Existence of Zero Sound in a Fermi Liquid*

N. D. MERMIN†
Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York
(Received 2 February 1967)

Sufficient conditions are derived for ordinary and for spin zero sound to exist in a Fermi liquid. In a pure neutral Fermi liquid, one or the other of these conditions must hold.

THE purpose of this paper is to point out that it follows from the basic notions of canonical Fermi-liquid theory[1–3] that at sufficiently low temperatures (i.e., in the collisionless regime) a Fermi liquid must sustain zero sound either in the form of density or spin-density oscillations.

Since the only candidate for a canonical Fermi liquid is liquid He_3, in which zero sound density waves have now been observed,[4,5] this conclusion should send no one rushing to the laboratory. It is reported here for three reasons:

(1) In the course of the proof simple sufficient conditions are derived for the existence of ordinary and spin zero sound, which are useful in studying the possibility of zero sound in anisotropic charged Fermi liquids[6] (i.e. metals).

(2) Should people succeed in current efforts[7,8] to find a kinetic equation to describe dilute solutions of He_3 in He_4, the method used to derive these sufficient conditions, if not the conditions themselves, should prove useful.

(3) The result has a place in the lore of normal Fermi systems, a self-sustaining body of knowledge with a quiet dignity and austere beauty, the paucity of examples notwithstanding.

The conclusion follows from these basic results:

(a) Landau's demonstration[2,3] that at sufficiently low temperatures a single-component spin-$\frac{1}{2}$ Fermi liquid with short-range interactions (here called a canonical Fermi liquid) will sustain collective oscillations with phase velocity ηv_F, provided the kinetic equation

$$(\eta-\cos\theta)\nu_s(\hat{n})=\cos\theta\int\frac{d\hat{n}'}{4\pi}\sum_{s'}F_{ss'}(\hat{n}\cdot\hat{n}')\nu_{s'}(\hat{n}') \quad (1)$$

has a solution for real η with $|\eta|>1$. Here $\nu_s(\hat{n})$ is proportional to the deformation in the mode of the Fermi surface of spin-s quasiparticles in the direction $\hat{n}$; θ is the angle between $\hat{n}$ and the direction of propagation; and $F_{ss'}(\hat{n}\cdot\hat{n}')$ is related to the forward scattering amplitude $a_{ss'}(\hat{n}\cdot\hat{n}')$ for two quasiparticles of spins s and s' and momenta $\hat{n}p_F$ and $\hat{n}'p_F$ by

$$F_{ss'}(\hat{n}\cdot\hat{n}')=A_{ss'}(\hat{n}\cdot\hat{n}')+\sum_{s''}\int\frac{d\hat{n}''}{4\pi}A_{ss''}(\hat{n}\cdot\hat{n}'')F_{s''s'}(\hat{n}''\cdot\hat{n}'), \quad (2)$$

where $A_{ss'}(\hat{n}\cdot\hat{n}')=N(0)a_{ss'}(\hat{n}\cdot\hat{n}')$, $N(0)$ being the density of states at the Fermi surface.

(b) The stability condition,[9]

$$0<\sum_s\int\frac{d\hat{n}}{4\pi}|\chi_s(\hat{n})|^2+\sum_{ss'}\int\frac{d\hat{n}}{4\pi}\int\frac{d\hat{n}'}{4\pi}\chi_s^*(\hat{n})F_{ss'}(\hat{n}\cdot\hat{n}')\chi_{s'}(\hat{n}'), \quad (3)$$

for any $\chi_s(\hat{n})$.

(c) The exclusion principle, which requires that[10]

$$A_{ss}(1)=0. \quad (4)$$

We shall prove from (2)–(4) and the assumption that A is continuous in the forward direction ($\hat{n}=\hat{n}'$) and bounded, that (1) has at least one solution for some real η greater than one in magnitude.[11]

This is done by first making the usual decomposition of (1)–(3) into separate equations describing spin symmetric and antisymmetric modes. If

$$\nu=\tfrac{1}{2}(\nu_\uparrow+\nu_\downarrow),\ \nu^-=\tfrac{1}{2}(\nu_\uparrow-\nu_\downarrow),$$
$$F=F_{\uparrow\uparrow}+F_{\downarrow\uparrow}=F_{\uparrow\downarrow}+F_{\downarrow\downarrow},$$
$$\tfrac{1}{4}G=F_{\uparrow\uparrow}-F_{\downarrow\uparrow}=F_{\downarrow\downarrow}-F_{\uparrow\downarrow},$$
$$B=A_{\uparrow\uparrow}+A_{\downarrow\uparrow}=A_{\uparrow\downarrow}+A_{\downarrow\downarrow},$$
$$\tfrac{1}{4}C=A_{\uparrow\uparrow}-A_{\downarrow\uparrow}=A_{\downarrow\downarrow}-A_{\uparrow\downarrow},$$

* Work supported by the National Science Foundation under Contract No. GP-5517.
† Alfred P. Sloan Foundation Fellow.
[1] L. D. Landau, Zh. Eksperim. i Teor. Fiz. **30**, 1058 (1956) [English transl.: Soviet Phys.—JETP **3**, 920 (1957)].
[2] L. D. Landau, Zh. Eksperim. i Teor. Fiz. **32**, 59 (1957) [English transl.: Soviet Phys.—JETP **5**, 101 (1957)].
[3] L. D. Landau, Zh. Eksperim. i Teor. Fiz. **35**, 97 (1958). [English transl.: Soviet Phys.—JETP **8**, 70 (1959)].
[4] B. E. Keen, P. W. Matthews, and J. Wilks, Phys. Letters **5**, 5 (1963); Proc. Roy. Soc. (London) **A284**, 125 (1965).
[5] W. R. Abel, A. C. Anderson, and J. C. Wheatley, Phys. Rev. Letters **17**, 74 (1966).
[6] L. P. Gor'kov and I. E. Dzyaloshinskii, Zh. Eksperim. i Teor. Fiz. **44**, 1650 (1963) [English transl.: Soviet Phys.—JETP **17**, 1111 (1963)].
[7] A. M. Badalyan and L. A. Maksimov, Zh. Eksperim. i Teor. Fiz. **50**, 783 (1966) [English transl.: Soviet Phys.—JETP **23**, 518 (1966)].
[8] G. Baym, Phys. Rev. Letters **18**, 71 (1967).
[9] P. Nozières, *Theory of Interacting Fermi Systems* (W. A. Benjamin, Inc., New York, 1964), p. 16.
[10] A. A. Abrikosov and I. M. Khalatnikov, Rept. Progr. Phys. **22**, 329 (1959); D. Hone, Phys. Rev. **125**, 1494 (1962).
[11] There are some subtle points relating to the validity of (3) and (4) and the continuity of A in the forward direction. These are taken up in Appendices A and B.

then (1)–(3) separate into two sets of equations:

$$(\eta-\cos\theta)\nu(\hat{n})=\cos\theta\int\frac{d\hat{n}'}{4\pi}F(\hat{n}\cdot\hat{n}')\nu(\hat{n}'),\qquad(1')$$

$$F(\hat{n}\cdot\hat{n}')=B(\hat{n}\cdot\hat{n}')+\int\frac{d\hat{n}''}{4\pi}B(\hat{n}\cdot\hat{n}'')F(\hat{n}''\cdot\hat{n}'),\qquad(2')$$

$$0<\int\frac{d\hat{n}}{4\pi}|\chi(\hat{n})|^2+\int\frac{d\hat{n}}{4\pi}\int\frac{d\hat{n}'}{4\pi}\chi^*(\hat{n})F(\hat{n}\cdot\hat{n}')\chi(\hat{n}'),\qquad(3')$$

and a set (1″)–(3″) differing from (1′)–(3′) only in the replacements $\nu\rightarrow\nu^-$, $F\rightarrow G/4$, and $B\rightarrow C/4$. Condition (4) becomes

$$B(1)+\tfrac{1}{4}C(1)=0.\qquad(5)$$

Equation (1′) describes modes in which the two spin populations oscillate in phase ($\nu_\uparrow=\nu_\downarrow=\nu$), while (1″) gives modes with the spin populations 180° out of phase ($\nu_\uparrow=-\nu_\downarrow=\nu^-$). The former type is called ordinary zero sound, and the latter, spin zero sound, or spin waves.

The next step is to demonstrate that (1′) has a solution for real η greater than one in magnitude, provided that $B(1)>0$. The identical reasoning evidently will also require that (1″) has an appropriate solution provided $C(1)>0$. But (5) requires that precisely one of these conditions is met,[12] and hence one must have either ordinary or spin zero sound. [Nothing excludes the possibility of both, since the conditions $B(1)>0$ and $C(1)>0$ are only sufficient conditions.]

To show that a sufficient condition for the existence of ordinary zero sound is that B be positive in the forward direction,[13] we adopt a compact operator notation:

$$\int\frac{d\hat{n}'}{4\pi}F(\hat{n}\cdot\hat{n}')\nu(\hat{n}')\rightarrow F\nu,$$

$$\int\frac{d\hat{n}''}{4\pi}F(\hat{n}\cdot\hat{n}'')F(\hat{n}''\cdot\hat{n}')\rightarrow F^2,\text{ etc.}$$

The kinetic equation (1′) becomes

$$\eta\nu=\cos\theta(1+F)\nu;\qquad(6)$$

the relation between F and the forward scattering amplitude, (2′) is

$$F=B(1+F),\qquad(7)$$

and the stability condition is the requirement that $1+F$ be positive-definite:

$$1+F>0.\qquad(8)$$

If we define

$$\bar{\nu}=(1+F)\nu,$$

then using (7) we can rewrite (6) as

$$\bar{\nu}=(\cos\theta/\eta)\bar{\nu}+B\bar{\nu}=H_\eta\bar{\nu}.\qquad(9)$$

The condition for zero sound is that H_η have the eigenvalue unity for some real η between 1 and ∞ (the range $-\infty$ to -1 need not be considered, since the spectrum of the kinetic equation is symmetric about 0).

Now since H_η is Hermitian when η is real and $B(\hat{n}\cdot\hat{n}')$ is assumed to be bounded, the maximum eigenvalue λ_η, of H_η is a continuous real function of η. Furthermore since $H_\infty=B$, all the eigenvalues of which are less than 1 as a result of (7) and (8), λ_∞ must be less than 1. Therefore, if for some $\eta_0>1$,

$$\lambda_{\eta_0}>1,\qquad(10)$$

then λ_{η_1} will equal 1 for some η_1 between η_0 and ∞, and zero sound will propagate with velocity $\eta_1 v_F$.

When $B(1)>0$, (10) is established by a variational calculation:

$$\lambda_\eta>(\chi,H_\eta\chi)=\eta^{-1}\int\frac{d\hat{n}}{4\pi}|\chi(\hat{n})|^2\cos\theta+\int\frac{d\hat{n}}{4\pi}\int\frac{d\hat{n}'}{4\pi}\chi^*(\hat{n})B(\hat{n}\cdot\hat{n}')\chi(\hat{n}'),\qquad(11)$$

for arbitrary χ satisfying the normalization condition

$$\int\frac{d\hat{n}}{4\pi}|\chi(\hat{n})|^2=1.$$

As a trial function for H_η take

$$\chi(\theta,\phi)=\frac{A}{\eta-\cos\theta},\quad 0<\theta<\theta_0$$
$$=0,\quad \theta_0<\theta<\pi,\qquad(12)$$

where A is a normalization constant and θ_0 is a small enough angle that

$$B(x)>B_{\theta_0}>0,\qquad\cos2\theta_0\le x\le1.\qquad(13)$$

The possibility of (13) is insured by the assumption that B is positive and continuous in the forward direction.

With this trial function the inequality becomes

$$\lambda_\eta>1+\eta^{-1}\left(\int_{\cos\theta_0}^1[dx/(\eta-x)]\Big/\int_{\cos\theta_0}^1[dx/(\eta-x)^2]\right)\times\left[B_{\theta_0}\tfrac{1}{2}\eta\int_{\cos\theta_0}^1\frac{dx}{\eta-x}-1\right].\qquad(14)$$

The coefficient of the bracket exceeds zero for every $\eta>1$, while the coefficient of B_{θ_0} within the brackets becomes arbitrarily large as η approaches 1 from above.

[12] Barring, of course, the possibility that $B(1)=C(1)=0$, which, being an accidental degeneracy, slight changes in pressure could eliminate.

[13] Landau pointed out [in a remark following Eq. (15) in Ref. (2)] that in the weak-coupling limit, zero sound exists provided F is positive in the forward direction. Our result reduces to his in this limit, since when $F\ll1$ $B\approx F$.

This establishes (10) for some $\eta_0>1$, and hence that $B(1)>0$ is a sufficient condition for the existence of zero sound.[14]

If

$$B(x)=\sum_l B_l P_l(x),$$
$$F(x)=\sum_l F_l P_l(x),$$

then B in the forward direction $(B(1))$ is just $\sum_l B_l$. Since (7) has $B_l=F_l/[1+F_l/(2l+1)]$, a sufficient condition for ordinary zero-sound waves is

$$0<\sum_l \frac{F_l}{1+F_l/(2l+1)}. \qquad (15)$$

Evidently the corresponding condition for zero-sound spin waves is

$$0<\sum_l \frac{\frac{1}{4}G_l}{1+\frac{1}{4}G_l/(2l+1)}. \qquad (16)$$

By virtue of the exclusion principle (5), one and only one of these conditions must hold, and therefore at least one type of zero sound exists.[15]

APPENDIX A

The stability condition (3) has a rather different nature from the other results on which our conclusion rests. Derived Fermi-liquid properties fall into two categories: microscopic and phenomenological. The former follow rigorously for normal Fermi systems assuming only the validity of the appropriate perturbation expansions and the smoothness of quantities one has no reason to consider other than smooth; the latter follow from simple kinetic and thermodynamic arguments, always allowing for a change in the quasiparticle energies as a consequence of changes in their distribution function. The difference in the two approaches is that between Landau's first two papers,[1,2] presenting the phenomenological theory, and his third,[3] which gives its microscopic justification.

All of Eqs. (1)–(4) can be extracted from the tangle of vertex parts and Ward identities except for (3) which has so far been derived only in the phenomenological theory, as a consequence of the statement that the free energy is minimum in equilibrium. It would be surprising if condition (3) did not hold for the microscopic F, but it has not yet been proved.

One might hope to bypass the microscopic theory by deriving (3) directly from the kinetic equation (1), especially since one feels that a static instability should be reflected in a dynamic one, and hence in a nonreal root η, which *is* ruled out on microscopic grounds. However although it is easy to show that (3) is sufficient for the stability of (1), it is not by itself necessary.

Equation (3) is sufficient because[16]

$$(\eta-\cos\theta)\nu=\cos\theta F\nu$$

requires that

$$\eta=(\nu,\,(1+F)\nu)/(\nu,\,\sec\theta\nu),$$

which since F is real is only possible for nonreal η if numerator and denominator both vanish, contradicting the assumption that $1+F$ is positive definite. It is trivially *not* necessary for stability, however, since the same argument works if $1+F$ is negative definite. We do know, however, that $1+F$ is certainly not negative definite, since $1+F_0$ and $1+F_1/3$ have been proved microscopically to be positive as consequences of the positiveness of the specific heat and compressibility. Conceivably a proof could be found deriving an instability when some but not all of the eigenvalues of $1+F$ were negative, but I have been unable to do so. Nor does a direct proof exploiting Ward identities seem likely, since the perturbations one must consider are of a much more general sort than the gauge transformations, constant potentials, and such that go into deriving other Fermi-liquid relations.

On the other hand, the conventional definition of a normal Fermi system is barely consistent with (3) being violated. As the interaction is slowly turned up to full strength, any F_l growing from 0 to a value violating (3) will pass through $-(2l+1)$ en route, leading for this value of the coupling constant to an infinite value of B_l, and a zero-frequency solution [namely $\nu=P_l(\cos\theta)$] to the kinetic equation. Both phenomena suggest that at this point a phase transition to something other than a normal Fermi system will set in, so that the value $F_l<-(2l+1)$ is never reached.

APPENDIX B

Condition (4) though apparently a direct consequence of the exclusion principle actually involves some rather subtle considerations.[17] The exclusion principle tells us that the vertex function[18] is antisymmetric in its first two arguments:

$$\Gamma_{s_1s_2;s_1's_2'}(p_1,\,p_2;\,p_1',\,p_2')=-\Gamma_{s_2s_1;s_1's_2'}(p_2,\,p_1;\,p_1',\,p_2'), \qquad (17)$$

and hence, in particular,

$$\Gamma_{ss;ss}(p,\,p;\,p+k,\,p-k)=0. \qquad (18)$$

(We use the customary 4-vector notation in which p

[14] It is not, however, a necessary condition. If B has a small negative value in a small region about the forward direction, and somewhat less than one everywhere else, then this is consistent with the requirement that the eigenvalues of B be less than unity. For this B the right side of 11 will exceed unity for η near one, with a trial function that is constant in the forward direction and zero in the backward direction.

[15] It follows from the form of the trial function that the type of mode that must exist is the $m=0$ kind $[\nu(\theta,\phi)\equiv\nu(\theta)]$, and hence the mode is either a density wave or a spin density wave. The condition tells us nothing about the transverse $(m=1)$ modes (although with F_1 greater than 6 in He^3, the spin symmetric transverse mode is very likely to exist there, especially at higher pressures) or the more subtle $m\geq 2$ modes.

[16] As in the text it is simplest here to deal separately with the spin symmetric and antisymmetric cases.

[17] The analysis in this Appendix grew out of some very stimulating conversations with H. Wagner.

[18] We use the notation of Ref. 3.

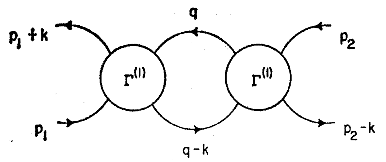

FIG. 1. A diagram contributing to the vertex function Γ.

stands for $\mathbf{p}$, p^0, etc.) If we define

$$\Gamma^\eta(\hat{n}\cdot\hat{n}') = \lim_{\substack{|\mathbf{k}|\to 0\\ k^0\to 0\\ k^0/|\mathbf{k}|=\eta v_F}} \times\{\Gamma_{ss;ss}(p_1,p_2;p_1+k,p_2-k)|_{\substack{\mathbf{p}_1=\hat{n}p_F,\mathbf{p}_2=\hat{n}'p_F\\ p_1^0=p_2^0=0}}\}, \quad (19)$$

then $A_{ss}(\hat{n}\cdot\hat{n}')$ is proportional to $\Gamma^0(\hat{n}\cdot\hat{n}')$ (Γ^k in the notation of Ref. 3), while $F_{ss}(\hat{n}\cdot\hat{n}')$ is proportional to $\Gamma^\infty(\hat{n}\cdot\hat{n}')$ (Γ^ω in the notation of Ref. 3).

Superficially, condition (4) would appear to follow from (19) with $\eta=0$, and (18). That this by itself is inadequate is best seen by noting that the same level of analysis entitles one to conclude that $F_{ss}(1)=0$. Thus we would have not only

$$\sum_l\left[\frac{F_l}{1+F_l/(2l+1)}+\frac{\frac{1}{4}G_l}{1+\frac{1}{4}G_l/(2l+1)}\right]=0, \quad (20)$$

but also

$$\sum_l[F_l+\tfrac{1}{4}G_l]=0. \quad (21)$$

However, subtraction of (20) from (21) yields

$$\sum_l\left[\frac{F_l^2/(2l+1)}{1+F_l/(2l+1)}+\frac{(\frac{1}{4}G_l)^2/(2l+1)}{1+\frac{1}{4}G_l/(2l+1)}\right]=0, \quad (22)$$

which is inconsistent with the stability condition (3) [the eigenvalues of F are $F_l/(2l+1)$].[19]

The trouble is that to conclude either (20) or (21), or for the validity of the analysis in the body of this paper, we require not

$$\Gamma^\eta(\hat{n}\cdot\hat{n})=0, \quad (23)$$

but

$$\lim_{\hat{n}'\to\hat{n}}\Gamma^\eta(\hat{n}\cdot\hat{n}')=0. \quad (24)$$

While (23) does indeed follow directly from (18) and (19), Eq. (24) is false unless $\eta=0$.

This can be seen by considering the contribution to $\Gamma(p_1,p_2;p_1+k,p_2-k)$ coming from a diagram such as that pictured in Fig. 1.[20] Here $\Gamma^{(1)}(p_1,p_2;p_1+k,p_2-k)$ is as defined in Ref. 3, and has the property that, as far as its contribution to the q integrations is concerned, its value as $\mathbf{k}$ and $k^0\to 0$, is independent of their ratio, $k^0/|\mathbf{k}|=\eta v_F$. The η dependence of this contribution to Γ in the limit of small k comes entirely from the two G lines connecting the two $\Gamma^{(1)}$'s, and is contained in a factor

$$\frac{\mathbf{q}\cdot\mathbf{k}/m^*}{k^0-\mathbf{q}\cdot\mathbf{k}/m^*}. \quad (25)$$

The diagram in $\Gamma(p_1,p_2;p_1+k,p_2-k)$ that cancels that shown in Fig. 1, when $p_1=p_2$, appears in Fig. 2. This diagram does not depend on the ratio of k^0 to $\mathbf{k}$, *except* in the case $p_1=p_2$, when it obviously must, since it is then identical to the diagram of Fig. 1. Thus the $p_1=p_2$ and $k=0$ limits do not commute for the diagram of Fig. 2, as a result of which our naive conclusions from the exclusion principle must be viewed with suspicion.

Suppose then we consider the case in which $p_1^0=p_2^0=0$, and k^0, $\mathbf{k}$, and $\mathbf{p}_1-\mathbf{p}_2$ are all very small. If the order in which these quantities vanish does not affect the contribution of $\Gamma^{(1)}$ to the integral, then the total contribution of Figs. 1 and 2 will contain a factor

$$\frac{\mathbf{q}\cdot\mathbf{k}/m^*}{k^0-\mathbf{q}\cdot\mathbf{k}/m^*}-\frac{\mathbf{q}\cdot(\mathbf{k}+\mathbf{p}_1-\mathbf{p}_2)/m^*}{k^0-\mathbf{q}\cdot(\mathbf{k}+\mathbf{p}_1-\mathbf{p}_2)/m^*}. \quad (26)$$

Now when $\mathbf{p}_1=\mathbf{p}_2$ this does indeed vanish as required by (18), but if the limit k^0, $\mathbf{k}\to 0$, $k^0/|\mathbf{k}|=\eta v_F$ is taken first, it becomes

$$\frac{\mathbf{q}\cdot\hat{k}/m^*}{\eta v_F-\mathbf{q}\cdot\hat{k}/m^*}+1, \quad (27)$$

which vanishes only when $\eta=0$. Thus

$$\lim_{\hat{n}'\to\hat{n}}\Gamma^\eta(\hat{n}\cdot\hat{n}')$$

will not in general vanish unless $\eta=0$.

It is not difficult to construct from the special cases of Figs. 1 and 2 an argument that one can still apply the exclusion principle to $\lim_{\hat{n}'\to\hat{n}}\Gamma^0(\hat{n}\cdot\hat{n}')$. The point is that if, in a diagram that cancels any singular[21] diagram in Γ when $p_1=p_2$, the limit $\mathbf{p}_1=\mathbf{p}_2$ follows the limit $k\to 0$, the result is always a contribution to Γ^0, regardless of the ratio of k^0 to $|\mathbf{k}|$, while the original diagram is contributing to Γ^η. Hence the cancellation occurs only when $\eta=0$.

Note added in proof: A. J. Leggett (private communication) has pointed out that a microscopic proof of the stability condition can be constructed out of his analysis in Phys. Rev. **140**, A1869 (1965). If one compares Eq. (18) of that paper with the usual spectral representation for K_ξ, one can show that if $1+F$ has a negative eigenvalue with eigenfunction $\chi(\hat{n})$, then the positivity of the long-wavelength spectral function will be violated when $\xi(\mathbf{p})=\chi(\hat{n})$, $\mathbf{p}=\hat{n}p_F$.

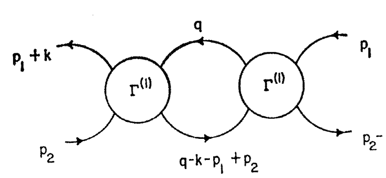

FIG. 2. The diagram contributing to Γ that cancels the diagram in Fig. 1, when $p_1=p_2$. The sign assigned to the diagram of Fig. 2 is opposite to that of Fig. 1.

[19] In the weak coupling limit one need not appeal to the stability condition to reject (22).

[20] We suppress the spin indices which do not play an important role in the analysis.

[21] One in which, like Fig. 1, the part of the diagram containing the incoming and outgoing lines for particle 1 can be separated from the part containing these lines for particle 2 by cutting just two G lines. Such diagrams depend on η.