\[ \frac{\partial T}{\partial t} = \frac{1}{2} \int \frac{\partial}{\partial t} \left( \rho v^2 \right) d^3x . \]

Using the summation convention, the time derivative of the kinetic energy density can be written

\[ \frac{1}{2} \frac{\partial \rho v^2}{\partial t} = \rho v^2 \frac{\partial v_i}{\partial t} . \]

From the Navier-Stokes equation for incompressible viscous fluids, we have

\[ \frac{\partial v_i}{\partial t} = -v_k \frac{\partial v_i}{\partial x_k} - \frac{1}{\rho} \frac{\partial p}{\partial x_i} + \frac{1}{\rho} \frac{\partial}{\partial x_k} \sigma_{ik} . \]

Making the appropriate substitutions leads to

\[ \frac{\partial T}{\partial t} = \int \left[ -\rho v_i v_k \frac{\partial v_i}{\partial x_k} - v_i \frac{\partial p}{\partial x_i} + v_i \frac{\partial \sigma_{ik}'}{\partial x_k} \right] d^3x \]

\[ = \int \left[ -\rho v_k \frac{\partial v_i}{\partial x_k} - v_k \frac{\partial p}{\partial x_k} + \frac{\partial v_i \sigma_{ik}'}{\partial x_k} - \sigma_{ik} \frac{\partial v_i}{\partial x_k} \right] d^3x , \]

* Really the somewhat artificial sounding restriction on \( \Theta \) amounts to a restriction on \( r \), since \( \Theta \) is merely a convenient variable with which to express \( r \) in the integration.
or
\[
\frac{dT}{dt} = \int \left[ -\rho \frac{\partial v_i}{\partial \kappa} \left( \frac{1}{2} v_i^2 + \frac{P}{\rho} \right) + \frac{\partial v_i}{\partial \kappa} \frac{\partial v_i}{\partial \kappa} - \sigma_{ik} \frac{\partial v_i}{\partial \kappa} \right] d^3x .
\]

Since \( \frac{\partial v_i}{\partial \kappa} = 0 \), this can be written
\[
\frac{dT}{dt} = \int \left[ -\frac{\partial}{\partial \kappa} \rho v_i \left( \frac{1}{2} v_i^2 + \frac{P}{\rho} \right) + \frac{\partial v_i}{\partial \kappa} \sigma_{ik} - \frac{\partial v_i}{\partial \kappa} \frac{\partial v_i}{\partial \kappa} \right] d^3x 
\]
\[
= \int \left[ -\frac{\partial}{\partial \kappa} \left[ \rho v_i \left( \frac{1}{2} v_i^2 + \frac{P}{\rho} \right) - v_i \sigma_{ik} \right] - \sigma_{ik} \frac{\partial v_i}{\partial \kappa} \right] d^3x 
\]
\[
= -\int \frac{\partial}{\partial \kappa} \left[ \rho v_i \left( \frac{1}{2} v_i^2 + \frac{P}{\rho} \right) - v_i \sigma_{ik} \right] d^3x - \int \sigma_{ik} \frac{\partial v_i}{\partial \kappa} \right] d^3x .
\]

By the use of Gauss's theorem the first integral is converted to a surface integral. Thus
\[
-\int \frac{\partial}{\partial \kappa} \left[ \rho v_i \left( \frac{1}{2} v_i^2 + \frac{P}{\rho} \right) - v_i \sigma_{ik} \right] d^3x 
\]
\[
= \int \left[ \rho v_i \left( \frac{1}{2} v_i^2 + \frac{P}{\rho} \right) - v_i \sigma_{ik} \right] dA_k ,
\]

where \( A_k \) is a directed differential surface element. It may now be interpreted as the net energy flux, kinetic, potential, and dissipative which flows through the bounding surface with the bulk fluid motion. Since there are no sources within the volume and since the fluid is incompressible, this term vanishes by conservation of energy. We are then left with the second integral which clearly represents the energy which has been dissipated during the time the particular element of fluid has occupied the control volume. We have for the rate of energy dissipation
\[
(2) \quad \frac{dT}{dt} = -\int \sigma_{ik} \frac{\partial v_i}{\partial \kappa} d^3x .
\]

Recalling that
\[
\sigma_{ik} = \gamma \left( \frac{\partial v_i}{\partial \kappa} + \frac{\partial v_i}{\partial \kappa} \right),
\]
we have for the integrand

$$\eta \frac{\partial v_i}{\partial x_\kappa} \left( \frac{\partial v_i}{\partial x_\kappa} + \frac{\partial v_k}{\partial x_i} \right) = \eta \left( \frac{\partial v_i}{\partial x_\kappa} \right)^2 + \eta \frac{\partial v_i}{\partial x_\kappa} \frac{\partial v_i}{\partial x_k} .$$

The indices $i$ and $k$ are dummy indices and can be relabeled. Thus we assert that

$$\left( \frac{\partial v_i}{\partial x_\kappa} \right)^2 + \frac{\partial v_i}{\partial x_\kappa} \frac{\partial v_i}{\partial x_i} = \left( \frac{\partial v_i}{\partial x_\kappa} \right)^2 + \frac{\partial v_i}{\partial x_i} .$$

It follows then that

$$\frac{\eta}{2} \left[ \left( \frac{\partial v_i}{\partial x_\kappa} \right)^2 + \frac{\partial v_i}{\partial x_\kappa} \frac{\partial v_i}{\partial x_i} + \frac{\partial v_i}{\partial x_i} + \frac{\partial v_i}{\partial x_i} \right] = \frac{\eta}{2} \left[ \left( \frac{\partial v_i}{\partial x_\kappa} \right)^2 + \frac{\partial v_i}{\partial x_i} \right] ,$$

or

$$\eta \frac{\partial v_i}{\partial x_\kappa} \left[ \frac{\partial v_i}{\partial x_\kappa} + \frac{\partial v_i}{\partial x_i} \right] = \eta \left( \frac{\partial v_i}{\partial x_\kappa} \right)^2 .$$

Equation (2) now becomes

$$\frac{\partial T}{\partial t} = -\eta \int \left( \frac{\partial v_i}{\partial x_\kappa} \right)^2 d^3x ,$$

and is the desired expression for the dissipation rate of kinetic energy.

An alternate form which is more intuitively obvious can be obtained by converting this last result to a surface integral. Thus

$$\frac{\partial T}{\partial t} = -\eta \int \left( \frac{\partial v_i}{\partial x_\kappa} \right)^2 d^3x = -\eta \int \frac{\partial}{\partial x_\kappa} \left[ v_i \frac{\partial v_i}{\partial x_\kappa} \right] d^3x + \eta \int v_i \frac{\partial}{\partial x_\kappa} \left( \frac{\partial v_i}{\partial x_\kappa} \right) d^3x .$$

One of the assumptions in the Navier-Stokes theory is that the viscous energy dissipation depends only on the first spatial derivatives of the velocity components and that higher derivatives vanish. Therefore the second integral on the right vanishes and we have

$$\frac{\partial T}{\partial t} = -\eta \int \frac{1}{2} \left[ v_i \frac{\partial v_i}{\partial x_\kappa} \right] d^3x .$$

Transforming this to a surface integral by Gauss's theorem we obtain

$$\frac{\partial T}{\partial t} = -\eta \int v_i \frac{\partial v_i}{\partial x_\kappa} dA_\kappa ,$$
which becomes in vector notation

\[ \frac{\partial T}{\partial t} = - \frac{\eta}{\rho} \int \nabla \left( \frac{\rho v^2}{2} \right) \cdot dA \]

Thus, we have the intuitively reasonable result that the rate of viscous energy dissipation is proportional to the gradient of the kinetic energy.

**B3.6g Classical Poiseuille Flow** Consider a circular pipe, as shown in figure 1, containing a flowing, viscous but incompressible fluid. We wish to determine the velocity distribution in the pipe under conditions of steady state laminar flow. As usual, begin with the Navier-Stokes equation

\[ \frac{\partial v}{\partial t} + (v \cdot \nabla)v = -\frac{\nabla P}{\rho} + \frac{\eta}{\rho} \nabla^2 v . \]

Because of the assumption of steady state laminar flow

\[ v_x = v_y = 0 \]

and

\[ \frac{\partial v_x}{\partial x} = \frac{\partial v}{\partial t} = 0 \]

Writing the equation of motion in component form, we have

\[ \frac{\partial v_i}{\partial t} + v_k \frac{\partial v_i}{\partial x_k} = -\frac{\partial P}{\partial x_i} + \eta \left[ \frac{\partial^2 v_i}{\partial y^2} + \frac{\partial^2 v_i}{\partial z^2} \right] . \]

**Figure 1.**

Because of the assumed conditions this reduces to

\[ -\frac{\partial}{\partial x} \left( \frac{P}{\rho} \right) + \frac{\eta}{\rho} \left[ \frac{\partial^2 v_x}{\partial y^2} + \frac{\partial^2 v_x}{\partial z^2} \right] = 0 , \]

from which it follows that

\[ \frac{\partial P}{\partial x} = \text{Constant} , \]

and

\[ \eta \left[ \frac{\partial^2 v_x}{\partial y^2} + \frac{\partial^2 v_x}{\partial z^2} \right] = \text{Constant} . \]

Thus we may write

\[ \frac{\partial P}{\partial x} = -\frac{\Delta P}{l} , \]
where $\Delta p$ is the pressure drop over a length $l$ of pipe and the equation to be solved is

$$
\left( \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} \right) = -\frac{\Delta p}{\eta l}.
$$

For this problem it is convenient to transform to cylindrical coordinates, which because of the symmetries involved yield the simple expression

$$
\frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial v}{\partial r} \right) = -\frac{\Delta p}{\eta l}
$$
or

$$
\frac{\partial^2 v}{\partial r^2} + \frac{1}{r} \frac{\partial v}{\partial r} + \frac{\Delta p}{\eta l} = 0.
$$

The solution is

$$
V = -\frac{\Delta p}{4\eta l} r^2 + a \log r + b.
$$

Since $V$ is finite at $r=0$, $a=0$. Also since $V=0$ at $r=\infty$, $b=-\frac{\Delta p}{4\eta l} (R^2-r^2)$. Therefore,

(1) \quad V = \frac{\Delta p}{4\eta l} (R^2-r^2)

is the desired solution.

Using this expression we may now solve for the kinetic energy loss per second due to the viscosity. The loss rate equation (B3.6f-4) is

$$
\frac{\partial T}{\partial t} = -\frac{1}{\rho} \int \nabla \cdot \left( \frac{\rho v^2}{2} \right) \cdot d\mathbf{A},
$$

and can be written

$$
\frac{\partial T}{\partial t} = -2\pi \eta \int_0^R \frac{\partial v}{\partial r} rdr.
$$

But from equation (1),

$$
\frac{\partial v^2}{\partial r} = \frac{3}{\eta l} \left[ \left( \frac{\Delta p}{4\eta l} \right)^2 (R^2-r^2)^2 \right]^{-1/2} = \left( \frac{\Delta p}{4\eta l} \right)^2 (R^2-r^2)(-2r),
$$
so,
\[
\frac{\partial T}{\partial t} = 8\pi \eta \left( \frac{\Delta P}{\alpha \eta} \right)^2 \int_{0}^{R} \left[ R^2 - r^2 \right] r^2 \, dr .
\]

Integrating we obtain for the rate of kinetic energy loss
\[
\frac{\partial T}{\partial t} = \pi \frac{(\Delta P)^2}{2 \eta} \left[ \frac{R^6}{3} - \frac{R^5}{5} \right]
\]
or
\[
(2) \quad \frac{\partial T}{\partial t} = \frac{\pi (\Delta P)^2}{15 \eta} R^5
\]

where \( \Delta P \) is the pressure drop per unit length of pipe.

B3.6h Irrotational Flow in Helium. Usually in the solving of the Schrodinger equation for a simple system, the center of mass motion is discarded as being of little interest and attention is directed toward the motions with respect to that center of mass. However, for our purposes, the center of mass motion is of interest and leads to a semi-quantitative specification of the wave function describing the irrotational flow of helium II.

Following Feynman, if the liquid moves as a body then one expects the wavefunction to have the form
\[
\psi = e^{i k \cdot \vec{r}} \frac{\delta}{\partial \vec{r}} \Phi = e^{i \frac{m}{\hbar} \vec{V} \cdot \vec{r}} \frac{\delta}{\partial \vec{r}} \Phi,
\]
where the summation is over the individual atoms and where \( \Phi \) is the ground state wavefunction discussed earlier. If, however, the motion of the system changes slowly with position then the wavefunction must be generalized. For example, one can write
\[
\psi = e^{i \frac{m}{\hbar} \vec{V} \cdot \vec{r} \delta_j \Phi},
\]
where \( \vec{V} \delta_j \) is a velocity function of position. Even more generally
\[
(1) \quad \psi = e^{i \vec{S}_j \delta_j \Phi},
\]
where $S(x)$ is a slowly changing phase function of position. Provided the phase changes are small over distances compared to the atomic spacing, it is reasonable to suppose that the velocity is given by

$$\mathbf{v} = \frac{m}{\hbar} \nabla S.$$

That is, the velocity is proportional to the gradient of the phase change. One can see this by considering the current flow. Thus,

$$\mathbf{j}(i_j) = \frac{\hbar}{2m_i} \left[ i_x S_j \nabla \cdot \mathbf{v} - i_y S_j \nabla \times \mathbf{v} \right]$$

which gives, finally,

$$\mathbf{j}(i_j) = \frac{\hbar}{2m_i} \left[ 2 i_x S_j \nabla \cdot \mathbf{v} + i_y S_j \nabla \times \mathbf{v} \right] = \frac{\hbar}{m} \mathbf{\Phi} \cdot \mathbf{\Phi} \nabla S_j(i_j)$$

But,

$$\mathbf{j}(i_j) = \rho \mathbf{v} i = \mathbf{\Phi} \cdot \mathbf{\Phi} \mathbf{v} :$$

which gives, finally,

$$\mathbf{v} = \frac{\hbar}{m} \nabla S.$$

From the fact the velocity can be expressed as the gradient of a scalar function, it follows that the velocity field is irrotational. If the medium is simply connected, $\nabla \times \mathbf{v} = 0$ everywhere. If it is multiply connected, then $\nabla \times \mathbf{v} = 0$ everywhere except at the singularities; that is, at those points which correspond to centers of vortices. Since the wave function must be single valued, these statements can be summarized by saying that the circulation around any closed path must be an integral multiple of $2\pi$. That is,

$$\oint_{\partial \mathcal{A}} \mathbf{v} \cdot d\mathbf{l} = \oint_{\partial \mathcal{A}} (\nabla \times \mathbf{v}) \cdot d\mathbf{n} = \frac{\hbar}{m} 2\pi n \quad n = 0, 1, \ldots.$$ 

Experimentally it is known that vortices do exist in flowing liquid helium II so the original assumption that
the motion of the system changes slowly everywhere cannot hold. It must change discontinuously at certain points; however, the number of places it can do so is finite and relatively small. This means a liquid helium II system can be viewed as a finite set of macroscopic velocity regimes where the motion is irrotational and where the velocity can be expressed as the gradient of the phase function of position.

B3.6i Rotational Flow in Helium. In the previous section, the assumption was made that the velocity of small elements of volume in helium II could be expressed as the gradient of a scalar function. This led to completely irrotational flow in a simply connected system and irrotational flow regimes in a multiply connected system; multiple-connectedness being the result of vortex formation. Despite the fact that experimental evidence suggests this to be a correct model, we seek further theoretical justification for the assumption, \( \mathbf{u} = \nabla \phi \). Why should the flow be irrotational at all? For example, why cannot liquid helium rotate as a rigid body?

To analyze rotational motion, it is convenient to think of a cylindrical vessel containing the helium at 0°K. Suppose the helium is under sufficient pressure to be solid and the cylinder is put into rotation with a specified angular velocity. The helium is thus given a definite angular momentum. Next the pressure is reduced so that the helium becomes liquid and we ask, what kind of motion of the liquid will ensue? Initially assume that it will be rigid body motion so that the question reduces to whether a single rigid body or many smaller rigid bodies will be most favorable, energetically, for a given angular momentum.

The angular momentum for a rigid body cylinder is given by

\[ L = I \omega = \frac{1}{2} I R^2 \omega. \]
The energy for such a body is

\[ E = \frac{1}{2} I \omega^2 = \frac{1}{2} L \omega \]  

It is clear that if we replace the large single body by four smaller bodies having radii \( \frac{R}{2} \), the velocity of each smaller body must be four times that of the original body to maintain the same total angular momentum. This implies directly that four times the energy is required to maintain the angular momentum. Hence a single rigid body is energetically much more favorable than many smaller rigid bodies.

Next let us analyze vortex motion. Quantum mechanically, this motion is a consequence of requiring that the circulation be given by

\[ \oint \mathbf{v} \cdot d\mathbf{l} = 2\pi \frac{\hbar}{m} n ; \quad n = 0, 1, \ldots \]

One can see that the above condition leads to vortex motion by carrying out the integration over a circular path and solving for the velocity. One gets

\[ v = \frac{n \hbar/m}{r} = \frac{\hbar}{r} , \]

which is just the velocity distribution associated with vortex motion.

First, we are interested in the maximum density of vortices one can expect to find in a rotating cylinder of liquid helium. The desired result follows immediately from Stokes' theorem. If we define a unit of circulation, \( h/m \), to be a "line", then

\[ \oint \mathbf{v} \cdot d\mathbf{l} = \oint (\nabla \times \mathbf{v}) \cdot d\mathbf{A} = n \frac{\hbar}{m} = \left\{ \text{Number of lines Enclosed by Path} \right\} . \]

Since \( \nabla \times \mathbf{v} = 2 \mathbf{\Omega} \) is the average circulation per unit area, the number of lines per square centimeter is given by

\[ n_0 = 2 \left( \frac{m}{h} \right) \mathbf{\Omega} \approx 2 \times 10^3 \mathbf{\Omega} \text{ lines/cm}^2 , \]

and the maximum radius of an individual vortex can be no
smaller than

\[ b \approx \frac{1}{2} \left(2 \times 10^3 \omega^2 \right)^{-\frac{1}{2}} a m. \]

The minimum radius \( a \) of an individual vortex is difficult to assess but it cannot be less than the atomic spacing and most probably is of the same order. Feynman assumes \( 4\hbar \).

For the purposes that follow, this dimension need not be accurately known.

Let us now derive an expression for the angular momentum per unit length of line. If \( a \) is the inner vortex radius and \( b \) the outer vortex radius, then

\[ L' = \int_a^b \rho \, dm \, v \, r \]
\[ = \int_a^b 2\pi \rho r \, \frac{n \, \hbar}{m} \, r \, dr \]
\[ = \hbar \rho \, n \left( \frac{\hbar}{m} \right) b^2 \]

since \( b^2 \gg a^2 \).

The energy per unit length stored in such a line is given by

\[ E' = \frac{1}{2} \int_a^b \rho \, dm \, v^2 \]
\[ = \frac{1}{2} \int_a^b 2\pi \rho r \left( n \, \frac{\hbar}{m} \right)^2 \, dr \]
\[ = \hbar \rho \left( n \, \frac{\hbar}{m} \right)^2 \left( \ln \frac{b}{a} \right). \]

B3.6j Critical Velocity and Flow Resistance. In the discussion on superfluidity, it was seen that the critical velocity below which no excitations would be formed was approximately given by \( \frac{\hbar \rho}{m} \). If no other energy absorbing mechanisms were available, then the flow would be superfluid. However, it is known experimentally that normal flow ensues at much lower velocities than the \( 6 \times 10^3 \text{ cm/sec} \), or so, suggested by the phonon-roton spectrum. One must conclude,
therefore, that another energy absorbing mechanism does exist. Further, it is reasonable to suppose that this mechanism is related to the formation of vortices.

To get a crude idea of how the critical velocity might be affected by vortex formation, consider the passage of liquid helium II through a narrow two dimensional slit as shown in figure 1. The jet emerges into a vacuum upon leaving the slit. If we take a line integral over the path shown, then the circulation is \( V \) and the number of lines per centimeter is given by

\[
\eta_0 = \frac{vm}{2\pi \eta}
\]

Now it takes energy to form these lines which must come from the fluid itself. The rate of line formation must approximately equal \( n_0 v \) since it is assumed that the lines in the slit leave it at this rate and must be replaced by new lines at the entrance of the slit. Thus, the energy per unit length of line per second removed from the fluid is

\[
\eta_0 v E = \frac{vm}{2\pi \hbar} \nu \pi \rho \left( \frac{\pi}{m} \right)^2 \ln \frac{d}{a} = \frac{\rho v^2}{2} \left( \frac{\pi}{m} \right) \ln \frac{d}{a},
\]

where the maximum vortex radius is assumed to be on the order of \( d \). On the other hand, the kinetic energy available from the helium flow per unit volume is

\[
\frac{\rho v^2}{2}
\]

Hence the kinetic energy per second available per unit line length is

\[
\nu d \frac{\rho v^2}{2}
\]

Equating this last result to the energy removed from the fluid per second we get

\[
(1) \quad \nu c d = \frac{\pi}{m} \ln \frac{d}{a},
\]
where $V_c$ is the critical velocity.

Bo.6k Rotons as Ring Vortices. Earlier we derived expressions for the energy and angular momentum per unit length of a linear vortex. We found

$$E' = \pi \rho \left( \frac{\hbar}{m} \right)^2 \ln \frac{b}{a}$$

and

$$L' = \pi \rho \frac{\hbar}{m} b^2 .$$

If the vortex, instead of being linear, closes upon itself and forms a ring, the expressions for the total ring energy and angular momentum become

\begin{align*}
(1) \quad E &= \left[ \pi \rho \left( \frac{\hbar}{m} \right)^2 \ln \frac{R}{a} \right] (2\pi R) \\
(2) \quad L &= \left[ \pi \rho \frac{\hbar}{m} R^2 \right] (2\pi R)
\end{align*}

where $R$ is the ring radius.

We can find the linear momentum carried by the ring by noting that $\mathbf{p} = \mathbf{r} \times \mathbf{p}$ so that for circular motion, the magnitude of $\mathbf{p}$ is given by $\mathbf{p} = r\rho$. We may write therefore,

$$p = \frac{L}{R} = \left[ 2\pi \rho \frac{\hbar}{m} \right] (\pi R^2)$$

If we now take the ratio $E/p$, we obtain

$$E/p = \frac{\hbar}{mR} \ln \frac{R}{a} ,$$

which is the same expression obtained for the critical velocity through a slit.
Appendix B

References


3. Kittel, p. 278.


7. Reif, Sect. 10.1.


