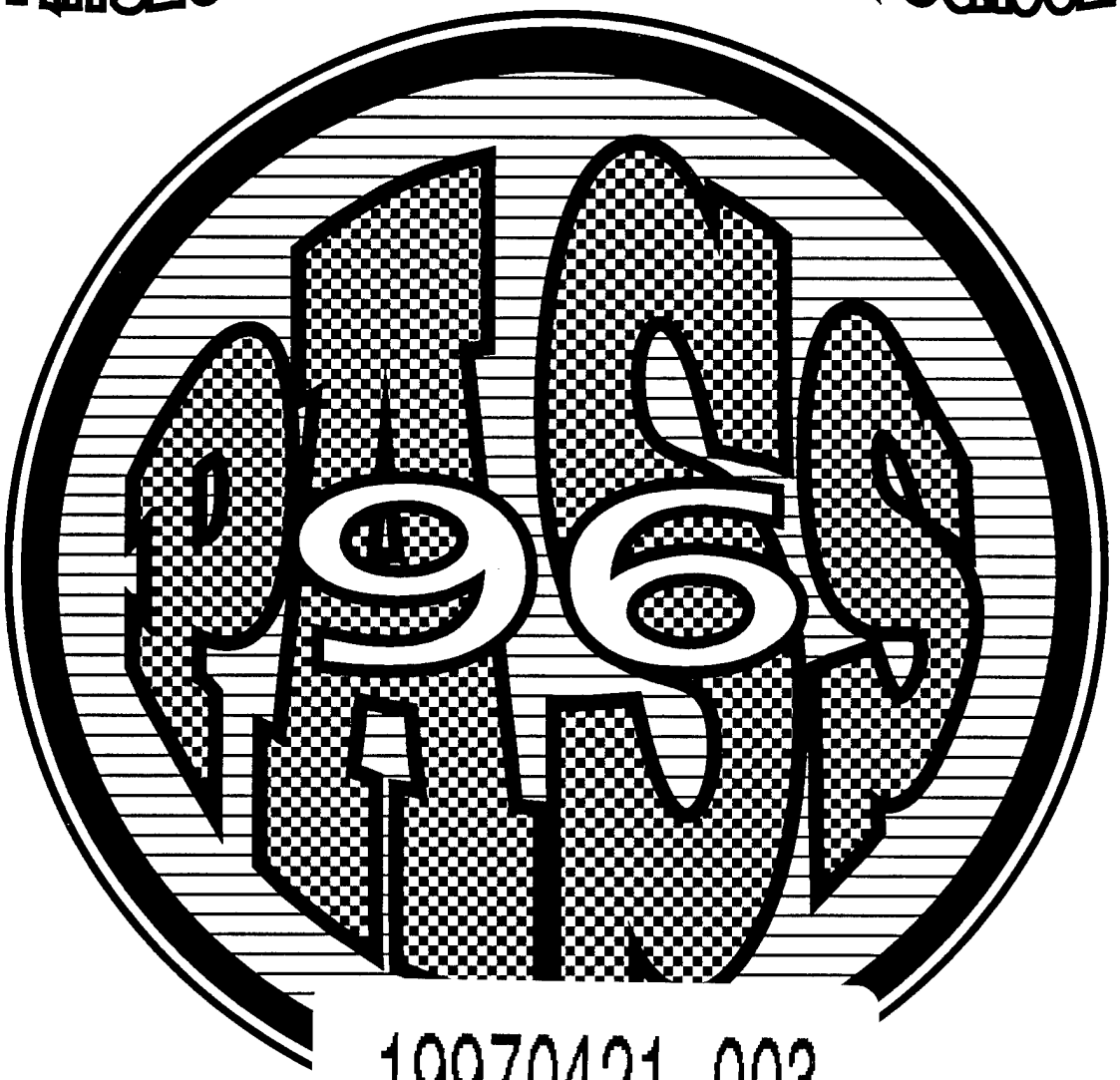
PHYSICAL ACOUSTICS SUMMER SCHOOL



19970421 003

ASILOMAR CONFERENCE CENTER

DTIC QUALITY INSPECTED 4

VOLUME III: BACKGROUND MATERIALS

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

1996 PHYSICAL ACOUSTICS SUMMER SCHOOL

VOLUME III: BACKGROUND MATERIALS

This work relates to Department of Navy Grant N00014-96-1-0033 issued by the Office of Naval Research. The United States Government has a royalty-free license throughout the world in all copyrightable material contained herein.

DTIC QUALITY INSPECTED 8

Copies of this three-volume proceedings can be obtained by contacting: Libby Cauthen, NCPA, University of Mississippi, University, MS 38677; voice: 601-232-5808; fax: 601-232-7494; e-mail: eacauthe@olemiss.edu.

TABLE OF CONTENTS

GENERAL BACKGROUND 1

MOLECULAR ACOUSTICS 19

ACOUSTICS DEMONSTRATIONS 89

RUS AND MATERIALS PHYSICS 91

PERIODIC, RANDOM AND QUASIPERIODIC MEDIA 144

SONOLUMINESCENCE 182

FUNDAMENTALS AND APPLICATIONS OF NONLINEAR ACOUSTICS 183

ATMOSPHERIC ACOUSTICS 235

SENSOR PHYSICS: SIGNALS AND NOISE 305

THERMOACOUSTIC ENGINES AND REFRIGERATORS 327

REPORT DOCUMENTATION PAGE 377

PHYSICAL ACOUSTICS SUMMER SCHOOL - 1996

General Background

Steven L. Garrett
United Technologies Corporation Professor of Acoustics
Graduate Program in Acoustics - Applied Research Laboratory
The Pennsylvania State University

ABSTRACT

This lecture is intended to provide students with a common background which would be useful in understanding the other eight lectures. Sound waves exist only in matter - sound does not propagate in a vacuum (although M. Greenspan got pretty close!). For that reason, this lecture will begin with the thermodynamic description of gases. Since gases also conduct heat and dissipate shear stresses, simple kinetic theory will be added to the description to permit analysis of the adiabatic and isothermal limits for sound speed and acoustical attenuation. In order to avoid detailed mathematics, the principle of similitude (dimensional analysis) will be exploited. The results of this analysis will be applied to loss mechanisms in planewave resonators, Helmholtz resonators and bubbles. This should provide the necessary background for Molecular Acoustics, Sonoluminescence, Signals & Noise, Nonlinear Acoustics and Thermoacoustics. Discussion of the isotropic elasticity of solids and the torsional, flexural and longitudinal modes of a long thin bar should provide an introduction to Resonant Ultrasound Spectroscopy, Materials Physics and Porous Media.

Precision acoustic gas analyzer for binary mixtures

E. Polturak, S. L. Garrett,[a] and S. G. Lipson

Department of Physics, Technion, Israel Institute of Technology, Haifa, Israel

(Received 18 June 1986; accepted for publication 3 July 1986)

We describe a cylindrical acoustic resonator which has been used for the determination of the molar concentration of a binary mixture of gases. We achieved an absolute accuracy better than 0.001 independent of the concentration or the type of gas. The practical factors limiting the accuracy are discussed in detail and their effect on the performance has been evaluated using gas samples prepared by precisely measured partial pressures. The device is routinely used to measure the concentration of ^{3}He–^{4}He mixtures.

INTRODUCTION

The measurement of sound speed in gas mixtures has been used to determine the concentration of binary and pseudo-binary[1] gas mixtures since 1884 when Forbes and Blaikley first made use of the subsequently widely adopted idea of passing mine exhaust air through an organ pipe to call attention to the presence of lighter (potentially explosive) gases such as hydrogen and methane.[2] Since that time several different schemes, both acoustic and electroacoustic, have been implemented and patented,[3] although no commerical unit using sound speed has ever been marketed. The appearance of a short note by Kagiwada and Rudnick[4] on the use of the sound speed to determine the isotropic concentration of the ^{3}He–^{4}He mixtures which are widely used in low-temperature experiments, generated renewed interest in sonic gas analysis within that community and has lead to the design of several devices for that and other applications related to low-temperature research. A cylindrical resonator capable of $\pm 1\%$ determination of ^{3}He–^{4}He concentration is described by Fraser[5] and a time-of-flight device which used a long tube and gave the same resolution is described by Guillon *et al.*[6] A cylindrical resonator and associated electronics, which are capable of continuously measuring air contamination as small as 1/4% in He gas recovered for reliquification[7] is another example of the application of this technique to cryogenic systems.

The first attempt to construct a precision analyzer for ^{3}He–^{4}He mixtures utilized a pair of spherical resonators.[8] That system had a quality factor $Q = 1200$ at 23 kHz which resulted in a sensitivity of 1 ppm and a stability of 10 ppm over 10 h, but it was only capable of an accuracy in isotopic ratio of 1×10^{-3} due to outgassing of plumbing, O-rings, transducers, etc. Another similar double-sphere system for the same application is described by Brooks and Hallock,[9] but their tests with a He–Ne mixture include no estimate of the accuracy obtained.

It is clear that the ultimate exploitation of this technique will require a gas-filled spherical resonator.[10] Such resonators have been extremely successful in the measurement of certain thermophysical properties,[11,12] but for most applications the additional complexities in machining and transduction introduced by this geometry are not justified, especially if nonacoustic factors such as outgassing are the dominant source of inaccuracy. It is the purpose of this paper to present the results of experiments which employ an optimized cylindrical resonator, in a constant temperature bath, to produce accuracies at the 0.1% level, which represent an order-of-magnitude improvement over previous designs,[5,6] using a system that can be built without a major investment in fabrication or instrumentation.

I. THEORY

The square of the adiabatic sound speed in an ideal gas is given by $C^2 = (\partial p/\partial\rho)_s = \gamma RT/M$, where R is the universal gas constant, γ is the ratio of the specific heats, and T is the absolute temperature; p, ρ, and s are the pressure, density, and entropy of the gas. For a binary gas mixture composed of two ideal gas species, with molar fraction x, the molecular weight of the mixture $M = M_1(1-x) + M_2x$, where M_1 is the molecular weight of the first components and M_2 is the molecular weight of the second. The expression for an effective specific heat ratio, $\gamma(x)$, for the mixture, is more complicated since it is the heat capacities themselves which are the extensive thermodynamic quantities, not the specific heat ratio[13]

$$\gamma(x) = \gamma_1 \frac{1 + (C_{p2}/C_{p1} - 1)x}{1 + (C_{v2}/C_{v1} - 1)x}. \tag{1}$$

For simplicity in expression of the sound speed in the mixture, and in the analysis of the results, we have chosen to treat the effective ratio of specific heats as the molar weighted average of the individual ratios: $\gamma(x) = \gamma_1(1-x) + \gamma_2 x$. Using this approximation, the sound-speed ratio for the mixture can be written

$$\frac{C(x)^2}{C(0)^2} = \frac{1 + [(\gamma_2 - \gamma_1)/\gamma_1]x}{1 + [(M_2 - M_1)/M_1]x}. \tag{2}$$

Here, $C(0)$ equals C_1, the speed of the pure (reference) gas. For mixtures which contain only ideal components with equal numbers of degrees of freedom, Eq. (2) is exact. For a monatomic/diatomic mixture, the maximum relative error in $\gamma(x)$ occurs at $x = 0.5$ and is 2.2%. For example, for the H_2–^{4}He mixture ($x < 0.1$) analyzed in this article the effects of the approximation are negligible.

The usefulness of Eq. (2) depends on the accuracy with which $C(x)$ can be determined and on the degree of exclu-

 0034-6748/86/112837-05$01.30 © 1986 American Institute of Physics

sion of impurity gases from the experimental setup. In addition, the derivation of Eq. (2) relies on the temperature being held constant. Obviously, the sensitivity of the method depends on the type of gases used. Expanding Eq. (2) for small x, we find

$$\frac{\delta C(x)}{C(0)} = \frac{\delta x}{2}\left(\frac{\gamma_2-\gamma_1}{\gamma_1} - \frac{M_2-M_1}{M_1}\right). \tag{3}$$

For any two gases, $0.3 > (\gamma_2-\gamma_1)/\gamma_1 > -0.3$. The mass difference factor in Eq. (3) gives greater sensitivity to gas mixtures whose components have very different masses. However, if M_1 and M_2 are relatively close, Eq. (3) indicates that the measurement can be seriously affected by small amounts of impurities having a very different value of M, to which $C(x)$ is much more sensitive. This consideration becomes important when applying the method to mixtures of the helium isotopes, because small amounts of air or other heavy gases as impurities can affect the result greatly. For ^{3}He–^{4}He mixtures, Eq. (3) translates the requirement of determining x with a 10^{-3} accuracy into a demand for an *absolute* accuracy and reproducibility of 10^{-4} in the sound velocity measurement. Note that an impurity content of less than 4×10^{-5} of air would change the velocity of sound by the same amount. It is our opinion that the impurity problem is the main cause for the disappointing results achieved by several groups while trying to implement techniques of this type.

An important consideration when designing a resonator is the optimization of its Q. The highest Q's, of the order of several thousand at room temperature, have been obtained using spherical resonators.[8–12] However, because of the sources of error that have been mentioned above, it is impossible to take advantage of such a high Q, and a cylindrical resonator with somewhat lower Q but much easier to construct becomes an attractive choice. The Q of the cylindrical resonator is limited by several energy dissipation mechanisms: heat loss at the walls due to heat conduction during an adiabatic compression–decompression cycle,[14] viscous damping of the motion of the gas in contact with the cylinder walls,[14,15] transducer losses, and losses through the openings in the resonator (fill tubes, etc.). If the fill tubes are located at pressure nodes, the losses through them are negligible. Assuming the first two mechanisms to operate independently, the total loss can be represented by the total quality factor $Q_{\rm tot}$ given by

$$1/Q_{\rm tot} = \delta_{\eta e}/a + 2\delta_K/L, \tag{4}$$

where a and L denote the radius and length of the resonator and $\delta_{\eta e}$ and δ_K are the effective viscous[15] and thermal[14] penetration depths. Varying Eq. (3) with respect to L (at fixed a), and remembering that $\delta_{\eta e}$ and δ_K depend on L through their independence on the resonant frequency, we obtain an aspect ratio condition for optimum Q:

$$2a/L = (\eta_e C_p/K)^{1/2}, \tag{5}$$

where

$$\eta_e = \eta\left[1 + (\sqrt{\gamma} - 1/\sqrt{\gamma})\sqrt{K/\eta C_p}\right]^2$$

is the effective viscosity[15] (which includes thermal wall losses), η is the ordinary shear viscosity, and K is the heat conductivity.

Substituting numbers, we find that the optimum choice gives $2a/L = 1.35$ for ^{4}He and 1.22 for air at room temperature. We chose $2a/L = 1$. Our choice of this value was made to guarantee that the first azimuthal mode would occur at a frequency greater than that of the gravest plane-wave mode. Around 20 kHz, we measured Q's of 300–400, which are some 30%–50% lower than predicted by Eq. (4).

II. RESONATOR

The length of the resonator, details of which are shown in Fig. 1, was chosen as 5.0 cm to give a convenient 10-kHz fundamental resonance frequency for ^{4}He gas at room temperature. Two identical resonator cavities were machined through a single aluminum block. The idea was to use one of them as a reference which would automatically compensate for ambient temperature changes.[8,9]

To enable low-temperature operation, the end caps were sealed with indium O-rings. The center electrodes were insulated from the caps using Stycast 1266 epoxy loaded with $CaCO_3$ powder, to match its expansion coefficient to that of the Al. Capacitive electret transducers mounted on the two end caps, were fabricated from 12-μm-thick Teflon, metallized on one side. To create trapped charge, the Teflon was sandwiched between two flat metal electrodes with an additional layer of Mylar to prevent breakdown, and polarized

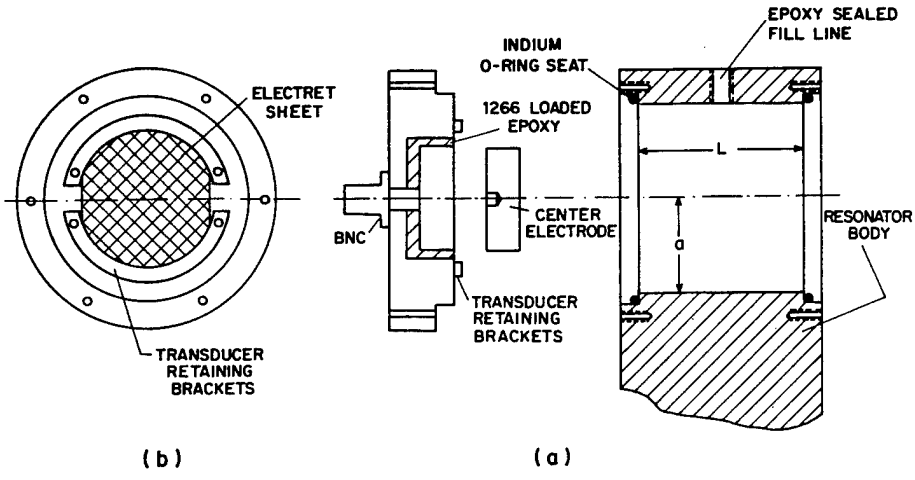


FIG. 1. (a) Schematic view of one of the cavities of the double resonator machined out of a single Al block. (b) Transducer clamping arrangement to reduce trapping of gas. All screws have vent holes.

for 1 h with a 10^6-V/cm field. The measured magnitude of the trapped charge was equivalent to that created by an external bias of 150 V. The Teflon strips were held down against the electrode by two semicircular brackets, as shown in Fig. 1. The capacitance of these transducers was about 1000 pF. We have also used transducers made from a commerical piezoelectric polymer film.[16] These seem to retain their polarization much better than the "homemade" electrets, which need to be repolarized every few months. It seems that carrying out the polarization at a temperature of about 100 °C and cooling the Teflon before removing the field gives better results regarding the stability of the trapped charge. The signal level available at resonance was 0.16 mV per volt of excitation.

III. OPERATION

The performance of the resonator with pure gases is shown in Fig. 2. In the region of high Q (around 20 kHz), the resonance could be located visually on an oscilloscope screen or with an ac voltmeter with an accuracy of 1×10^{-5}

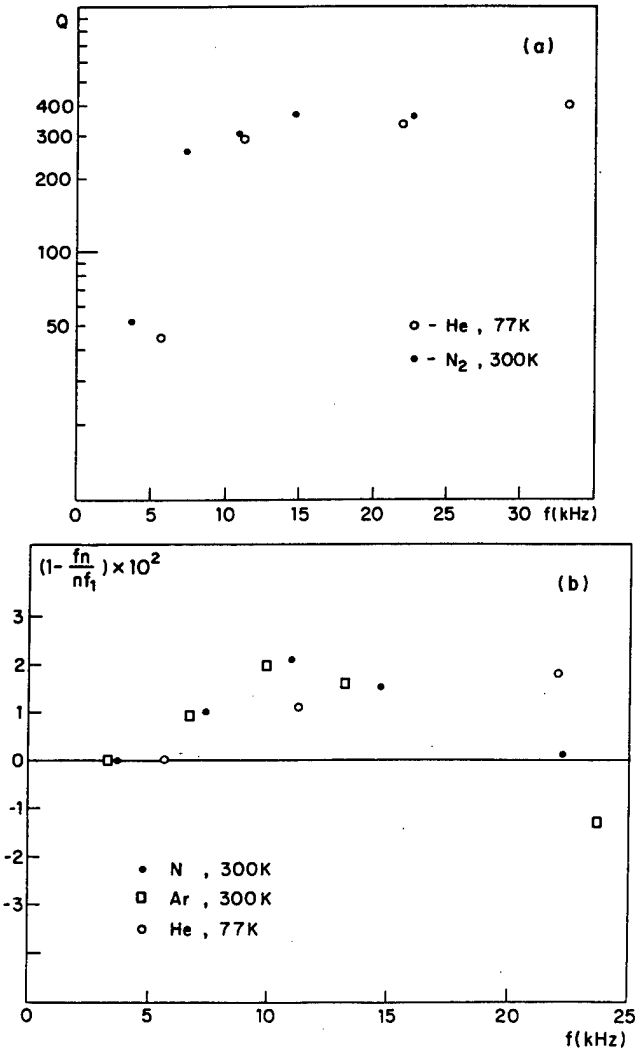


FIG. 2. (a) Q of the resonator vs frequency for helium and nitrogen. (b) Deviations of the resonator from perfect harmonicity for different gases. f_1 is the fundamental resonant frequency, f_n is the resonant frequency of the nth harmonic.

TABLE I. Typical sources of error.

Source of error	Magnitude parts of f	Temperature of test	Type of gas used in test
Thermal drift	5×10^{-6}	77 K	^{4}He
Thermal drift	5×10^{-7}	273 K	N_2
Outgassing	$<1\times10^{-6}$	77 K	^{4}He
Pressure cycling	2×10^{-5}	77 K	^{4}He
Nonharmonicity	$5\times10^{-4}x_p$	273 K	O_2–N_2
Nonharmonicity	$<2\times10^{-4}x_p$	77 K	H_2–^{4}He

of f.[17] Changes in the resonant frequency could be followed with a 3×10^{-7} precision. A convenient way of doing this was to make the resonator part of a phase-locked loop, with the quadrature component of the detected signal used to program a VCO which, when amplified, then drives the resonator. We could not maintain this precision under practical operating conditions involving moving gas in and out of the resonator, since when the resonator was evacuated, the signal disappeared and the phase-locked loop lost track. The absolute accuracy with which the resonance could be repeatedly located was within 10^{-5} of f. (This precision could perhaps be utilized in an application which required continuous monitoring of the composition changes in flowing gas.[7]) We now describe the various tests to which the resonator was subjected in order to determine the accuracy of the method.

When the resonator contains a fixed sample of gas, the resonant frequency is affected by temperature drift. The change of temperature needed to shift the frequency by 1×10^{-4} is only 60 mK at 300 K or 15 mK at 77 K. One way to correct for this effect is to have two resonators machined into a single block of high thermal conductivity metal. One contains a reference gas and the other the unknown mixture. Since their temperatures are the same, their resonant frequency will drift at the same rate and can thus be corrected. The other alternative is to use a single resonator and to regulate its temperature. The latter method saves the electronics needed to run the second resonator, and the extra machine shop time needed to construct it. We found that putting the resonator in an unregulated liquid-nitrogen (or ice water) bath gave a thermal drift rate which was both very small and monotonic. Thus, we ended up using a single resonator only. The errors caused by thermal drift are listed in Table I.

Another, more serious, source of drift was a slow change in the composition of the gas due to the outgassing of the inner surfaces of the resonator. Since the gas emanating from the walls is mostly air, the resulting frequency drift is proportional to $[(M_{\text{air}}-M_1)/M_1]x$, where M_1 is the molecular mass of the gas inside the resonator, and x is the air concentration. (The effect of outgassing can be distinguished from that of the temperature drift, since it always lowers the resonant frequency with light gases.) One major source was gas trapped behind the Teflon electrets. To minimize this source of outgassing the electret transducers were clamped over only a part of their circumference (see Fig. 1), so that any gas trapped between the Teflon and the electrode could be pumped out. An all-metal gas handling system was constructed, and the whole assembly was given a mild bake,

which the epoxy parts and the transducer could withstand, and pumped out for a day with a diffusion pump. This reduced the outgassing sufficiently that heavier gases (N_2, O_2, Ar, etc.) could be measured without any significant error. Cooling the resonator to 77 K reduced the outgassing further, and only at this temperature could the light gas mixtures be measured with the same precision (see Table I).

In a typical application, pure gas is introduced into the resonator for calibration. After the pure gas has been removed, the resonator is filled with an unknown mixture at the same pressure, and the resonant frequency measured. The process involves a pressure excursion of the order of one bar and the accuracy depends on the dimensional stability of the resonator: using ^{4}He gas, we found a relative shift of the resonant frequency by about 3.5×10^{-6} P/Torr. From the virial coefficient for ^{4}He we estimate that almost all of this effect is caused by the pressure dependence of the velocity of sound.[18] We used a pressure gauge capable of 5-Torr resolution to reset the pressure. With the aid of this gauge, we found that f would reproduce to an accuracy of 2×10^{-5} after repeated pressure excursions of 1–1.5 bars.

Finally, the results have to be corrected for the imperfect harmonic behavior of the resonator, an example of which is shown in Fig. 2(b). The deviations from harmonicity arise because of imperfect cylindrical geometry, viscous and thermal conductivity corrections to the speed of sound, and coupling between the plane-wave modes and various other modes of the resonator having resonances near frequencies of interest. These corrections are quite small, but still significant at the level of accuracy required of the device. We can write empirically

$$f=[(1+\alpha(f_p,M)]f_p, \tag{6}$$

where f_p is the resonant frequency one would measure with a perfect resonator and $\alpha(f_p,M)$ is a small empirical correction function which depends both on the frequency and on the molecular mass of the gas. Values of α for different gases and harmonics are shown in Fig. 2(b).

Filling the resonator with a mixture shifts the resonant frequency by δf from f, the frequency measured with pure reference gas. The relative change, $\delta f/f=\delta C/C(0)$, can be substituted into Eq. (3) to give an apparent concentration x. Now, by varying both sides of Eq. (6) and dividing by f, we get

$$\frac{\delta f}{f}=\frac{\delta f_p}{f_p}+\frac{\delta\alpha}{1+\alpha}. \tag{7}$$

Similarly to $\delta f/f$, $\delta f_p/f_p$ can be expressed in terms of the true concentration x_p. Expanding $\delta\alpha$ to first order in M and f, and neglecting α in the denominator on the RHS of Eq. (7), we obtain after some algebra

$$x_p=x-\left[\left(\frac{\partial\alpha}{\partial f}\right)_M f-2\left(\frac{\partial\alpha}{\partial M}\right)_f M\right]x_p. \tag{8}$$

The partial derivatives are estimated from the measured values of several resonance frequencies and for several gases at f and M appropriate to $x=0$. Using f instead of f_p introduces a negligible error. From our calibration data, we found that for O_2–N_2 mixtures used in this work, this difference was $x_p-x=7.5\times10^{-3}x_p$. For H_2–^{4}He mixtures and $x<0.1$ it was negligible. The uncertainty of this correction is listed in Table I. It is important to point out that if this correction is to be applied at all, both terms on the RHS of Eq. (8) are separately significant, although they almost offset each other. One needs, therefore, to calibrate the resonator with several gases.

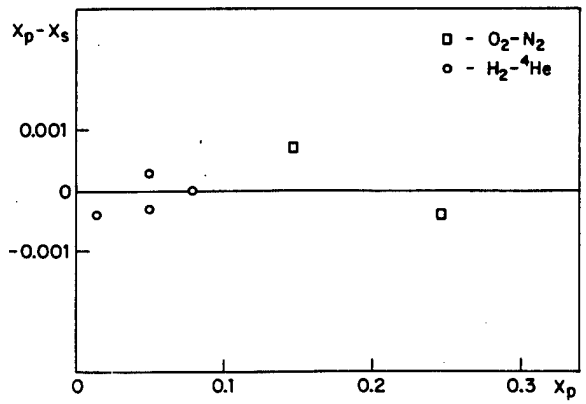


FIG. 3. Performance of the resonator with preprepared gas mixture. x_s is the concentration determined acoustically. x_p by partial pressures.

IV. EXPERIMENTAL RESULTS

To test the resonator we prepared a series of "standard" mixtures of H_2–^{4}He and O_2–N_2. These were made using a precise pressure gauge.[19] There are several points which are noteworthy: first, it is not always appreciated that without stirring it takes nearly a day for two gases to become thoroughly mixed by diffusion inside a few liter volume. Second, cold traps containing adsorbents located in the path of the gas entering the resonator tend either to desorb unwanted gas or to adsorb preferentially one of the species of the mixture, and should not be used. We found a simple U-tube trap to work very well. For ^{3}He–^{4}He mixtures we cooled the trap to liquid-helium temperature and for H_2–^{4}He mixtures to liquid-nitrogen temperature. For O_2–N_2 mixtures, the trap was located in the ice + water bath. The results of these tests are shown in Fig. 3. One can see that a uniform accuracy, better than 10^{-3}, was achieved independent of x and gas species. The combined uncertainties in Table I correspond to about half that value. Since the nominal accuracy of the pressure gauge used to prepare the "standards" was 0.025% the combined error of the three measurements necessary to prepare a mixture can easily account for the difference. To date, the resonator has primarily been used to identify the multitude of ^{3}He–^{4}He mixtures of unknown composition usually found in low-temperature laboratories.

In conclusion, we have built and tested a cylindrical acoustic resonator capable of determining the composition of a binary gas mixture with an absolute accuracy better than 0.001 independent of the concentration.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the expert technical assistance of M. Ayalon. This work was supported by the Fund for Basic Research of the Israel National Academy of Sciences and by the U.S.–Israel Binational Science Foundation.

S.L.G.'s visit to the Technion was supported by the Rosen Fund for Solid State Physics.

[a] On a sabbatical leave from Naval Postgraduate School, Monterey, CA 93943.

[1] A pseudobinary gas mixture is composed of two gases, one of which is itself a mixture, such as air, which has a relatively stable composition.
[2] P. W. Mullen, *Modern Gas Analysis* (Interscience, New York, 1955), Chap. X.
[3] W. Mikelson, U.S. Patent 2,283,750 (19 May 1942).
[4] R. S. Kagiwada and I. Rudnick, J. Low Temp. Phys. **3**, 113 (1970).
[5] J. C. Fraser, Rev. Sci. Instrum. **11**, 1692 (1972).
[6] F. Guillon, J. P. Harrison, and A. Tyler, J. Phys. E **14**, 1148 (1981).
[7] S. L. Garrett, G. W. Swift, and R. E. Packard, Physica **107B**, 601 (1981).
[8] R. Keolian, S. Garrett, J. Maynard, and I. Rudnick, J. Acoust. Soc. Am. **64**, S61 (1978); Bull. Am. Phys. Soc. **24**, 623 (1979).
[9] J. S. Brooks and R. B. Hallock, Rev. Sci. Instrum. **54**, 1199 (1983).
[10] M. R. Moldover, J. B. Mehl, and M. Greenspan, J. Acoust. Soc. Am. **79**, 253 (1986).
[11] M. R. Moldover, M. Waxman, and M. Greenspan, High Temp.–High Press. **11**, 75 (1979).
[12] J. B. Mehl and M. R. Moldover, J. Chem. Phys. **74**, 4062 (1981); **77**, 455 (1982).
[13] R. Holmes and W. Tempest, Proc. Phys. Soc. (London) **75**, 898 (1960).
[14] L. Landau and E. Lifshitz, *Fluid Mechanics* (Pergamon, New York, 1959), problem 1, p. 302.
[15] L. Kinsler and A. Frey, *Fundamentals of Acoustics*, 2nd ed. (Wiley, New York, 1958), Chap. 9.
[16] "Kynar" piezoelectric PVF_2 film is manufactured by Pennwalt Corp., 900 First Ave., P.O. Box C, King of Prussia, PA 19406.
[17] We used a HP-3325A synthesizer to drive the resonator and detected the signal with either a vector lock-in amplifier or simply an oscilloscope.
[18] See, for example, K. Herzfeld and T. Litovitz, *Absorption and Dispersion of Ultrasonic Waves* (Academic, New York, 1959). Values of the virial coefficients were taken from J. Dymond and E. Smith, *The Virial Coefficients of Gases: A Critical Compilation* (Clarendon, Oxford, 1969).
[19] "Digiquartz Pressure Transducer" is made by Paroscientific Inc., 4500 148th Ave., N.E. Redmond, WA 98052.

Resonant acoustic determination of elastic moduli

Steven L. Garrett
Physics Department, Code 61 Gx, Naval Postgraduate School, Monterey, California 93943

(Received 11 December 1989; accepted for publication 23 February 1990)

A long rod of circular or elliptical cross section can be selectively excited in the torsional, longitudinal, and flexural resonant modes using a single pair of electrodynamic transducers consisting of coils glued to each end of the bar and placed in the field of a magnet. The elastic moduli of the material can be simply determined to high accuracy by measuring the resonance frequencies of these modes and the mass and physical dimensions of the rod. Since the longitudinal and flexural modes both yield values for the Young's modulus, the system has a built-in redundancy that makes the interpretation of the data more robust. The large signal amplitudes make automated measurement of the temperature dependence of the moduli simple through the use of a phase-locked loop to track the change in the resonance frequency with temperature.

PACS numbers: 43.40.Cw, 43.20.Ye, 62.20.Dc, 07.10 + i

INTRODUCTION

The accurate measurement of the elastic constants of materials and their dependence on temperature, static pressure, and other ambient parameters is important in many fields of science and engineering research, as well as in product design and quality control. This is particularly true for hydrophones, since most transduction schemes involve the measurement of the deformation of some solid material in response to a change in pressure. The elastic modulus relates these strains to the applied stresses, so its value and its temperature dependence are important design parameters. Interest in the measurement of these moduli in the acoustical transduction community, particularly for castable polymers, has increased recently due to the development of interferometric fiber-optic hydrophones,[1-2] which measure the strain in the optical fiber induced by pressure changes in the fluid. Castable polymers are attractive for several hydrophone designs that use shell structures[3-7] or encapsulation[8-11] for strain enhancement, since the optical fibers can be cast directly into those materials as well as for more conventional piezoelectric hydrophones.[12,13]

Manufacturer's specifications for elastic constants of castable polymers are not particularly useful for the hydrophone designer since they are usually determined by static techniques, rarely contain more than one modulus (two are the minimum required to uniquely specify the elastic response of an isotropic material[14]), vary widely depending upon sample preparation (e.g., catalyst) and cure temperature, and never contain information about the temperature dependence of the moduli. It is also important in hydrophone design to measure the elastic moduli at the frequencies of intended operation, since it is well known that the static and dynamic moduli of plastics can differ substantially[15] due to the existence of relaxation time effects that can make the modulus measured by conventional quasi-static stress–strain curves significantly lower than the dynamic modulus. Since it is the dynamic modulus that determines the acoustic sensitivity, and the static modulus that determines the deformation due to increased operating depth, it is important that both are known to the hydrophone designer.

A recent series of measurements on 31 elastomeric samples by Lagakos *et al.*[16,17] illustrate how the assumption of a simple form for the frequency dependence of the Young's modulus can lead to unacceptable disagreement between the measured moduli and the predicted value. For the case of Uralite 3130, a polyurethane used in a prototype fiber-optic planar flexible hydrophone design,[9] the variation between the model and measured modulus was in excess of 750%. One must question the validity of their temperature coefficient of the modulus measured at 1 MHz when extrapolated to operating frequencies of less than a kilohertz.

Because of the number of variations in the preparation that are available to the experimentalist (cure, catalyst, filling fraction for composites, etc.), it is important to have a convenient technique for measuring at least two moduli for reasonably small samples. Since the derivation of a third modulus from the measurement of two others can be very sensitive to the values of those moduli (particularly for samples with Poisson's ratio close to 0.5), it is important that the measurement technique have very high precision. In the following sections, a technique is described that is convenient, accurate, precise, and economical, that depends upon the measurement of the frequencies of the longitudinal, flexural, and torsional resonant modes of a single rod-shaped sample of circular cross section using the same two transducers to excite and detect all three modes.

The fact that the technique is resonant insures high signal-to-noise ratio while the fundamental measurement being a frequency means that one can obtain extremely high precision with an inexpensive instrument (i.e., a frequency counter). The technique can be used with both insulating or conducting samples that are not ferromagnetic. An additional attractive feature of this technique is the fact that Young's modulus can be determined independently in two ways by the measuring the frequencies of the longitudinal and flexural modes. This provides the experimentalist with immediate feedback on the self-consistency of his/her data and analysis. (A man with one watch thinks he knows the time; a man with two is never sure!)

The technique for measurement of the torsional mode is a refinement of one developed first by Barone and Giacomini[18] to study the modes of vibration of bars having vari-

able cross section. A similar arrangement was used later by Leonard[19] to disprove the existence of the "Fitzgerald effect"[20–22] by measuring the attenuation of torsional waves in Teflon™. I was first introduced to the technique described here, which excites and detects the flexural and longitudinal modes in addition to the torsional mode with the same transducer, by Professor Isadore Rudnick who used a similar apparatus in a teaching lab experiment[23] in an advanced undergraduate acoustics laboratory at UCLA.

The technique can be extended to measure the complex modulus in a manner similar to that used by Barmatz *et al.*,[24] if the quality factor Q, or free decay time is measured in addition to frequencies or if velocity measurements are made over a large enough frequency span that the Kramers–Kronig[25] relations can be integrated. The measurement of Q introduces additional experimental difficulties, since care must be taken to insure that losses through the suspension system used to support the rod are not significant compared to those intrinsic to the sample material under study. Because the strain distribution is known for each mode, these suspension losses can be reduced to insignificant levels for all but the fundamental torsional and longitudinal modes, since the two support points for the bar can be adjusted to occur arbitrarily close to velocity nodes.

I. MODES OF A SOLID ROD OF CIRCULAR OR ELLIPSOIDAL CROSS SECTION

A uniform, rod-shaped sample of a homogeneous, isotropic solid having circular cross section of diameter d and length L, which is significantly greater than its diameter, will propagate three independent waves if their wavelengths λ are much greater than d. These modes will exhibit resonances at appropriate frequencies, depending upon the boundary conditions imposed on the ends of the rod. For this application, the simplest and most reproducible boundary condition to impose on the rod in a laboratory measurement is that of zero stress and zero moment at both ends, which is commonly referred to as the free–free boundary condition.

The solution to the wave equation for all three modes, subject to the free–free boundary condition, is available in a variety of textbooks, although surprisingly few treat all three modes and those that do[26–31] were all written before 1970. The more popular contemporary textbooks in acoustics such as Morse,[32] and Morse and Ingard[33] treat only flexural waves; Ingard,[34] and Feynman[14] cover only longitudinal and torsional waves; Kinsler *et al.*[35] treat only flexural and longitudinal modes. The more recent graduate level texts[36,37] ignore the subject altogether!

A. Nondispersive modes

The displacements associated with the longitudinal and torsional modes satsify an ordinary second-order wave equation[14] so, for a free–free boundary condition, the resonances are harmonically related and correspond to integral numbers of half-wavelength contained within the length of the rod. This assumed boundary condition does not take the added mass of the transducers and their adhesive into account, but this is generally a small effect since their additional mass is rarely more than a few percent of the bare mass of the bar. For the moment, we will neglect this effect which will be discussed in the Appendix. The transducer mass loading will lead to the introduction of an effective length, L_{eff}, which provides a first-order correction if accuracy of greater than a few percent is required.

The phase speed for the longitudinal waves c_L is given by Young's modulus E, and the mass density ρ, of the rod material:

$$c_L = \sqrt{E/\rho}. \tag{1}$$

This result is independent of the cross-sectional shape of the rod as long as the initial assumptions ($\lambda \gg d \ll L$, homogeneous, isotropic) are met. Imposition of the free–free boundary condition leads to a series of harmonic modes whose frequencies f_n^L, are given by

$$f_n^L = nc_L/2L; \quad n = 1,2,3\ldots . \tag{2}$$

The previous two equations can be combined to provide an expression for the Young's modulus of the bar in terms of its density, length, and the ratio of the frequency of the nth mode to its mode number n:

$$E = 4\rho L^2 (f_n^L/n)^2. \tag{3}$$

Similarly, the phase speed for the torsional waves c_T, is given by the shear modulus G, and the mass density ρ, of the rod material:

$$c_T = \sqrt{G/\rho}. \tag{4}$$

From stability requirements, this speed is always smaller than the longitudinal phase speed by a factor between $\sqrt{2}$ and $\sqrt{3}$ depending upon the value of Poisson's ratio.[26] This equation for the torsional wave phase speed is strictly true only for a rod of circular cross section. If the rod were elliptical, with major and minor radii a and b, respectively, the speed is modified[30] by the multiplicative factor $[2ab/(a^2 + b^2)]$. Imposition of the free–free boundary condition leads to a series of harmonic modes whose frequencies f_n^T are given by

$$f_n^T = nc_T/2L; \; n = 1,2,3,\ldots . \tag{5}$$

The previous two equations can similarly be combined to provide an expression for the shear modulus of the bar in terms of its mass density, length, and the ratio of the frequency of the nth mode to its mode number n:

$$G = 4\rho L^2 (f_n^T/n)^2. \tag{6}$$

At this point, if one can measure the frequencies of these modes along with the mass and the physical dimensions of the sample, sufficient information exists to completely specify the elastic properties of the isotropic, homogeneous material.

B. Flexural mode

The measurement of the flexural mode is not necessary, but it does provide another estimate for the Young's modulus, and its fundamental frequency is generally an order-of-magnitude lower than the longitudinal modes. It can also be strongly excited by this transduction scheme, so it is easy to observe and can remove ambiguity when torsional and longitudinal modes occur at nearly the same frequency.

Unlike the torsional and longitudinal modes, the flexural waves of the bar obey a fourth-order differential equa-

tion and the flexural wave phase speed c_F is dispersive. At frequencies low enough that the effects of rotary inertia and shear deformations associated with the flexure[26] can be neglected, the flexural wave phase speed c_F can be expressed as

$$c_F = \sqrt{2\pi f \kappa c_L}, \tag{7}$$

where κ is the radius of gyration[38] given by

$$\kappa^2 = \left(\frac{1}{S}\right)\int z^2\, dS, \tag{8}$$

S is the cross-sectional area of the rod, and z is the distance of an element above the neutral axis in the direction of flexure. For a rod of circular cross section, $\kappa = d/4$. It is worth noting that for a rod of elliptical cross section, the two possible polarizations of this mode are no longer degenerate and the phase speed will also depend upon whether the flexure is along the major or minor axes of the ellipse, in which case the appropriate axis length is substituted for d.

It is also important to recognize that the phase speed does not increase without limit as the frequency becomes infinite. At high frequencies, the assumption that $\lambda \gg d$ fails, and the contributions of the rotary inertia and shear are no longer negligible. For the case of $\lambda < d$, the phase speed for flexural waves asymptotically approaches the Rayleigh wave speed[28,30] which varies from 87.4% of c_T to 95.5% of c_T for values of Poisson's ratio between zero and one-half. All of the measurements made using this technique reported here are at frequencies that are well below this limit and Eq. (7) is entirely adequate.

The application of free–free boundary conditions in this case leads to a series of modes that are not harmonic. The frequency of the nth mode f_n^F, is given by

$$f_n^F = \pi n^2 c_L \kappa / 8L^2; \quad n = 3.0112, 4.9994, 7, 9, 11\ldots . \tag{9}$$

Combining the previous two equations and assuming a rod of circular cross section, the Young's modulus of the rod can be expressed as

$$E = \frac{1024}{\pi^2}\frac{\rho L^4}{d^2}\left(\frac{f_n^F}{n^2}\right)^2, \tag{10}$$

where, as in Eq. (9), n takes the values 3.0112, 4.9994, 7, 9, 11... . For elliptical samples the reader is again cautioned to substitute the length of appropriate axis for d depending upon the polarization of the specific mode excited.

II. APPARATUS

As demonstrated in the previous section, the ability to measure the resonance frequencies of a free–free bar and the knowledge of the mass and dimensions of the bar are sufficient to completely determine the elastic constants of the material of the bar if it is homogeneous, isotropic, and of circular or elliptical cross section. The following subsection will describe an electrodynamic transduction scheme that can be used to detect and excite all three modes using a single coil of wire attached to each end of the bar placed, in various orientations, within the field of a magnet. An overall view of the apparatus used to obtain the data in this article, consisting of radar-style magnets and chemistry ring stand supports, is shown in Fig. 1. A complete commercial version of the measurement apparatus is now available.[39] The cylindrical samples are typically 25–35 cm long and 1.2 cm in diameter although there is no reason why larger or smaller samples should present any difficulties as long as $L \gg d$.

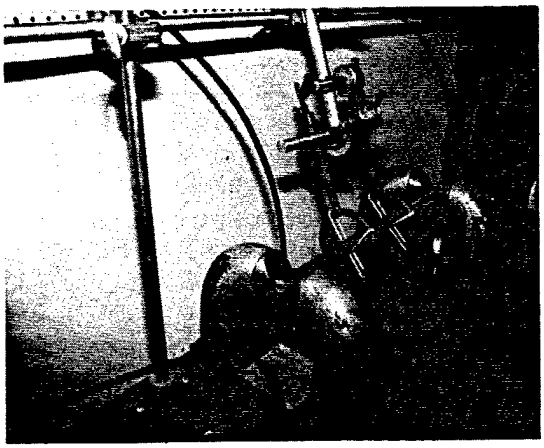


FIG. 1. Photograph of the entire physical apparatus shown with a bar sample that is approximately 30 cm long and 1.2 cm in diameter.

A. Transducer coils and magnets

The transducers are formed out of two individual coils of "magnet" wire that are glued to the two ends of the bar and placed in the magnetic field of two magnets. The coils that were used to produce the data discussed in the following section were fabricated from No. 32 gauge copper wire (nominal diameter 0.008 in. = 0.2 mm), about 1 m long, that were formed into coils of approximately ten turns each. After winding the coil, an additional length ($\approx$12 cm) of the same wire is wound helically around the bundle of turns to keep the coil together while it is being deformed to fit on the bar and then glued. That wire is provided only for packaging and has no electrical function. These coil parameters are not particularly important. It is advantageous to use as small a gauge wire as possible to reduce the mass loading of the transducer on the bar modes. The wire for the transmitter coil must have sufficient current carrying capacity (typically 50–200 mA_{rms} in most measurements).

The coils are positioned in the magnetic gap formed by two pole pieces that have rectangular faces that are 3 cm long and 1 cm wide. The poles are typically separated by approximately 2 cm, although an adjustable magnet structure is convenient since it allows the gap to be varied for larger or smaller samples. When configured with a 2 cm gap, the maximum magnetic induction **B** at the center of the gap is typically 2.0 $\pm$ 0.2 kG (0.20 $\pm$ 0.02 T). Since nominally identical transducers are used for both excitation and detection of the modes, an increase in the magnitude of the magnetic field improves the magnitude of the signal in proportion to the square of the field.

When the coils are positioned in the gap, they can be modeled as a resistor and inductor in series that, together, are in parallel with a capacitance. For the measurements reported in Sec. III, the equivalent coil resistance was 1.6 Ω, and the coil self-inductance was 13.7 μH. The value of the equivalent parallel capacitance was about 250 pF. This sets the coil self-resonant frequency at several megahertz, well above the highest frequencies of interest. The transmitter

coil generates an oscillating force due to the cross product of the current i and the magnetic field, $\mathbf{F} = \mathbf{B} \times N\mathbf{i}$, where N is the number of turns. The impedance seen by the current generator is that of the equivalent series resistor and inductor. The receiver generates an emf due to the changing flux Φ, within the area of the coil, emf $= d\Phi/dt$. The output impedance of the equivalent generator is again the equivalent coil impedance, which means that the output should be transformer coupled or a transimpedance (current) preamplifier should be used before presenting the signal to standard measurement electronics (spectrum analyzers, voltmeters, oscilloscopes, etc.) that have a typical input electrical impedance of 1 MΩ. The electronic instrumentation used in our measurements will be described in Sec. II C.

B. Selective mode excitation and detection

The key to the simplicity and economy of this acoustical method of modulus measurement is the ability to selectively and strongly excite the three resonant modes independently using the same inexpensive (hence disposable) transducer. Figure 2 illustrates the arrangement of the magnet and coil required to transduce the torsional mode. The bar is placed with its axis at the center of the pole-piece faces that are aligned along the bar axis. The normal to the plane of coil is perpendicular to the magnetic field lines. This makes the net force on the coil zero while producing the maximum moment to efficiently generate the torsional waves. The bar should be arranged so that the straight section of the coil, which goes over the end of the bar, is extending beyond the end of the pole pieces so that the minimum amount of longitudinal force is generated. For transduction of the torsional mode, both magnet–coil transducers should, of course, be oriented in the aforementioned fashion. The receiver generates its emf because the plane of the coil is "rocking" back and forth in the region of strong, fairly uniform magnetic field. This change in angle modulates the flux in the coil since the projection of the area, which the coil presents in the direction of the magnetic field, is varying harmonically.

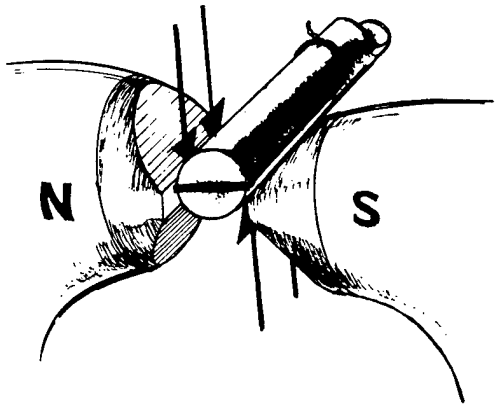


FIG. 2. Orientation of the coil and magnet pole pieces for the transduction of the torsional mode. The arrows indicate the magnitude and direction of the electromagnetic forces on the two dominant sections of the coil for a given phase of driver current. The forces produce the moment that excites the torsional oscillations.

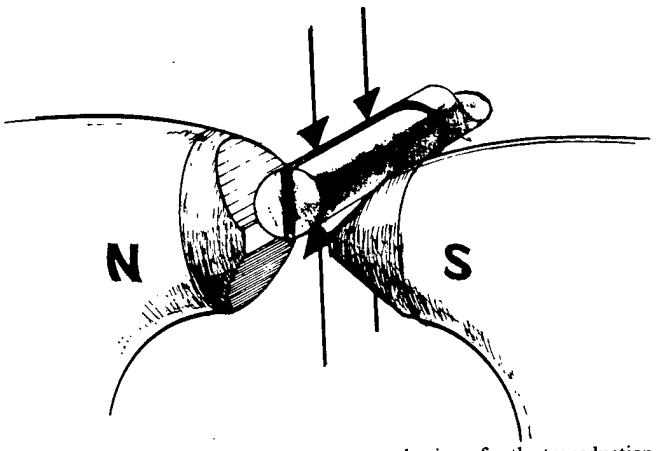


FIG. 3. Orientation of the coil and magnet pole pieces for the transduction of the flexural mode. The arrows indicate the magnitude and direction of the electromagnetic forces on the two dominant sections of the coil for a given phase of driver current. The lower section of the coil is in the stronger region of the magnetic field and hence exerts a greater force on the bar than the upper section of the coil, as indicated by the shorter arrows above the bar. This leads to a net force that excites the flexural mode.

The flexural mode can be observed by rotating the bar by 90° and translating it up or down by the diameter of the bar as shown in Fig. 3. This places one of the long coil sections near the top of the rectangular pole faces while the other section is well above or below the pole face and hence in a position of weaker magnetic field. The forces generated by these two sections of coil are then in opposition but, because one of the sections has a force which is of greater magnitude, the net force causes the bar to flex. The receiver coil is then raised and lowered through the field gradient changing the flux through the coil and generating the emf.

The longitudinal mode can be excited and detected by rotating the pole pieces so that the long edge of each pole piece is now vertical. The separation between the two magnet structures at either end of the bar is increased so that the strong region of magnetic field is primarily felt by the short section of coil that crosses the end of the bar along its diameter. This configuration is shown in Fig. 4. The currents in

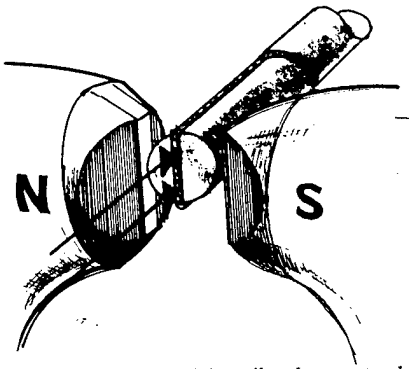


FIG. 4. Orientation of the coil and magnet pole pieces for the transduction of the longitudinal (extensional) mode. Note that the pole pieces have been rotated by 90° to concentrate the magnetic field just beyond the end of the bar. The arrows indicate the magnitude and direction of the electromagnetic forces on the dominant section of the coil for a given phase of driver current. Since this section of the coil at the end is shorter than the other sections used to drive the torsional and flexural modes, it is useful to reduce the amount of stray magnetic field that links the other portions of the coil if the longitudinal mode spectrum is not to be "polluted" by the presence of the torsional modes.

that short section generate longitudinal forces on the end of the bar that excite the longitudinal wave mode. Similarly, at the receiver, the coil is moved in and out of the strong field region by the wave-induced motion and thereby generates the emf.

Since the torsional and flexural modes use the longer section of the coil, they are the most strongly excited and (by reciprocity[40]) detected. The longitudinal mode is less responsive but can usually be detected with little difficulty. It is generally wise to measure the torsional and flexural modes first, since there are usually some torsional modes observed when longitudinal modes are being excited. If this is the case, there will be no confusion if the torsional frequencies have already been cataloged. In particularly difficult cases, the expected frequencies of the longitudinal modes can be calculated based on the Young's modulus derived from the flexural modes since those modes are always easy to observe and occur at several frequencies that are lower than even the lowest torsional mode. This procedure can help in the positive identification of the longitudinal modes when there is any ambiguity.

C. Instrumentation

An additional advantage of this technique is that it can be performed with common laboratory instrumentation. Figure 5 provides a block diagram of a typical measurement setup. As previously mentioned, the only unusual feature is the choice of a transimpedance configuration for the preamplifier. Since the output impedance of the transducer coils are typically only a few ohms, a current-to-voltage converter is dictated for low noise performance. The alternative would be a stepup signal transformer such as the PAR 181, Ref. 41, but the availability of good quality, low current drain, operational amplifiers for about a dollar[42] (e.g. LF 441,LM308, etc.) make a battery operated preamplifier with a feedback resistor in the 1- to 10-kΩ range more attractive. Since the bar can vibrate at low frequency due to the soft suspension, a high-pass filter with a 10-Hz rolloff frequency will remove annoying jitter on the oscilloscope due to the experimenter jostling the work bench by turning knobs on the instruments. On the transmitter side, any power amplifier capable of delivering a watt to a 1-Ω load, such as the HP 467, or any good audio transformer is entirely adequate to couple signal sources of typically 50- or 600-Ω output impedance to the low-impedance transducer coil.

Beyond this point, the choice of instrumentation will be dictated by the equipment on hand. It is particularly convenient to have a swept-sine spectrum analyzer with a CRT display of amplitude (linear or log) versus frequency to setup the system, adjust the coils and pole pieces, and identify the various resonant modes. Due to the large signal-to-noise ratio, excellent results with precision in excess of 0.1% can be obtained using only an ac voltmeter, a signal generator, and a frequency counter. A reciprocal counter or an ordinary counter used in the period mode is preferable since the frequencies of the flexural modes are typically 100–1000 Hz. We have found the HP 3580A spectrum analyzer (5 Hz–50 kHz) particularly useful both in our research and teaching laboratories. Still better is the HP 3562A dynamic signal analyzer (0–100 kHz) that will automatically create a fit to as many as ten resonances, providing their frequency and quality factor in a pole-zero table and displaying the fit function overlayed on the acquired spectrum for comparison. [Figures 8–10 were generated by the HP 3562A directly using a digital pen plotter (HP 7470A).]

D. Phase-locked resonance tracking

As mentioned in the Introduction, one of the most important material measurements, particularly for hydrophone design, is the change in modulus with temperature. In several hydrophone designs,[3–13] the elastic properties of the materials dictate hydrophone sensitivity and some auxiliary element (optical fiber, PVDF, piezoresistive strain gauges etc.) provides the actual transduction mechanism. The technique described here is particularly well suited to such temperature dependance measurements. Figure 6 is a block diagram of the instrumentation we use to produce a

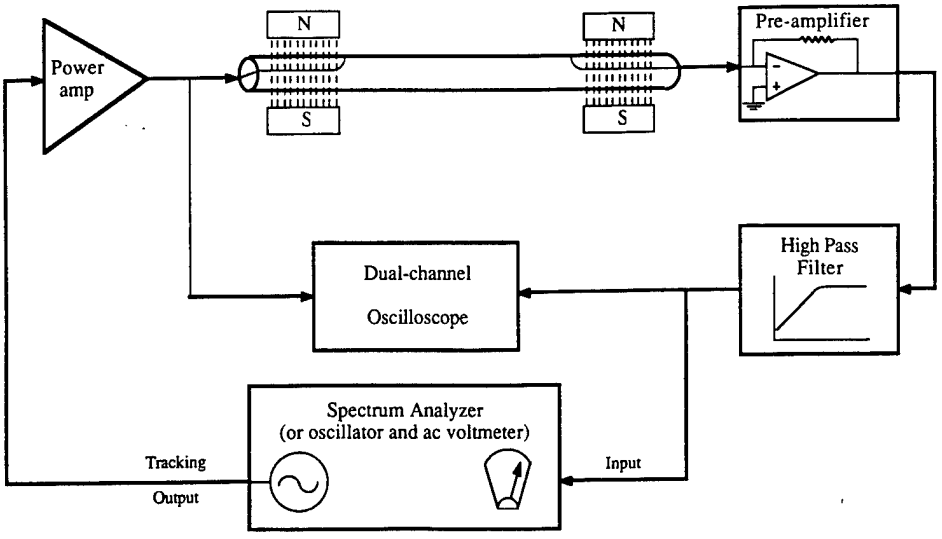


FIG. 5. Block diagram of the typical instrumentation used to observe the modes of the bar. Transformers can be substituted for the two amplifiers if desired. A swept-sine spectrum analyzer with a CRT display, such as an HP 3580A or PH 3562A, is recommended to optimize the coil and magnet setup and to simplify the identification of the various modes.

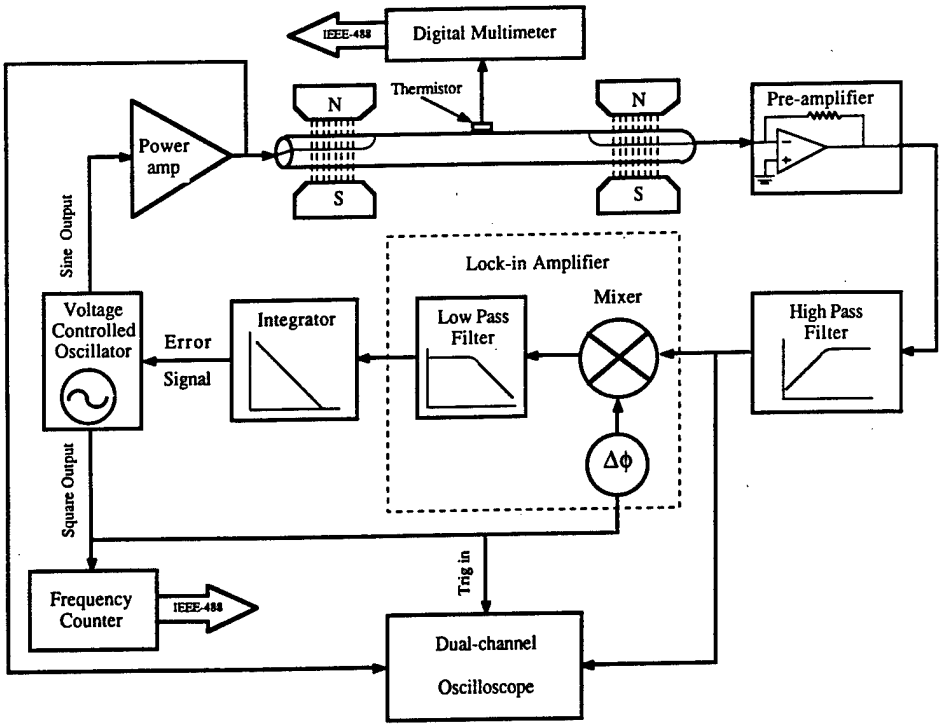


FIG. 6. Block diagram of the phase-lock-loop used to automatically track the change in the resonance frequency of a particular mode as a function of temperature. Only the digital multimeter (thermometer) and the frequency counter have to be "bus compatible" for computerized data acquisition, display, and analysis since the resonance tracking is performed entirely by the analog electronics.

phase-locked-loop[42] to facilitate the automation of such measurements.

Our usual procedure is to identify the modes of the bar at room temperature, as discussed previously, and then select a given mode for automatic tracking with temperature. This is done by turning the voltage controlled oscillator (VCO) manually to resonance with the integrator shorted and the error signal feedback path open so that there is no error signal presented to the voltage control (feedback) point. The phase shifter is then adjusted so that there is zero output from the quadrature signal channel of the lock-in amplifier. The integrator can then be opened and the control loop completed. If the resonance "runs away" the signal (either transmitter or receiver) needs to be inverted. This is most easily accomplished by shifting the phase an additional $\pm 180°$ if a lock-in analyzer is being used as the mixer/phase-shifter/low-pass filter. It may also be necessary to adjust the filter time constant and amplifier gain if the signal oscillates. Once the control loop is locked and stable, the temperature may be varied.

One may elect to eliminate the error signal integrator and place the output of the low-pass filter that follows the mixer directly into the VCO control. Without the integrator, there must always be some phase error present to provide the required VCO offset from its free-running frequency. In cases with high Q resonances, the loop gain can be made so large and the time constant so long (1–10 s) that the phase error is entirely negligible.

We generally use a small thermistor[43] with a mass of only 25 mg to monitor the temperature of the sample since it introduces minimal perturbation (none if mounted at the center of the bar and odd modes are monitored), and it can be read directly in degrees Celsius or Fahrenheit to a precision of millidegrees by our HP 3456A digital multimeter. If the multimeter and frequency counter are "bus compatible,"

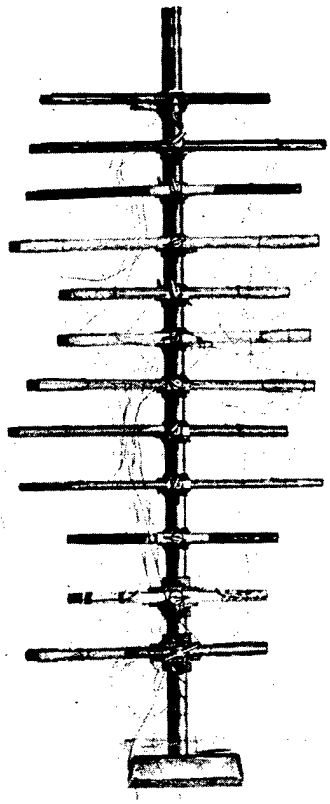


FIG. 7. Photograph of one of our "sample trees." The use of standard chemistry ring stands and clamps for sample storage is particularly convenient since the transducer leads do not tend to tangle or break and the labels are easily readable.

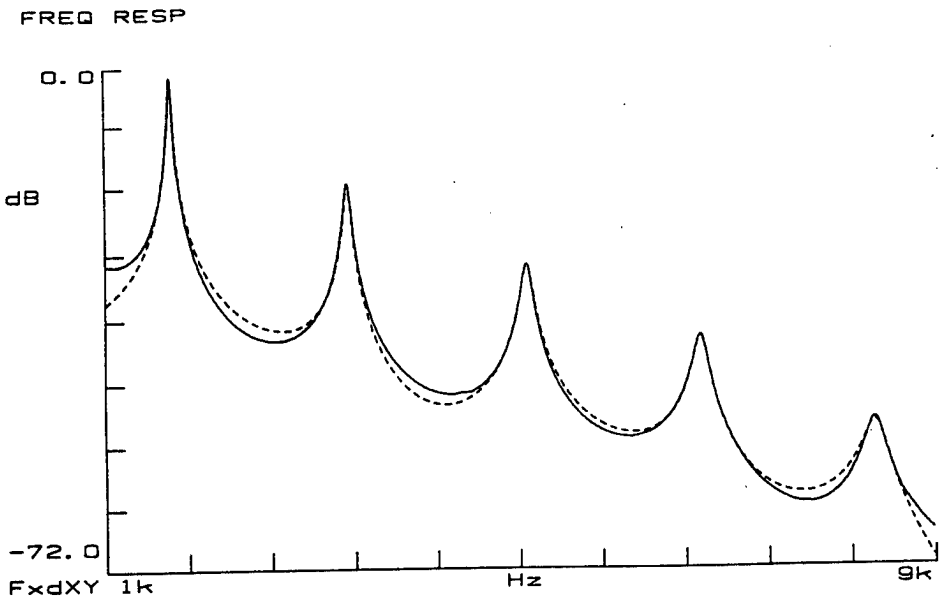


FIG. 8. Spectrum of the first five torsional modes of the sample bar between 1 and 9 kHz. The solid line is the measurement and the dashed line is a fit to the data created by the HP 3562A Dynamic signal analyzer. Notice that the dynamic range is in excess of 70 dB, there is no noise, the modes are harmonically related, and neither the longitudinal nor flexural modes are apparent.

the entire set of temperature measurement data can be acquired, displayed, and analyzed by computer, since the actual "resonance tracking" is performed entirely by the analog electronics.

III. DATA ANALYSIS

In this section the complete analysis of the elastic constants of a single sample will be described in the hope of making the actual use of this technique more transparent. The choice of this particular composite epoxy sample was dictated only by convenience and by the fact that its cross section had a small, but non-negligible amount of ellipticity. This allows discussion of the required corrections for an "imperfect" sample. Results are considerably better for metallic samples since their resonances have higher quality factors and their densities are higher than plastics. The higher density reduces the effects of the added transducer mass (typically 1 g).

A. Sample preparation and geometry

The sample that will be used as an example here is a high tensile strength epoxy[44] which has been filled with 1/32 in. glass fibers to 20% by weight during mixing. It was cast in a plastic tube that had been coated prior to casting with a mold release agent.[45] The mold was cured at room temperature. The sample was pushed out of the tube and the ends were sawed straight using a bandsaw. Its length was measured to be 33.60 ± 0.02 cm and it was found to have a mass of 55.723 ± 0.001 g. As can be seen from the dozen samples, shown in Fig. 7, the actual sample size is not critical. The maximum and minimum diameters were measured at five points along the sample and were found to be 13.17 ± 0.02 mm and 12.77 ± 0.02 mm, respectively, yielding a mean diameter of 12.97 ± 0.20 mm and an average density of 1.256 ± 0.004 g/cc ($\pm 0.3\%$). It is not ususual to have this degree of ellipticity in a sample which is cast in a plastic tube. Glass tubes are more circular but are more dangerous to

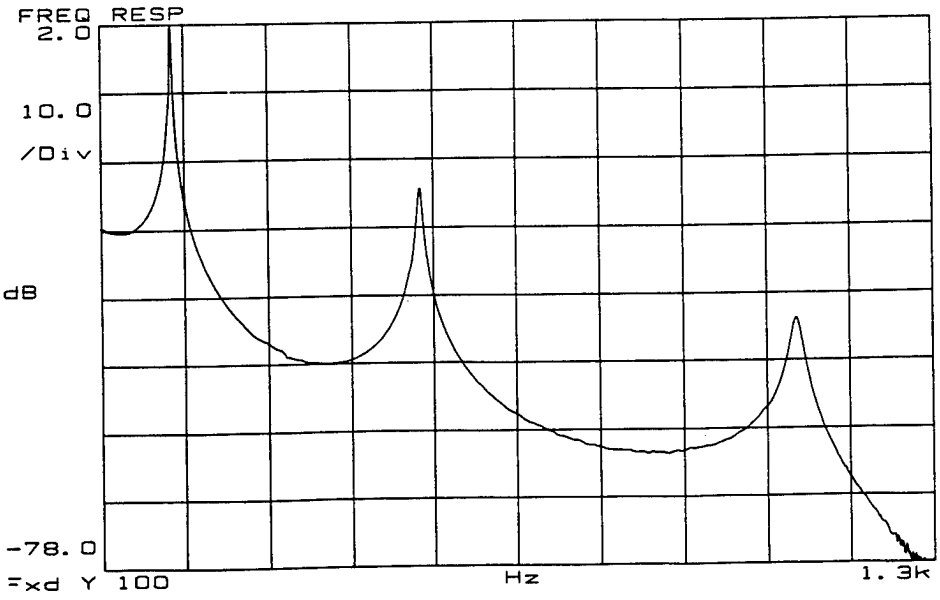


FIG. 9. Spectrum of the first three flexural modes of the sample bar between 100 and 1300 Hz. Notice that these three modes are not harmonically related. The dynamic range is 80 dB, and the first appearance of noise is at least 75 dB below the amplitude of the first mode.

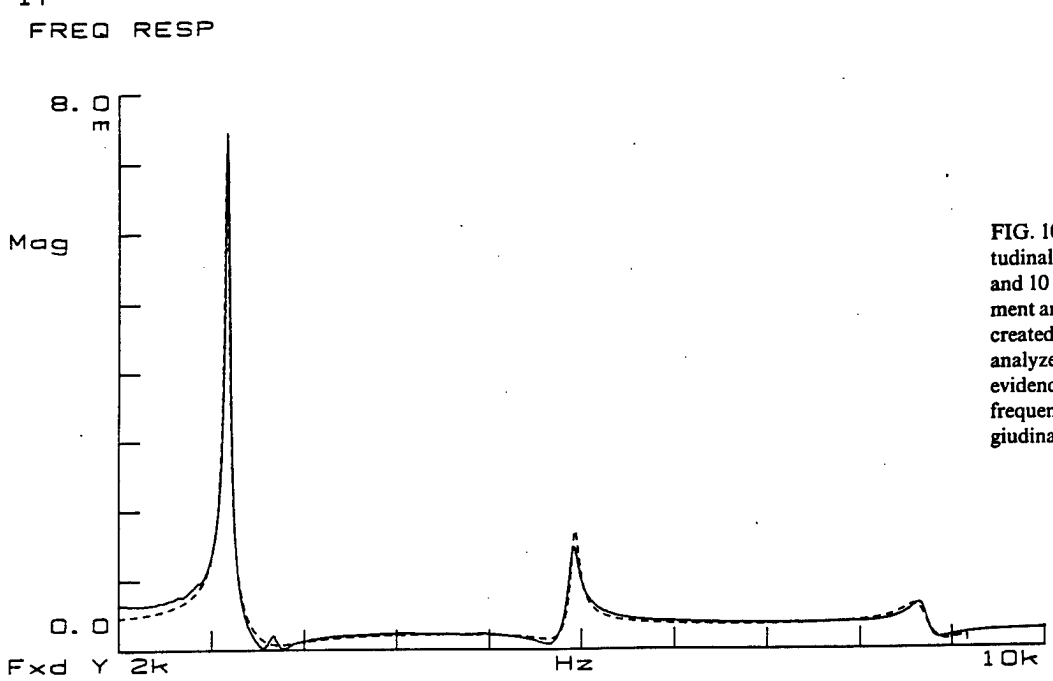


FIG. 10. Spectrum of the first three longitudinal modes of the sample bar between 2 and 10 kHz. The solid line is the measurement and the dashed line is a fit to the data created by the HP 3562A Dynamic signal analyzer. Notice that there is some slight evidence of the second torsional mode at a frequency just above the fundamental longiudinal mode.

remove since they are seldom removed in one piece.

Transducer coils were then attached with Duco™ cement as described previously. For the case of a sample with an elliptical cross section, it is preferable to orient the coils along either the major or minor diameter so that the polarization of the flexural mode is known and the analysis can be more specific. For this sample, the major and minor diameters differ by 3% corresponding to a difference of 6% in the frequencies of the two polarizations. After attachment, the bar should be reweighed to determine the added mass of the coils and their adhesive. Although this measurement can be made to very high precision, the fact that the amount of lead wire which participates in the vibration is not known means that the added mass is only approximate. For this sample I have taken the added mass to be 1.5 ± 0.1 g. This uncertainty does not significantly degrade the accuracy of the results since it produces an error in the effective length of only about $\frac{1}{2}$%.

B. Modal analysis

Figures 8–10 show the spectrum of the gravest torsional, flexural, and longitudinal modes of the bar. The individual spectra are described in the figure captions and the frequencies of the modes are summarized in Table I. The standard deviation of the normalized frequencies are shown next to their average values followed by the relative uncertainty of the mean values expressed as a percent. The relative uncertainty in the mean is derived from the ratio of the standard deviation to the mean divided by the square root of the number of measurements minus one. One can see from those relative uncertainties that the technique is indeed precise and the assumed frequency ratio [Eqs. (2), (5), and (9)] are correct.

C. Calculated moduli

The shear modulus is derived directly from Eq. (6) using the measured density and the appropriate effective length [Eq. (A9)], which for this sample is 35.41 cm (± 0.4%). The result is $G = 1.77 \pm 0.02$ GPa, with a relative uncertainty of ± 1%. The major contribution to the uncertainty arises from the 0.6% relative uncertainty in f_n/n based on the assumed harmonicity of the lowest five modes. The correction to the torsional wave speed, and hence the resonance frequencies, due to the ellipticity of the sample, is less than four parts in ten thousand.

The Young's modulus derived from the measurement of the resonance frequencies of the first three longitudinal modes [Eq. (3)] is $E^L = 5.23 \pm 0.05$ GPa, with a relative

TABLE I. Summary of the frequencies of the three modes of vibration of the sample bar described in the text. The number following the average of the normalized frequency is the standard deviation. The number following the standard deviation in parentheses is the relative uncertainty in the mean value reported as a percent. It is equal to the ratio of the standard deviation to the mean divided by the square root of the number of entries in the table less one.

	Modal frequency summary	
Mode number [n]	Frequency (Hz) [f_n]	Normalized frequency (Hz) [f_n/n]
Torsional		
1	1644	1644
2	3341	1670
3	5080	1693
4	6750	1688
5	8446	1689
Average		$1677 \pm 20(\pm 0.6\%)$
Flexural		
$(3.0112)^2$	203.1	22.40
$(4.9994)^2$	560.4	22.42
$(7.0000)^2$	1104	22.53
Average		$22.45 \pm 0.07(\pm 0.2\%)$
Longitudinal		
1	2941	2941
2	5936	2968
3	8896	2965
Average		$2958 \pm 15(\pm 0.4\%)$

uncertainty of $\pm 0.9\%$, with again the dominant correction being due to the $\pm 0.4\%$ relative uncertainty in f_n/n based on the assumed harmonicity of the lowest three modes. The effective length used for this mode, as well as the flexural mode, is 34.51 cm $\pm 0.2\%$.

The flexural mode measurements produce a Young's modulus based on Eq. (10) of $E^F = 5.37 \pm 0.09$ GPa with a relative uncertainty of $\pm 1.7\%$. If one considers only the error introduced by the uncertainty in the effective length correction, the error would be about 1%, but the coils were not exactly aligned with the major diameter and the increased uncertainty is increased to reflect that fact. In any case, the determinations of the Young's modulus differ by about 2.7% and would correspond to a rotation of the coil axis by about 30° from the assumed orientation. Better agreement could, of course, be obtained if a sample of circular cross section was used, or if the orientation of the coils were determined more accurately and chosen to excite the mode polarized so that the minor diameter would have dominated the flexural resonance, since this is the lower energy polarization, hence the preferred mode.

ACKNOWLEDGMENTS

The author wishes to acknowledge Professor Isadore Rudnick of the UCLA Physics Department for teaching me this method while I was his graduate student, and David A. Brown for his careful reading of the manuscript and thoughtful comments on its organization. The author appreciates the assitance of Lt. B. L. Beaton, Lt. J. Serocki, and Lt. K. Wetterskog in the preparation of the samples (Fig. 7) and the construction of the apparatus (Fig. 1). This research was supported by the Naval Sea Systems Command and the Naval Postgraduate School Direct Funded Research Program. The writing of the paper was supported in part by Sperry Marine, a subsidiary of Newport News Shipbuilding.

APPENDIX: EFFECTIVE LENGTH CORRECTION

In the development presented in the body of this article, the added mass of the coils and their adhesive was neglected since its introduction at that point would have detracted from the development of the technique. Since the mass of both of these transducers, m, can be incorporated into the modulus equations [Eqs. (3), (6), and (10)] through the introduction of an effective length of the bar L_{eff}, the accurate value of the modulus can be determined if an expression for the effective length can be obtained. The results presented below were first obtained by Rayleigh[26] for the longitudinal (Sec. 155) and flexural (Sec. 185) cases, although not for the torsional case. A simple derivation of the length correction for all three cases is presented here in the hope that the potential users of this technique will be more confident in the application of these corrections and also be aware of the limitations inherent in the assumptions used to derive them.

All three corrections will be obtained by means of Rayleigh's method[26] that obtains the resonance frequency by equating the maximum kinetic and potential energies of the vibrating system. This result is most easily seen for the case of a lumped parameter, simple harmonic oscillator where the kinetic energy is $(1/2)mv_0^2$ and the potential energy is $(1/2)kx_0^2$. For simple harmonic motion $v_0 = \omega x_0$, and the well-known result for the frequency of oscillation is obtained:

$$\frac{1}{2}mv_0^2 = \frac{1}{2}m\omega^2 x_0^2 = \frac{1}{2}kx_0^2 \Rightarrow \omega^2 = \frac{k}{m} \quad . \tag{A1}$$

Since the perturbation is applied at the free ends of the bar in all three cases, only the kinetic energy of the system will be effected at lowest order, which is adaquate since $m/M \ll 1$. The square of the relative shift in the resonance frequency $(\delta f/f)^2$, due to the added mass, therefore depends only on the relative shift in the kinetic energy caused by the addition of the transducers:

$$\frac{\delta f}{f} = \frac{\delta\omega}{\omega} = \frac{-1}{2}\frac{\delta KE}{KE} \quad . \tag{A2}$$

The effective length is written in terms of the actual length of the bar before attachment of the coils L, and the ratio of the total mass added by both of the coils and their adhesive m, to the bare mass of the rod M. The total mass of the bar after the transducer coils have been attached is then $M_{\text{tot}} = m + M$. In all of the following calculations it is assumed that half of the total added mass is applied at a point at both ends of the bar. Although this is not strictly the case, since the length of either coil when mounted on the bar is about 5%–10% of a wavelength for the fundamental mode, the difference in the motion of the various parts of the coil is negligible. This is because the effective length correction $\delta L = L_{\text{eff}} - L$, is only a few percent of the total length and the neglect of the variability in the coil motion modifies that correction by only about 5%, making its effect on the order of a part per thousand of the length.

1. Longitudinal mode

The calculation for the effective length of the bar in the lowest frequency longitudinal mode is simplified if one exploits the symmetry of that mode. Since the center of the bar is at rest, the entire left half of the bar can be replaced by a rigid end at $x = 0$ as diagramed in Fig. A1. The displacement of the bar due to the acoustic oscillations can be written as

$$\xi(x,t) = \xi_0 \sin kx \sin \omega t, \tag{A3}$$

where ξ_0 is the displacement of the free end, $k = 2\pi/\lambda$ is the wavenumber, $\omega = 2\pi f_1^L$ is the frequency of the fundamental longitudinal mode, and $\omega/k = c_L$ is the phase speed of longitudinal waves. The kinetic energy of the unloaded bar in

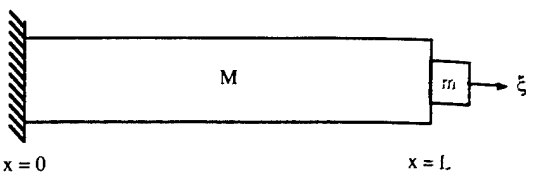


FIG. A1. Schematic diagram of a fixed-free, mass loaded bar of half the length of the original free–free bar. The mass of the unloaded bar is $M = (\pi/4)d^2L\rho$, and the added mass is m. In this figure L, M, and m refer to one-half of the corresponding parameters for the free–free case. The acoustic particle displacement of the end of the bar is indicated by ξ.

longitudinal vibration KE_0^L can be obtained by integration as shown below:

$$KE_0^L = \frac{1}{2}\rho S\omega^2\xi_0^2\int_0^L \sin^2 kx\,dx = \frac{1}{4}M\omega^2\xi_0^2, \quad \text{(A4)}$$

where temporarily we have taken m to be the added mass due to a single coil and the length L to be equal to half of the original bar length, since we are now concerned with only half of the bar. As before, S is the cross-sectional area of the bar and ρ is its mass density.

The kinetic energy of the coil placed at the end δKE is simply $(1/2)mv^2$ under the assumption that all of the mass is in fact concentrated at the end. The relative shift is simply $\delta KE/KE = 2m/M$. From Eq. (2) we have that the length of the bar and its resonance frequency are inversely proportional, hence,

$$\frac{\delta L}{L} = \frac{-\delta f}{f} = \frac{1}{2}\frac{\delta KE}{KE} = \frac{m}{M}. \quad \text{(A5)}$$

The effective length for the longitudinal mode is therefore given as

$$L_{\text{eff}}^L \cong L(1 + m/M). \quad \text{(A6)}$$

The above result applies to the entire bar since the length would be twice that in Fig. A1 as would be the added mass, since there are now two transducer coils. This simple result can be interpreted as saying that the frequency of the mass-loaded bar, for small m/M is the same as an unloaded bar of equal cross section that has been lengthened, so that its mass is now $M_{\text{tot}} = M + m$.

2. Torsional mode

The same style of argument applies to the torsional case except that the location of the added mass with respect to the axis of the bar is important, since the torsional oscillations occur due to the restoring torque and the moment of inertia of the bar's mass. For this case the kinetic energy is due to rotation so that $KE = (1/2)I(\omega\theta_0)^2$, where I is the moment of inertia and $\omega\theta_0$ is the maximum angular velocity. The moment of inertia[38] of a disk of mass m and diameter d is $(1/8)md^2$. The unloaded kinetic energy of rotation, KE_0^T, for the bar undergoing torsional oscillations can be obtained by integration as in the longitudinal case, since the angular displacement at each point along the bar is given by Eq. (A3) if θ_0 is substituted for ξ_0:

$$KE_0^T = \frac{1}{2}\rho S\omega^2\theta_0^2\frac{d^2}{8}\int_0^L \sin^2 kx\,dx = \frac{M\omega^2\theta_0^2 d^2}{32}. \quad \text{(A7)}$$

For an ellipse, d^2 would be replaced by the product of the major and minor axes.

Since the coil mass is located primarily at the disk radius, its moment of inertia is $(1/4)md^2$, and the change in kinetic energy would be $(1/8)m(d\omega\,\theta_0)^2$. (Again, the fact that a small portion of the coil is placed across the end of the bar, rather than on its surface, is neglected since this would be a small correction to a small correction.) The relative change in the kinetic energy for this case would therefore be $\delta KE/KE = 4m/M$. For a sufficiently elliptical sample it could be important to arrange that the plane of the coil coincide with either the major or minor diameter of the ellipse, since this would effect the length correction because the d^2 terms in the two kinetic energy expressions would not necessarily cancel.

The frequency of the torsional mode and its length are inversely related [Eq. (5)], the relative frequency shift and hence the relative length change is given by

$$\frac{\delta L}{L} = \frac{-\delta f}{f} = \frac{1}{2}\frac{\delta KE}{KE} = \frac{2m}{M}, \quad \text{(A8)}$$

and the effective length is,

$$L_{\text{eff}}^T \cong L(1 + 2m/M). \quad \text{(A9)}$$

This suggests that the mass of the coil is twice as effective in reducing the frequency of the torsional mode as the same added mass would have been in reducing the frequency of the longitudinal mode.

3. Flexural mode

The unloaded transverse displacement function $y(x,t)$ of the bar for the gravest flexural mode is also known exactly[26,35]

$$y(x,t) = y_0[\cosh kx + \cos kx - B(\sinh kx + \sin kx)]\sin \omega t, \quad \text{(A10)}$$

where, for the lowest frequency mode $k = 4.73$ for a bar of unit length, $B = -0.985$, and y_0 is half the amplitude of either end of the bar. Based on this function, one could calculate the kinetic energy of transverse vibration of the bar from the well-known[32] expression:

$$KE_0^F = \frac{1}{2}\pi a^2\rho\int_0^L \left(\frac{\partial y}{\partial t}\right)^2 dx. \quad \text{(A11)}$$

To simplify the calculation, a polynomial approximation to the exact solution $y^*(x,t)$ will be used. In contrast to the exact expression Eq. (A10) for $y(x,t)$, which is written with the origin of the coordinate system at the left end of the bar, the origin of the coordinate system for $y^*(x,t)$ will be translated to the center of the bar to exploit the symmetry of the fundamental mode:

$$y^*(x,t) = (A/L^4)(cL^4 - 3L^2x^2 + 2x^4)\sin \omega t. \quad \text{(A12)}$$

For the approximate expression, the limits of integration for the kinetic energy integral [Eq. (A11)] must then become $\pm L/2$.

This functional form has the advantage that it automatically satisfies the boundary conditions on the bar since $(\partial y^*/\partial x)_{x=0} = 0$, and $(\partial^2 y^*/\partial x^2)_{x=\pm L/2} = 0$. The value of the constant $c = 0.2363$, is set by forcing the two functions to have equal values at their centers and their end points. Since the added mass is applied at the ends of the bar, and the kinetic energy is to be recalculated to obtain the frequency shift, this seems to be the reasonable choice. Here, A is an abitrary constant that determines the amplitude of the motion and hence is not significant in the linear theory. For this case, its value is 2.57 times greater than the transverse displacement of either end points.

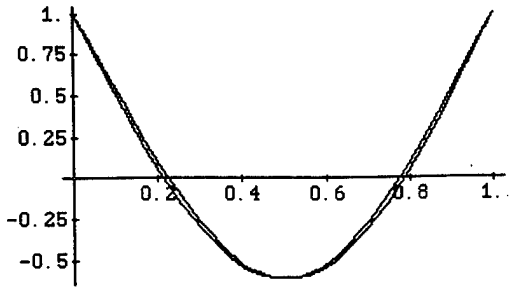


FIG. A2. Plot of the exact solution (A10) for the transverse displacement of a free–free bar executing flexural oscillations and the polynomial approximation (A12) used to calculate the kinetic energy and its mass-loaded correction. The nodes of the exact solution are located 0.224 L from the ends of the bar and the nodes of the approximate solution are at 0.211 L. The length of the bar has been normalized to unity as has the transverse displacement of the free ends of the bar.

The polynomial approximation Eq. (A12) and the exact solution Eq. (A11) are plotted together in Fig. A2, where it is apparent that their shapes are similar. For the exact solution, the nodes are located 0.224 L from either end. The nodes of the approximate solution are 0.211 L from the ends. Substitution of the polynomial approximation into the integral for the kinetic energy [Eq. (A11)] yields $KE_0^F = 1.85\times10^{-2}MA^2\omega^2$. The maximum velocity at the end of the bar is simply $\omega y^*(\pm L/2)$, so the change in kinetic energy $\delta KE = (1/2)mv^2 = 7.56\times10^{-2}mA^2\omega^2$. The relative change in kinetic energy $\delta KE/KE$ is calculated to be 4.1 m/M. Since, for the flexural mode, the frequency is proportional to the square root of the energy ratio and inversely proportional to the length of the bar squared [Eq. (9)], the required effective length correction δL, is given by

$$\frac{\lambda L}{L} = \frac{-l}{2}\frac{\delta\omega}{\omega} = \frac{-1}{2}\left(\frac{-1}{2}\frac{\delta KE}{KE}\right) \cong \frac{m}{M}. \qquad \text{(A13)}$$

It was reasonable to let $4.1 \approx 4.0$, since the Rayleigh method always generates an overestimate and the exact transverse displacement function would also have shifted slightly due to the added mass that would have changed the condition that $(\partial^2 y/\partial x^2)_{x=\pm L/2} = 0$. This suggests that the effective length correction for the flexural mode is identical to that for the longitudinal mode.

$$L_{\text{eff}}^F \cong L(1 + m/M). \qquad \text{(A14)}$$

The above result agrees with that obtained by Rayleigh[26] in Sec. 186 of *The Theory of Sound* where he makes the following statement after a more rigorous calculation for the flexural mode: "If the load be at the end, its effect is the same as lengthening of the bar in the ratio $M:M + dM$."

[1]L. M. Lyamshev and Yu. Yu. Smirnov, "Fiber-optic acoustic sensors (review)," Sov. Phys. Acoust. **29**(3), 169–180 (1983).

[2]V. I. Busurin, A. S. Semenov, and N. P. Udalov, "Optical and fiber-optic sensors (review)," Sov. J. Quantum Electron. **15**(5), 595–621 (1985).

[3]M. Layton and E. Moore, "Extended fiber optic hydrophones for at-sea testing," Litton Guidance and Control Systems Division, June, 1988, Document No. 19078A.

[4]G. F. McDearmon, "Analysis of a push-pull fiber-optic hydrophone," J. Lightwave Tech. **LT-5** (5), 647–652 (1987).

[5]D. L. Gardner and S. L. Garrett, "High sensitivity fiber-optic compact directional hydrophone (U)," U. S. Navy J. Underwater Acoust. (JUA USN) **38**(1), 1–22 (1988).

[6]D. A. Brown, T. Hofler, and S. L. Garrett, "High sensitivity, fiber-optic, flexural disk hydrophone with reduced acceleration response," Fiber Integrated Opt. **8**(3), 169–191 (1989).

[7]D. A. Danielson and S. L. Garrett, "Fiber-optic ellipsoidal flextensional hydrophones," J. Lightwave Tech. **7**(12), 1995–2002 (1989).

[8]G. E. MacDonald, "Fiber optic gradient hydrophone construction and calibration for sea trials," Masters thesis, Naval Postgraduate School, March, 1985, DTIC Rep. No. AD A 156 469.

[9]N. Lagakos, P. Ehrenfeuchter, T. R. Hickman, A. Tveten, and J. A. Bucaro, "Planar flexible fiber-optic interferometric acoustic sensor," Opt. Lett. **13**(9), 788–790 (1988).

[10]D. W. Gerdt and N. T. Richardson, "Fiber optic hydrophones based on variable ratio couplers," in *Fiber and Laser Sensors VI*, Proc. Soc. Photo-Optical Inst. Eng. (SPIE) **1169**, 249–260 (1989).

[11]D. W. Gerdt and L. H. Gilligan, "Variable coupler fiber optic sensor," Proc. Soc. Photo-Optical Inst. Eng. (SPIE) **835**, 25–31 (1987).

[12]T. Sullivan and J. Powers, "Piezoelectric polymer flexural disk hydrophone," J. Acoust. Soc. Am. **63**, 1396–1401 (1978).

[13]D. Ricketts, "Model for a piezoelectric polymer flexural plate hydrophone," J. Acoust. Soc. Am. **70**, 929–935 (1981).

[14]R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison–Wesley, Reading, 1964), Chap. 38.

[15]L. H. VanVlack, *Materials Science for Engineers* (Addison–Wesley, Reading, 1970), p. 195.

[16]N. Lagakos, A. Dandridge, and J. A. Bucaro, "Optimizing fiber coatings for interferometric acoustic sensing (U)," U. S. Navy J. Underwater Acoust. (JUA USN) **37**(3), 233–254 (1987).

[17]N. Lagakos, J. Jarzynski, J. H. Cole, and J. A. Bucaro, "Frequency and temperature dependence of elastic modulii of polymers," J. Appl. Phys. **59**(12), 4017–4031 (1986).

[18]A. Barone and A. Giacomini, "Experiments on some electrodynamic ultrasonic vibrators," Acoustica **4**, 182–184 (1954).

[19]R. W. Leonard, "Attenuation of torsional waves in teflon," J. Acoust. Soc. Am. **40**, 160–162 (1966).

[20]E. R. Fitzgerald, "Simple method for observing audiofrequency resonances and sound beams in crystals," J. Acoust. Soc. Am. **36**, 2086–2089 (1964), and references therein.

[21]R. W. Leonard, "Comment on the existance of the Fitzgerald effect," J. Acoust. Soc. Am. **38**(4), 1–2 (1965).

[22]I. Rudnick, "Acoustics goofs or irreproducible effects in acoustics," J. Acoust. Soc. Am. (Suppl. 1) **83**, S38 (1988).

[23]S. L. Garrett, "Modes of a bar," Am. J. Phys. (submitted).

[24]M. Barmatz, H. J. Learny, and H. S. Chen, "A method for the determination of Young's modulus and internal friction in metallic glasses," Rev. Sci. Instrum. **42**, 885–887 (1971).

[25]M. O'Donnell, E. T. Jaynes, and J. G. Miller, "Kramers–Kronig relationship between ultrasonic attenuation and phase velocity," J. Acoust. Soc. Am. **69**, 696–701 (1981).

[26]J. W. Strutt (Lord Rayleigh), *The Theory of Sound* (Dover, New York, 1948), Vol. I, 2nd ed., Chap. VII and VIII.

[27]A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity* (Dover, New York, 1944), 4th ed., Chap. XX.

[28]R. B. Lindsey, *Mechanical Radiation* (McGraw-Hill, New York 1960), Chap. 7.

[29]H. Kolsky, *Stress Waves in Solids* (Dover, New York, 1963), Chap. III.

[30]L. D. Landau and E. M.Lifshitz, *Theory of Elasticity* (Pergamon, New York, 1970), Sec. 25.

[31]W. T. Thomson, *Theory of Vibration with Applications* (Prentice-Hall, Englewood Cliffs, NJ, 1981), 2nd ed., Chap. 7.

[32]P. M. Morse, *Vibration and Sound* (American Institute of Physics, New York, 1976) 2nd ed., Chap. IV.

[33]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Chap. 5.

[34]K. U. Ingard, *Fundamentals of Waves and Oscillations* (Cambridge U. P., Cambridge, England, 1988), Chap. 20.

[35]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982), 3rd ed., Chap. 3.

[36]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981).

[37]S. Temkin, *Elements of Acoustics* (Wiley, New York, 1981).

[38]R. J. Roark and W. C. Young, *Formulas for Stress and Strain* (McGraw-Hill, New York, 1975), 5th ed., Chap. 5.

[39]Mechanical Measurement Devices, P. O. Box 8716, Monterey, CA 93943.

[40]J. W. Strutt (Lord Rayleigh), "Some general theorems relating to vibra-

tions," Proc. Math. Soc. (London) 4, 357–368 (1873); *Collected Works* (Dover, New York, 1964), Vol. I, 170–181.
[41]EG and G Princeton Applied Research, P. O. Box 2565, Princeton, NJ 08543.
[42]P. Horowitz and W. Hill, *The Art of Electronics* (Cambridge U. P., Cambridge, England, 1989), 2nd ed.
[43]Fenwal Electronics, Model UUA35J84, 63 Fountain Street, Framingham, MA 01701.
[44]E-Cast F-28 with Hardner 215, United Resin Corp., 2730 S. Harbor Blvd., Suite F, Santa Ana, CA 92704.
[45]Eccoslip H-48, Emerson & Cuming, Canton, MA.

Molecular Acoustics

By

Henry E. Bass

Physical Acoustics Research Group

The University of Mississippi

TABLE OF CONTENTS

1. Absorption and Dispersion in Gases (Beyer) (40 min.)

1.1 Wave Equation

1.2 Classical Absorption/Dispersion

1.3 Relaxation Absorption/Dispersion

Break

2. Relaxation Absorption at the Molecular Level (60 min.)

2.1 Experimental Techniques

2.2 Typical Results

2.3 Relaxation in a Polyatomic Molecule

2.4 Relaxation in CS_2 Gas and Liquid

2.5 Example-Absorption of Sound in Air

Break

3. Classical Absorption at Low Pressure (20 min.)

3.1 Diffusion

3.2 Mixtures of Gases w/large Mass Disparity

3.3 Boltzmann's Equation for Dilute Gases

4. **Discussion** (20 min.)

4.1 Negative Absorption

4.2 Electrical Discharges

1.0 Absorption and Dispersion in Gases

Note: Much of the following was taken verbatim from the book Physical Ultrasonics by Robert T. Beyer and Stephen V. Letcher published by Academic Press in 1969. It is my understanding that a newer version of this classic is in preparation. I would like to thank Dr. Beyer for his permission to use this material. I have made changes to conform to my own notation and to emphasize selected points. I accept responsibility for errors introduced in the process.

1.1 Wave Equation

In the early part of the nineteenth century, Stokes[1] in England analyzed the effect of viscosity on the propagation of a sound wave in a gas. This analysis yielded an absorption coefficient, as we shall see later. In 1868, Kirchhoff[2] considered the effect of heat conduction in dissipating the energy of a sound wave, and presented a derivation which took both viscosity and heat conduction into account.

It proved impossible to test either of these theories quantitatively until the present century. By that time, these theories had been so long established in the literature that they had been given the name *classical.* Although the presence of both of these processes has now been proved experimentally, they are usually not the most important causes of sound absorption. Nevertheless, they can be treated simply, and can also serve as an introduction to the more complicated processes with which we must deal later.

Hopefully, you have all seen the equations governing the propagation of a plane wave in a dissipation-free medium. I will repeat many of those equations for completeness, but I am going to be especially interested in a microscopic interpretation of individual terms. Let us recall some of the features of that calculation. We begin with the equation of continuity (or conservation of mass):

$$-\partial(\rho\dot{\xi})/\partial x = \partial\rho/\partial t. \tag{1}$$

We also introduce the equation of motion (or conservation of momentum):

$$\rho_0 \ddot{\xi} = -\partial p/\partial x. \tag{2}$$

The first law of thermodynamics can be written in the form

$$\Delta Q = dU - \Delta W, \tag{3}$$

where ΔQ is the heat added (per mole) to the system in an infinitesimal process, dU is the corresponding increase in the internal energy of the system, and ΔW is the work done on the system during the same process. In particular, we use this equation under the adiabatic condition; that is, we assume that no heat enters or leaves the system during the process. In such a case, $\Delta Q = 0$. Furthermore, ΔW can be replaced by

$$\Delta W = -p\, dV = M(p/\rho^2)\, d\rho, \tag{4}$$

Where M is the gram molecular weight of the gas, while $dU = C_v dT$, where C_v is the heat capacity per mole at constant volume.

In the general case, the equation of state is an expression of the form

$$p = p(\rho, T). \tag{5}$$

In the case of an ideal gas, $p = \rho RT/M$, where R is the gas constant (per mole). In many cases however, the more general form of Eq. (5) is satisfactory.

Now let us suppose that a plane harmonic wave travels through the medium in the $+x$ direction. We shall write the expressions for the changes in the pressure (p_e),condensation [$s = (\rho - \rho_0)/\rho_0$], temperature ($\theta = T - T_0$), and velocity ($u = \dot{\zeta}$) associated with this wave in the complex form:

$$\begin{aligned} p_e &= p_{e0}\, e^{i(\omega t - kx)} & \theta &= \theta_0\, e^{i(\omega t - kx)} \\ s &= s_0\, e^{i(\omega t - kx)} & u &= u_0\, e^{i(\omega t - kx)}. \end{aligned} \tag{6}$$

We now substitute these values in eqs. (1) – (5). We shall make use of the linear approximation: that is, we shall neglect all products of small quantities. We then obtain the four equations

$$\text{(a)} \quad -\frac{\partial u}{\partial x} = \frac{\partial s}{\partial t} \qquad \text{(c)} \quad C_v \theta = \frac{M p_0}{\rho_0} s$$

$$\text{(b)} \quad \frac{\partial u}{\partial t} = -\frac{1}{\rho_0}\frac{\partial p_e}{\partial x} \qquad \text{(d)} \quad p_e = \left(\frac{\partial p}{\partial \rho}\right)_T \rho_0 s + \left(\frac{\partial p}{\partial T}\right)_\rho \theta \tag{7}$$

and therefore

$$\begin{aligned} iku \qquad & - i\omega s \qquad & = 0 \\ i\omega u \qquad & - (ik/\rho_0) p_e \qquad & = 0 \\ -M(p_0/\rho_0)\, s \qquad & + C_v \theta & = 0 \\ -(\partial p/\partial \rho)_T \rho_0 s \qquad & + p_e - (\partial p/\partial T)_\rho \theta & = 0. \end{aligned} \tag{8}$$

A nontrivial solution of Eq. (7) exists if the determinant of the coefficients vanishes. This yields the following expression for the velocity of propagation c_0:

$$c_0^2 = \omega^2/k^2 = (\partial p/\partial \rho)_T + (M p_0/\rho_0^2 C_v)(\partial p/\partial T)_\rho . \tag{9}$$

By making a number of thermodynamic transformations, it is possible to show that, to terms in first order, Eq. (8) is indeed equivalent to the general expression for the square of the sound velocity in a nondissipative medium, namely $c_0^2 = (\partial p/\partial \rho)s$. In the particular case of an ideal gas,

$$(\partial p/\partial \rho)_T = p_0/\rho_0 \qquad \text{and} \qquad (\partial p/\partial T)_\rho = p_0/T = \rho_0 R/M \tag{10}$$

so that

$$c_0^2 = (p_0/\rho_0)(1 + R/C_v) = \gamma(p_0/\rho_0). \tag{11}$$

1.2 Classical Absorption/Dispersion

Now, in what way will this development be altered by the presence of viscous drag? To answer this, we must look again at the equation of motion (2). The presence of viscosity adds new force terms to the right–hand side of this equation. In the most general form, the resultant equation of motion is known as the Stokes–Navier equation. For the one–dimensional case, this can be written

$$\partial(\rho\dot{\xi})/\partial t = -\partial p/\partial x + (\tfrac{4}{3}\eta + \eta')\,\partial^2\dot{\xi}/\partial x^2, \tag{12}$$

where η is the shear viscosity coefficient. The quantity η' is known as the bulk viscosity, and corresponds to the viscous drag that would be experienced in a pure

volume dilatation, in which no shearing motions can occur.

The nature and value of η' forms one of the most interesting problems in the historical development of ultrasonic wave propagation. Its discussion, however, will be postponed until the next hour. It was assumed by Stokes that η' was identically zero, and, to a large extent, this assumption marks the difference between classical and modern theories of ultrasonic absorption and dispersion.

Another question often raised at this point is why the shear viscosity should enter into the description of the motion of a plane wave of large extent, since no shearing motions are immediately obvious. The answer to this question lies in the fact that one cannot restrict all the motion to one direction without doing it as a combination of deformations in all the coordinate directions.

However, in a plane longitudinal wave the particle motion is in just one dimension, and a one–dimensional motion in either an elastic or viscous deformation must be considered as a combination of deformations in all the coordinate directions. The analysis has been given by Kittel[3] for the elastic case. The elasticity–viscosity analogy can then be used to obtain the desired result.

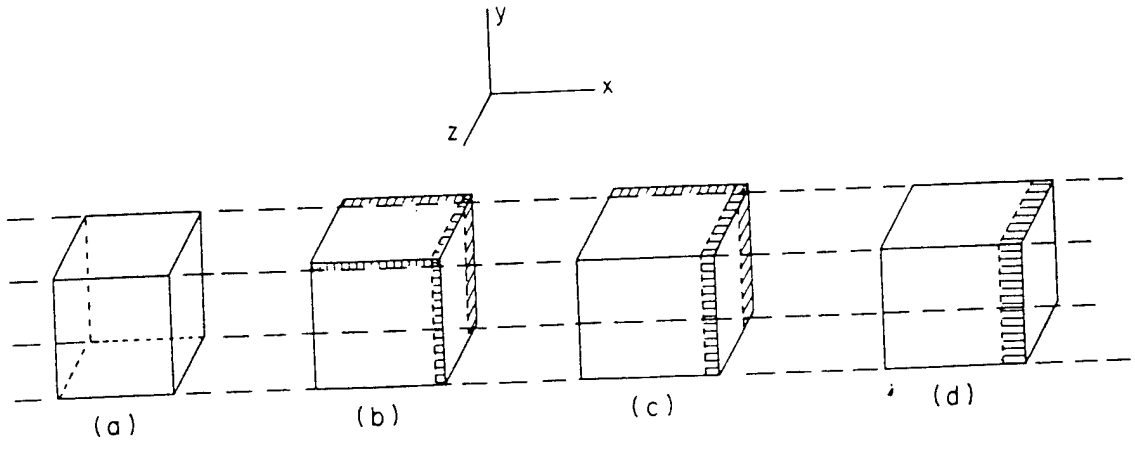


Fig. 1. Steps showing that a longitudinal dilation involves both compressional and shear moduli. See text for description.

Let us consider a unit cube of material (Fig 1a) that is to be stretched in the x direction to a strain ε_{11}, without any net change in either the y or z directions. This will be done in three steps:

1. Uniform strain by an amount $\varepsilon_{11}/3$ (Fig. 1b).
2. Extension in the x direction to a total strain $2\varepsilon_{11}/3$, with simultaneous compression to zero in the y direction, preserving a constant volume in the process (Fig. 1c).
3. Extension in the x direction to a total ε_{11} and simultaneous compression in the z direction to zero, again preserving a constant volume.(Fig. 1d).

In the first step, the stresses are uniform and given by

$$T_{11} = B\frac{\Delta V}{V} = \frac{B}{3}(\varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33}) = B\varepsilon_{11}, \tag{13}$$

where B is the bulk modulus, since $\varepsilon_{11} = \varepsilon_{22} = \varepsilon_{33}$ in the present case.

In the second step, we have a process that can be described as a shear[4] of angle $2\varepsilon_{11}/3$, so that the stress in the x direction will be $\mu \cdot 2\varepsilon_{11}/3$, where μ is the shear modulus.

The situation in the third step is identical with that in the second except that we have replaced y by z. Hence the total stress in the x direction will be equal to $(B + \frac{4}{3}\mu)$.

Now elementary viscosity theory follows directly as an analog of elasticity, with strains replaced by the corresponding velocity gradients and elasticity coefficients replaced by the corresponding viscosity coefficients, η', η. Hence we have for the viscous stress in the x direction

$$T_{11} = (\eta' + \tfrac{4}{3}\eta)\,\partial v_x/\partial x, \tag{14}$$

To proceed to a discussion of the effect of the shear viscosity on the propagation of

sound, we replace Eq. (2) by (5) with $\eta' = 0$, and make the linear approximation, thus obtaining

$$\partial u/\partial t = -\rho_0^{-1}\partial p_e/\partial x + \tfrac{4}{3}(\eta/\rho_0)\,\partial^2 u/\partial x^2, \tag{15}$$

or, for harmonic waves of frequency ω,

$$(\omega - \tfrac{4}{3}i(\eta/\rho_0)k^2)u - (k/\rho_0)p_e = 0. \tag{16}$$

Equation (16) is the appropriate modification of the second of Eq. (8), so that the corresponding determinant of the coefficients will now be

$$\begin{vmatrix} k & -\omega & 0 & 0 \\ \omega - i\tfrac{4}{3}(\eta/\rho_0)k^2 & 0 & -k/\rho_0 & 0 \\ 0 & -p_0 M/\rho_0 & 0 & C_v \\ 0 & -(\partial p/\partial\rho)_T\,\rho_0 & 1 & -(\partial p/\partial T)_\rho \end{vmatrix} \tag{17}$$

This leads to the result

$$k^2/\omega^2 = [c_0^2 + i\tfrac{4}{3}(\eta/\rho_0)\omega]^{-1}. \tag{18}$$

The presence of the imaginary term on the right side of Eq. (18) makes k complex. We therefore set $k = k_r - i\alpha$, so that all the Eqs. (6) can be written in the form

$$u = u_0 \exp -\alpha x \exp i(\omega t - k_r x), \tag{19}$$

in which α is the amplitude absorption coefficient, while k_r is the real wave number,

equal to ω divided by the phase velocity c.

The substitution of $k_r - i\alpha$ in Eq. (14) leads to the two equations

$$k_r^2 - \alpha^2 = \frac{\omega^2 c_0^2}{c_0^4 + (16/9)(\eta^2/\rho_0^2)\omega^2} \tag{20}$$

$$2k_r\alpha = \frac{(4/3)(\eta/\rho_0)\omega^3}{c_0^4 + (16/9)(\eta^2/\rho_0^2)\omega^2}. \tag{21}$$

In virtually all cases, $k_r^2 \gg \alpha^2$ (and hence $(16/9)(\eta^2\omega^2/\rho_0^2) \ll c_0^4$) so that, to an excellent approximation, we have

$$k_r = \omega/c_0$$
$$k_r\alpha = \tfrac{2}{3}(\eta\omega^3/\rho_0 c_0^4) \qquad \text{or} \qquad \alpha = \tfrac{2}{3}(\eta\omega^2/\rho_0 c_0^3). \tag{22}$$

It can now be seen that the approximation in parentheses above reduces to $k_r^2 \gg 4\alpha^2$. For comparison, α in oxygen, at 20° and at atmospheric pressure, is $1.68 \times 10^{-13}\nu^2$ in cgs units, while the corresponding value for water [both computed from Eq. (22) is $8 \times 10^{-17}\nu^2$. Hence, for O_2,

$$\alpha/k_r = 0.88 \times 10^{-8}\nu \tag{23}$$

(where ν is frequency) and, for water,

$$\alpha/k_r = 1.9 \times 10^{-12}\nu. \tag{24}$$

Each ratio is usually safely small at frequencies commonly used.

It may also be observed that the presence of viscosity leads to the irreversible production of heat, so that our use of the first law of thermodynamics must also be reconsidered. In the case of a reversible process, an adiabatic process ($\Delta Q = 0$) is also necessarily an isentropic one ($dS = 0$). In the irreversible process, however, there will be a production of entropy, even though the process is adiabatic and $\Delta Q = 0$. This will in turn alter Eq. (7c). Since the rate of energy dissipation is equal to the product of the viscous force $\frac{4}{3}\eta\, \partial^2 u/\partial x^2$ and the particle velocity u, we now have

$$C_v\, d\theta - (Mp/\rho_0{}^2)\, d\rho + \tfrac{4}{3}u\eta\, \partial^2 u/\partial x^2\, dt = 0 \tag{25}$$

or

$$C_v\, d\theta/dt - (Mp/\rho_0{}^2)\, d\rho/dt + \tfrac{4}{3}\eta u\, \partial^2 u/\partial x^2 = 0. \tag{26}$$

It is immediately clear, however, that the new term is of higher order of smallness than the first two terms, since it represents a product of u and $\partial^2 u/\partial x^2$. Under the linear approximation that we are using, such a term is to be neglected..

Nevertheless, there are cases in which this irreversible heat production must be taken into account. These include heat conduction, and the relaxation processes to be discussed later.

In the case of heat conduction, a term must be added to the energy conservation equation which represents the heat removed per second per mole ($\partial Q'/\partial t$). From the basic equation of heat conduction, we have

$$\partial Q/\partial t = (M\kappa/\rho_0)\, \partial^2\theta/\partial x^2, \tag{27}$$

where κ is thermal conductivity. Hence Eq. (7c) becomes

$$c_v \frac{\partial \theta}{\partial t} - \frac{p_0}{{\rho_0}^2} \frac{\partial \rho}{\partial t} - \frac{\kappa}{\rho_0} \frac{\partial^2 \theta}{\partial x^2} = 0, \tag{28}$$

where we have written c_υ for the heat capacity per unit mass (at constant volume), and where we have divided through by the gram molecular weight M.

For a harmonic wave we then have

$$[c_v - (i\kappa k^2/\rho_0 \omega)]\theta - (p_0/\rho_0)s = 0 \tag{29}$$

so that another change must be made in the determinant of (8). If we carry out the derivation in the usual way, we obtain

$$\frac{k^2}{\omega^2} = \frac{1}{{c_0}^2} \frac{1 - i\zeta}{1 - i(\zeta/{c_0}^2)(\partial p/\partial \rho)_T}, \tag{30}$$

where

$$\zeta = \kappa k^2/\rho_0 \omega c_v . \tag{31}$$

As written, Eq. (30) is not an explicit solution for k^2, since k is also contained in ζ. The quantity ζ is small however; for water, $\zeta = 6 \times 10^{-4}$. It is customary, however, to replace k in ζ by ω/c_0 so that $\zeta = k\omega/\rho_0 {c_0}^2 c_v$. Then, if we set $k = k_r - i\alpha$, we have, for the case of thermal conduction (after some algebra),

$$k_r^2 = \omega^2/c_0^2$$

$$2k_r\alpha = \frac{\omega^2}{c_0^2}\left[1 - \left(\frac{\partial p}{\partial \rho}\right)_T \frac{1}{c_0^2}\right]$$

$$\alpha = \frac{\omega^2 \kappa}{2c_v\, \rho_0\, c_0^3}\left[1 - \frac{1}{c_0^2}\left(\frac{\partial p}{\partial \rho}\right)_T\right]. \tag{32}$$

For the case of an ideal gas, $(1/c_0^2)(\partial p/\partial \rho)_T = 1/\gamma$, and

$$\alpha = \frac{\omega^2 \kappa}{2c_p\, \rho_0\, c_0^3}(\gamma - 1), \tag{33}$$

where $c_p = \gamma c_v$ is the heat capacity per unit mass (at constant pressure).

1.3 Relaxation Absorption/Dispersion

Studies of acoustic absorption and dispersion have proven to be an effective tool in developing a physical description of molecular interactions in gases and liquids. Experimentally, absorption or velocity dispersion is measured as a function of frequency (or, in gases as a function of frequency/pressure). These same quantities are computed using the theory outlined below which requires some insight into the particular physical processes which gives rise to the absorption or dispersion (e.g. vibrational relaxation). Computed values are compared to experiment and the assumed physical process or interaction parameters are varied until agreement is achieved. The theoretical model which

gives rise to predicted absorption and dispersion which agree with experiment are assumed to correctly represent the interactions taking place at the microscopic level.

Some of the microscopic processes studied acoustically can also be studied using optical techniques. In those cases, the physical model and numerical values for the parameters describing the interactions are found to be consistent provided a proper theoretical link between microscopic model and measured quantities is used.

For the purpose of developing an understanding of relaxation processes, first consider an ideal gas made up of diatomic molecules. The individual molecules are free to move translationally in three directions, rotate about two perpendicular axes (actually three but the third has zero moment of inertia so has no energy), and vibrate along the bond joining the atoms. Some energy is associated with each of these allowed motions.

Translational motion can be considered non-quantized. Any energy is allowed. As the molecules translate, they collide exchanging energy with their collision partners. At atmospheric pressure, assuming a hard sphere molecular model, molecular collisions take place at a rate of about 10^{11}/second. A single collision is typically sufficient to transfer translational energy from one molecule to another. However, a certain period of time is required to randomize energy associated with excess velocity in a particular direction. This time is often referred to as the translational relaxation time. Kohler following Maxwell[5], associates viscosity with this relaxation time writing $\tau_{tr} = \eta/p = 1.25\tau_c$ where p is gas pressure and τ_c is the time between collisions.

As the pressure is lowered, the rate of collisions decreases proportionately. At 1 torr (1/760 atm), the translational relaxation time is about 10^{-8} sec; at 1 millitorr 10^{-6} sec. These time regimes can be effectively studied using ultrasonics.

Unlike translational motion, rotation and vibration are noticeably quantized. During a collision, a change in rotational or vibrational state can only occur if the change in energy of another state is sufficient to allow at least one quantum jump. For rotational energy

transfer, the spacing between energy levels is given by 2(J+1)B where J is the rotational quantum number, $B = \hbar^2/2I$ and I is the effective moment of inertia. If one assumes J is the most probable value (from a Boltzman distribution), the value of 2(J+1)B for a typical molecule (say N_2) in units of kT is about 1°. This means that in a gas above 1K, essentially all collisions will have sufficient translational energy to cause multiple changes in J. As a result, rotation rapidly equilibrates with translation.

An exception is hydrogen which has much larger rotational energy level spacing due to the small moment of inertia. On the average, as many as 350 collisions[6] may be necessary to transfer a quantum of rotational energy in H_2. At a pressure of one atmosphere, this gives a relaxation time of about 2×10^{-8} sec. It should be noted that a given collision does or does not transfer a quantum of energy. The 350 collision average means that only one collision in three hundred and fifty has the proper geometry and energy to cause a transfer of one quantum of rotational energy. The number of collisions necessary, on the average, to transfer one quantum of energy is referred to as the collision number Z. When rotational energy is involved, the subscript "rot" is typically added (Z_{rot}). The inverse of this dimensionless quantity is the probability of transferring a quantum in a collision (P_{rot}). Since rotational energy level spacings are unequal, a $1 \rightarrow 2$ transition should be more probable than a $2 \rightarrow 3$ transition. These events are distinguished by using the symbols P_{rot}^{1-2} or P_{rot}^{2-3}.

Generally speaking, the probability for transferring a quantum of energy in a collision decreases rapidly with the size of the quantum transferred. Since vibrational levels are much more widely spaced than rotational energy levels, vibrational relaxation times are much longer than rotational. Vibrational levels in a single vibrational mode are approximately equally spaced. This means that energy can be exchanged between levels (i.e. the vibrational quantum number goes up in one molecule and down in the other) with very little energy exchanged between vibration and translation. The result is that such v-v

exchanges take place very rapidly. The vibrational relaxation time is controlled by the time it takes energy to transfer between translation and the lowest lying vibrational level. Since this energy level varies greatly for different molecules so do the probabilities of vibrational energy transfer during a collision. During N_2 collisions with N_2, Z_{10} is near 1.5×10^{11} so the relaxation time is near 1.5 seconds[7] (Z_{10} is the number of collisions needed to transfer energy from the lowest vibrational level to translation). Large molecules have vibrational energy levels which are very close together. A molecule such a C_2H_6 requires only 100 collisions to transfer a quantum of vibrational energy from the first vibration level into translation.[8] This very wide range of relaxation times presents interesting experimental challenges.

To this point, there has been no attempt to rigorously define relaxation times in terms of energy transfer probabilities. In fact, such a relation is possible in a simple form for only the few cases where gases exhibit a single relaxation time. Nevertheless, the simple case provides valuable insight into the behavior of more complex systems and deserves a detailed description.

Consider the case where the population of a vibrational state is excited to an energy E_v which is greater than the energy $E_v(T_{tr})$ which it could have in Boltzman equilibrium with translation. In this case, the excess vibrational energy will equilibrate with translation according to a standard relaxation equation,

$$-\frac{dE_v}{dt} = \frac{1}{\tau}\left[E_v - E_v(T_{tr})\right] \tag{34}$$

The return to equilibrium occurs due to energy transfer during individual collisions.

The rate of energy transfer k_{10} is defined as the rate at which molecules go from the first excited state to the ground state due to collisions at a pressure of one atmosphere. This rate is just the collision frequency, M, times the probability of energy transfer, $P^{1\rightarrow 0}$, times

the mole fraction of molecules in the first excited state, x_1. During some collisions, the reverse process will occur, that is some molecules in the ground state will become excited at a rate k_{01}. In equilibrium, equal numbers of molecules go in both directions so

$$k_{10}\, x_1 - k_{01}\, x_0 = 0\,. \tag{35}$$

As explained above, energy is quickly shared from the first excited level of the vibrational mode to higher levels of the mode by v-v exchanges. Assuming quantum mechanical laws hold for probabilities of energy exchanges between vibrational levels of a harmonic oscillator, Landau and Teller showed[9]

$$-\frac{dE_v}{dt} = k_{10}(1-e^{-h\nu/kT})\left[E_v - E_v(T_{tr})\right], \tag{36}$$

where ν is the vibrational frequency of the relaxing mode. By comparison with Eq. (34),

$$\tau = \frac{1}{k_{10}(1-e^{-h\nu/kT})}\,. \tag{37}$$

The link between the relaxation time and ultrasonic absorption and dispersion is understood by noting that the relaxation process makes the specific heat of the gas time (or frequency) dependent. This time dependence can be obtained from the energy relaxation equation. Consider, as above, that the specific heat of a simple gas can be divided up into translational, rotational, and vibrational contributions. For now, assume that the translation and rotational energy both equilibrate rapidly enough to follow any acoustically induced temperature variations. In this case, the effective specific heat can be written as

$$(C_v)_{eff} = C_v^{\infty} + C' \frac{dT'}{dT_{tr}}\,, \tag{38}$$

where C_v^{∞} is the sum of rotational and translational specific heats, C' is the relaxing specific heat, and T' is the instantaneous temperature of the relaxing mode (in this case, vibration). From the energy relaxation equation (Eq.36), for small periodic variations in T_{tr} and T_v about their equilibrium values,

$$(C_v)_{eff} = C_v^{\infty} + \frac{C'}{1 + i\omega\tau} , \tag{39}$$

where ω is the angular frequency of the acoustic wave.

The acoustic propagation constant can be written in the form

$$\frac{k^2}{\omega^2} = \left(\frac{1}{c} - \frac{i\alpha}{\omega}\right)^2 = \frac{\rho_0 \kappa_T}{\gamma_{eff}} , \tag{40}$$

where c is the acoustic velocity, α is the attenuation, ρ_0 is the equilibrium density, κ_T is the compressibility, and

$$\gamma_{eff} = [(C_v)_{eff} + R] / (C_v)_{eff} \tag{41}$$

with R the gas constant. For this simple single relaxation, assuming $\alpha/\omega << 1/c$,[10]

$$\alpha\lambda = \pi \left(\frac{c}{c_0}\right)^2 \varepsilon \frac{\omega\tau_s}{1 + (\omega\tau_s)^2} \tag{42}$$

and

$$\left(\frac{c_0}{c}\right)^2 = 1 - \frac{\varepsilon \omega^2 \tau_s^2}{1 + \omega^2 \tau_s^2} \tag{43}$$

where

$$\varepsilon = \left(\frac{c_\infty^2 - c_0^2}{c_\infty^2} \right)$$

and where λ is the wavelength, c_0 is the speed of sound for $\omega\tau_s << 1$, c_∞ is the speed of sound at frequencies much greater than the relaxation frequency. The adiabatic relaxation time τ_s, is related to the isothermal relaxation time τ used earlier by

$$\tau_s = (C_v + R)/(C_v^\infty + R)\,\tau\ . \tag{44}$$

The relaxation frequency, f_r, defined as the frequency at which the maximum absorption per wavelength occurs is related to τ_s by

$$f_r = 1/2\pi\tau_r = c_\infty/c_0\ (1/2\pi\tau_s) \tag{45}$$

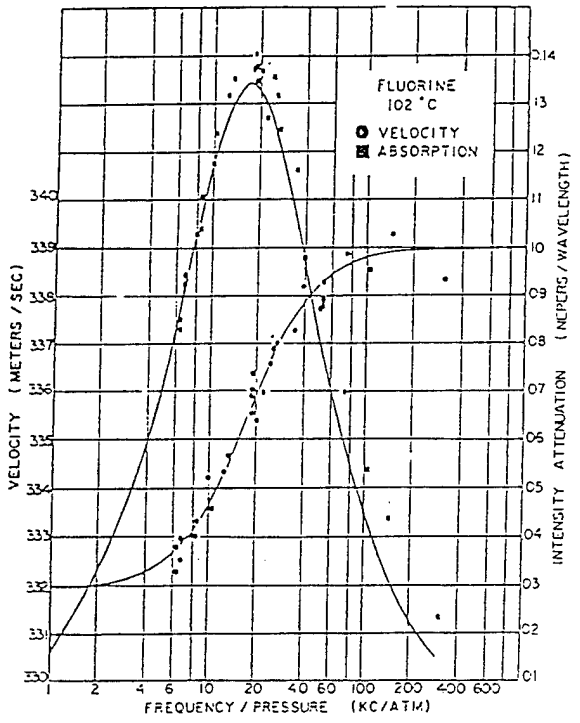


Figure 2 shows typical curves for absorption per wavelength and velocity dispersion to a single relaxation process. The example here is Fl_2 at 102°C. The figure compares curves representing the above theory with measured values.[11]

In the case of polyatomic gases or mixtures of relaxing diatomic gases, the different relaxing modes can be coupled together by v-v exchanges. Such complex or multiple relaxation processes exhibit the general behavior given by Eqs. (42) and (43) but the magnitude of the absorption and dispersion and the relaxation frequencies can take on new meaning. In this case of multiple relaxing internal energy modes, Eq. (42) and Eq. (43) take on the form [12]

$$\left(\frac{c}{c^{\infty}}\right)^2 = 1 + \Sigma_j \frac{\delta_j \, k_s/k_s^{\infty}}{1 + (\omega\tau_{s,j})^2} \tag{46}$$

$$\alpha\lambda \left(\frac{c}{c^{\infty}}\right)^2 = -\pi \, \Sigma_j \frac{\delta_j \, k_s/k_s^{\infty}}{1 + (\omega\tau_{s,j})^2} \; \omega\tau_{s,j} \tag{47}$$

where $\delta_j \, k_s/k_s^{\infty}$ is a relaxing adiabatic compressibility (negative) and j indicates that there might be more than one relaxation process. In these complex cases, $\tau_{s,j}$ can no longer be related to a single energy transfer reaction and $\delta_j \, k_s/k_s^{\infty}$ can no longer be related to relaxing energy of a specific mode. Instead, the various modes and reaction pathways are coupled. The sums in Eq. (46) and Eq. (47) are over eigenvalues of the energy transfer matrix which simultaneously accounts for all reactions. Equations (46) and (47) have been used to calculate the sound absorption in moist air as a function of frequency and temperature. The standard for such calculations is now based upon these equations.

Not only can Eqs. (46) and (47) be used to calculate sound absorption and velocity, but the reverse process is also possible, i.e., the transition rates can be extracted from measured values of absorption and velocity. However, the number of possible relaxation paths multiplies rapidly with the increase in the number of relaxing modes, and the identification of specific rates becomes a tedious process and usually involves some

assumptions. It has been done for only a few special cases. During the next hour we will discuss one example, i.e. SO_2/O_2 mixtures.

So far, only relaxation processes which involve the exchange of energy in gases have been considered. Conceptually, relaxation processes in liquids result in similar relaxation equations but there are important differences in detail. These differences are brought about by the greater density of molecules which make multi-body interactions typical. In this case, the idea of a rate equation is less precise but the existence of a relaxation time as a measure of the time required for the system to return to equilibrium following a perturbation remains valid.

Various interesting relaxation processes occur in liquids. In some cases, such as CS_2 and a number of organic liquids, the relaxation mechanism seems to be the same as that described above for gases, i.e. the internal energy of the individual molecules is being excited in "collisions." Such liquids generally have a positive temperature coefficient of absorption and are called Kneser liquids.[10,5] It is more common, however, that in liquids, molecules temporarily bond to form large groups which may re-configure themselves when an ultrasonic wave passes through. Such configuration relaxations tend to be very fast and the frequency dependence of the absorption and dispersion may indicate a distribution of relaxation times. H_2O is an example of such a liquid, generally referred to as associated liquids.[5]

A further complication arises from chemical reactions. For a reversible chemical reaction with heat of reaction ΔH, ΔH enters into the relaxation equations in a manner similar to ΔE for vibrational relaxation. A major difference is that chemical reactions allow the possibility that the number density of molecules can change. Such changes bring about additional relaxation absorption and dispersion.

2.0 Relaxation Absorption at the Molecular Level

2.1 Experimental Techniques

Before modern techniques were available for measuring transients most ultrasonic absorption and velocity measurements were made with ultrasonic interferometers.[13,14,15,16] This instrument continues to be used with refined precision and modern methods of control and measurement.[17] Both velocity and absorption can be measured with the interferometer. A column of gas or liquid of varying length forms the load on a quartz crystal vibrating at its resonant frequency. The loading effect of the gas or liquid column increases whenever the length of the gas column is a whole multiple of a half-wavelength of the sound. This loading effect is reflected in the driving circuit of the crystal. The separation between the peak values determines the sound wavelength and the variation of the magnitude of the peaks with length of the column allows the determination of the absorption coefficient of the sound in the test medium.

Another continuous-wave-method for measuring velocity and absorption in gases uses a source and receiver whose separation can be varied that are mounted in a tube so as to avoid standing waves.[18,19] See Figure 3. This method has been used to make measurements at audible frequencies and reduced pressures, and therefore, is capable of measurements over the f/p range where many of the interesting relaxation processes occur in gases and gas mixtures. The absorption is determined from the decrease in sound amplitude at the receiver as the source-receiver separation is increased. The wavelength of the sound is equal to the distance between points of equal phase in the received sound. This change in phase is easily observed by comparing the receiver and source signals.

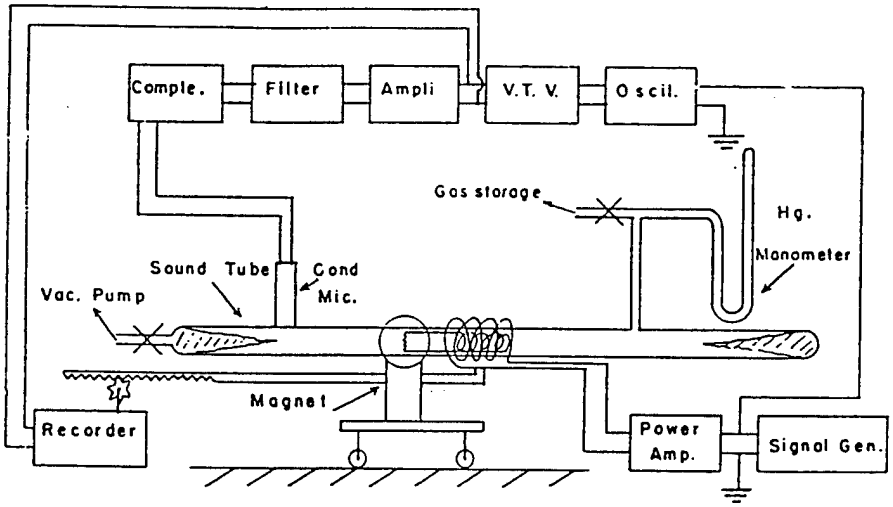


Figure 3. Schematic diagram of the apparatus for measuring the attenuation of sound.

With the development of the capability for generating and measuring tone bursts it has been possible to use the pulse-echo technique[23,24] In this case the sound velocity is determined from time of flight and the absorption from the variation of the tone burst amplitude with source and receiver separation.

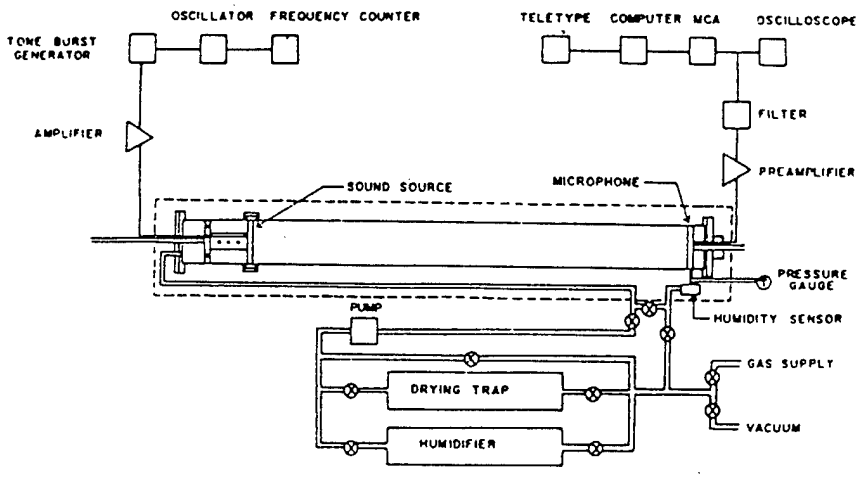


Fig. 4. Diagram of the experimental system for measuring sound absorption.

At lower frequency-to-pressure ratios resonant tubes are used.[7] The velocity is determined from the value of the resonant frequency and the absorption from the width of the resonant peak.

In the case of liquids, special techniques have been developed to measure absorption and velocity at very high frequencies where many of the interesting relaxation effects occur.[5]

When relaxation effects are studied in gases the measurements of sound absorption are designed to provide absorption as a function of frequency divided by pressure over a range of values where the period of the sound wave is approximately equal to the relaxation time. When studying relaxation in molecular N_2, this requires measurements of very small absorption (less than 1 dB/m) at f/p values as low as a fraction of a Hertz per atmosphere. On the other hand, measurements of absorption in UF_6 involve very large attenuation (~100 dB/m) at f/p values as high as 10 MHz/atm.[2]

In order to separate the effects due to vibrational relaxation, the measured absorption must be corrected for:

1. viscous and thermal losses in the body of the gas
2. viscous and thermal losses to the measuring chamber walls
3. radiation or leakage losses from the chamber
4. spreading losses
5. losses due to rotational relaxation.

These different effects are generally small enough to be additive. The quantity usually reported is the absorption coefficient, α, in terms of which the amplitude of the plane wave is written as:

$$A = A_0 e^{-\alpha(x_2-x_1)}. \tag{48}$$

Plotting log A versus x gives α as a slope in nepers/m. The measured α is due to a combination of the losses listed above. Careful design of the experimental apparatus is necessary if accurate corrections are to be made for wall and/or spreading losses.

A common geometry for sound absorption measurements is a cylindrical tube. This method generally avoids losses 3 and 4 (due to spreading and leakage) above and involves well known corrections for 1 and 2. In this case, the total absorption is given by

$$\alpha_{tot} = \alpha_{cl} + \alpha_{rot} + \alpha_{tu} + \alpha_{v} , \tag{49}$$

where α_{cl} (classical absorption) is the absorption due to viscosity and thermal conductivity, α_{rot} is absorption due to relaxation processes not of interest to the study (rotational relaxation for example), α_{tu} is the tube absorption, and α_{v} is the absorption due to the relaxation process of interest. The term α_{cl} can be computed from known or measured values of viscosity and thermal conductivity;[23] it is proportional to f^2. The term α_{rot} generally is estimated based upon other studies. At frequencies well below the rotational relaxation frequency it also is proportional to f^2. The tube absorption, α_{tu} , depends upon the viscosity, thermal conductivity and tube radius (r). It varies approximately as[24] $[f/(pr^2)]^{1/2}$.

Only plane waves will be present in the tube if the sound frequency is maintained below the cut-off frequency for the first non-plane mode, i.e. $f < 0.586c/d$, where c is the sound speed and d the tube diameter. Plane waves can be maintained for higher frequencies if the transducer generating the waves fills the tube, is carefully maintained perpendicular to the tube axis and its surface vibrates as a piston.[25] Interferometers used for high frequency measurements in liquids and gases will generally satisfy these conditions. As the wavelength becomes small compared to the tube diameter, wall losses become negligible.

Low frequency absorption measurements in liquids are difficult. Because of the high

sound velocity and small absorption large volumes are required. If the measurements are made in an unbounded liquid, spreading losses will likely dominate. If they are made in an enclosure the small compressibility of the liquid makes invalid the rigid wall assumption and it is difficult to correct for wall losses.

In the above discussion of relaxation in gases the relaxation absorption and dispersion of sound resulted when a finite time was required for the passage of the energy of translation into the vibrational energy of the gas molecules. Relaxation mechanisms can also be studied in special cases using the reverse processes. Laser light of a particular frequency is used to excite a vibrational mode. This excess vibrational energy then relaxes into translational energy generating an increase in translational temperature and pressure. If the laser light is chopped a sound wave is generated. This phenomena is referred to as the "optoacoustic effect" and the device as a "spectrophone".[26] The relaxation time can be determined from the phase relations between modulated laser light and the resulting sound pressure.

The primary attraction of optoacoustic measurements is the ability to excite a specific internal mode and observe how energy from that mode makes its way to translation. Experimentally, only a limited number of internal modes can be excited with available lasers which limits the list of systems which can be studied.

2.2 Typical Results

A comprehensive review of all studies of ultrasonic relaxation is far beyond the scope of this article. No such review has been published in recent years. The best available is that by Herzfeld and Litovitz which is now 37 years old.[5] For the purposes of this article specific studies will be selected which illustrate the physics involved. Three cases will be treated. The first is the halogen family of diatomic molecules, specifically F_2, Cl_2, Br_2 and I_2. This example illustrates the functional forms presented earlier for absorption due to

relaxation processes, gives typical values for probability of energy transfer, $P^{1\rightarrow 0}$, and provide some indication of how $P^{1\rightarrow 0}$ varies with temperature, molecular weight, and vibrational frequency. The second example is a more complex molecule, SO_2, in mixtures with Ar and O_2. This example illustrates the complexity of the relaxation process when different vibrational modes exchange energy. The final example will be for CS_2 which illustrates the case where relaxation has been studied in both the liquid and gas phase.

2.2.1 Simple Relaxation in Halogen Gases

Shields used an acoustic traveling wave tube similar to that developed by Angona[18] to measure acoustic absorption in F_2, Cl_2, Br_2 and I_2 as a function of frequency, pressure and temperature.[11, 27] Results for Cl_2 at five different temperatures are plotted in Fig. 5. The absorption and velocity in F_2 at 102 °C are shown in Fig. 2. The curves drawn through the experimental points were calculated using the theory discussed previously.

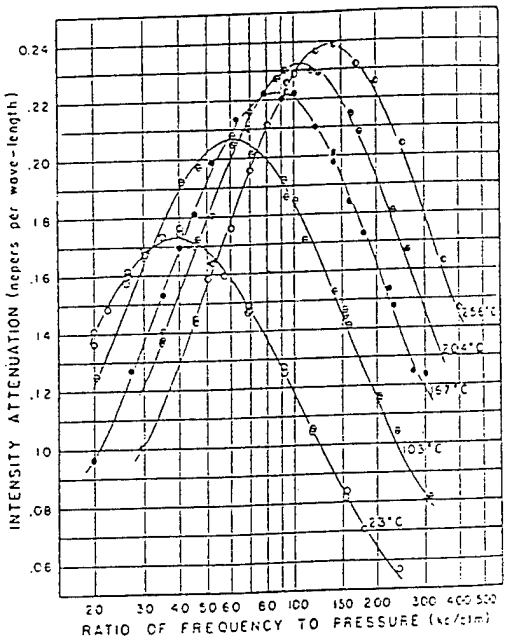


Fig. 5. Relaxation absorption coefficient per wavelength vs log (frequency/pressure) for chlorine. The solid curves are theoretical absorption with values of Am and fm adjusted to give the best fit of the experimental points.

Several features of these curves are interesting. First, when more than one point is plotted for the same value of f/p, it means that absorption was measured at two different pressures but at different frequencies so that the ratio was the same. The agreement between such measured values confirms that the classical and tube corrections were being made correctly and that the relaxation absorption varied as f/p. Second, note that the relaxation time decreases as temperature increases (the relaxation frequency increases). This means that although the density is lower for a given pressure, the probability of energy transfer increases more rapidly than the collision frequency decreases. Finally, note that the maximum absorption increases as the temperature increases. This is a results of increasing vibrational specific heat. The Plank-Einstein relation predicts that

$$C' = R\left(\theta/T\right)^2\left[e^{\theta/T} / \left(e^{\theta/T} - 1\right)^2\right]. \tag{50}$$

where C' is the vibrational specific heat, θ is the vibrational temperature, and T is the gas temperature.

Results of Shields' study are summarized in Figure 6 which shows the probability for de-exciting the first excited vibrational level in a collision. The log of this probability is plotted vs $T^{-1/3}$. This type of plot was suggested by an early theory by Landau and Teller.[9] They predicted an approximately linear relationship between log $P^{1\rightarrow 0}$ versus $T^{-1/3}$.From similar measurements on a great many other gas one can expect the following trends for v–t transitions:

1. The log of the transition probability increases roughly linearly with $T^{-1/3}$ for a particular molecular collision pair. The reason for this dependance is complex but can be thought of in terms of the speed molecules are traveling when they collide. The more rapidly they are moving, the more energy they carry into the collision. As the temperature increases, the average speed of each molecule

increases hence, there is a greater probability of energy transfer.

2. The transition probability is very sensitive to the amount of energy that must be transferred between vibration and translation in the transition, increasing rapidly as this energy decreases. This can be thought of the same as the temperature dependance. When two molecules collide with translational energy less than a quantum of vibration, the probability of a transition is classically zero.
3. The transition probability is very sensitive to the time involved in the molecular collision increasing rapidly as this time decreases. Wheat this means is that the transition probability increases with increasing temperature (see 1 above) and with decreasing molecular mass. This effect is due to the requirement that momentum as well as energy be conserved in the collision.
4. For polyatomic molecules, bending vibration are more easily excited than stretching vibration (as one might expect from the geometry of the collision process). There is a larger collision cross section for excitation of a bending mode.
5. Water vapor and a few other molecules with low moments of inertia, and therefore, high rotational velocities are very efficient in shortening relaxation times when added to gases, even in a small amounts. For such mixtures the relaxation time is much less sensitive to temperature change and can even increase with temperature. This effect is attributed to the coupling between vibration and rotation in such collisions.[28–32]
6. The probability of the transition is sensitive to the nature of the intermolecular potential. Figure 7 shows some single potentials. These can give reasonable results for atom-diatom collisions and for molecules that have electron clouds which are reasonably spherical. For these cases, the probability of a transition depends upon the depth of he potential and the steepness of the repulsive part of the potential.

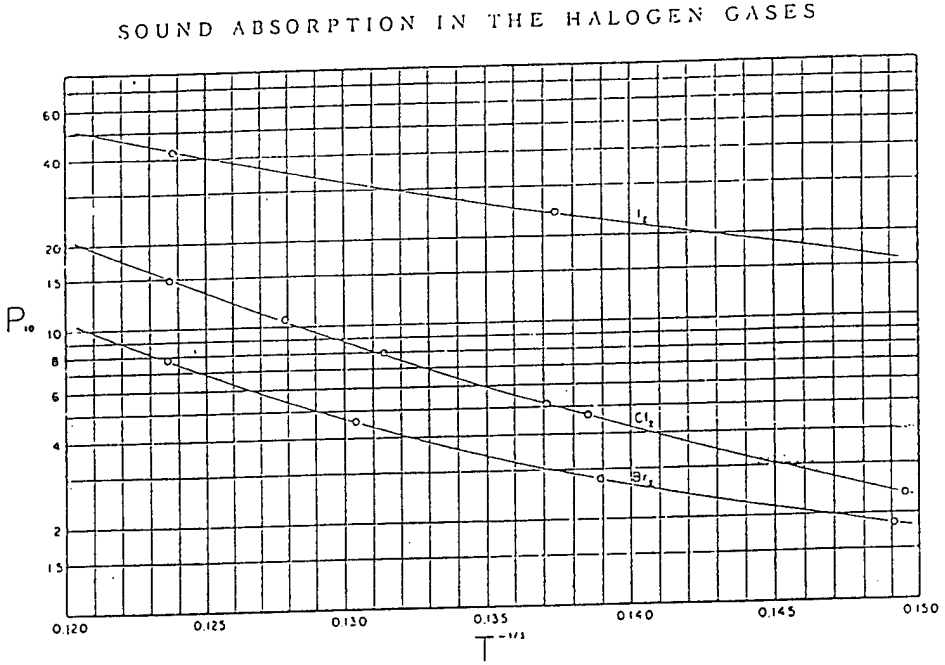


Fig. 6. Log of collision efficiency vs absolute temperature to the minus one-third power. The values of the ordinate should be multiplied by 10^{-5} for Cl_2 and by 10^{-4} for Br_2 and I_2.

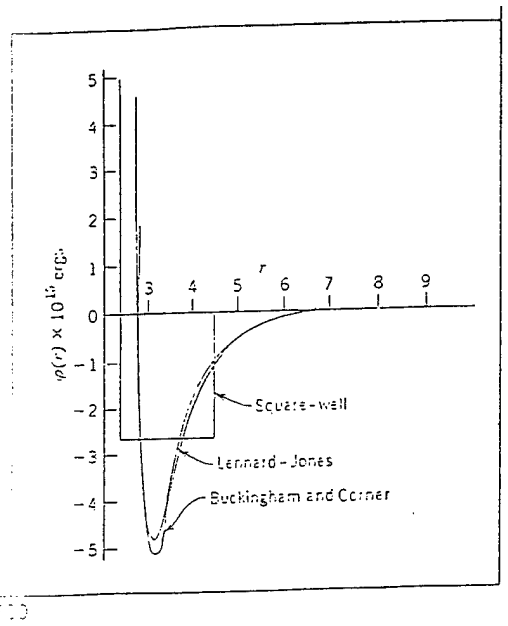


Fig. 7. Potential energy of interaction.

2.2.2 Relaxation in a Polyatomic Molecule

As an example of the use of sound absorption measurements to determine the relaxation scheme in a polyatomic gas mixture consider the case of SO_2/O_2 mixtures.[33] SO_2 has three

vibrational modes. These can be classified as a bending mode with a fundamental vibrational frequency of 518 cm^{-1}, a symmetrical stretching mode, and an asymmetrical stretching mode with frequencies of 1151 and 1361 cm^{-1} respectively. O_2 has a vibrational frequency of 1580 cm^{-1}. SO_2 was one of the first gases in which the sound absorption versus frequency curve evidenced more than one relaxation time. Figure 8 shows measurements in three different SO_2/O_2 mixtures at 500 °K. Twelve different transition rates were adjusted to make the theoretical curves simultaneously fit these data and the data for pure SO_2, SO_2/Ar mixtures,[34,35] and pure O_2.[36]

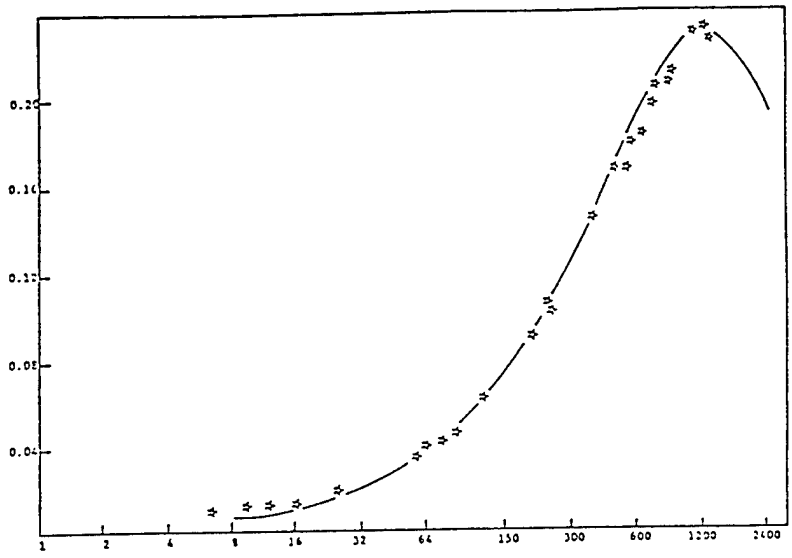


Fig. 8a Relaxation absorption in 75% SO_2/25% O_2.at 500°K. The solid curve is the theoretical curve for the series relaxation process and was obtained from the transition probabilities in Table II.

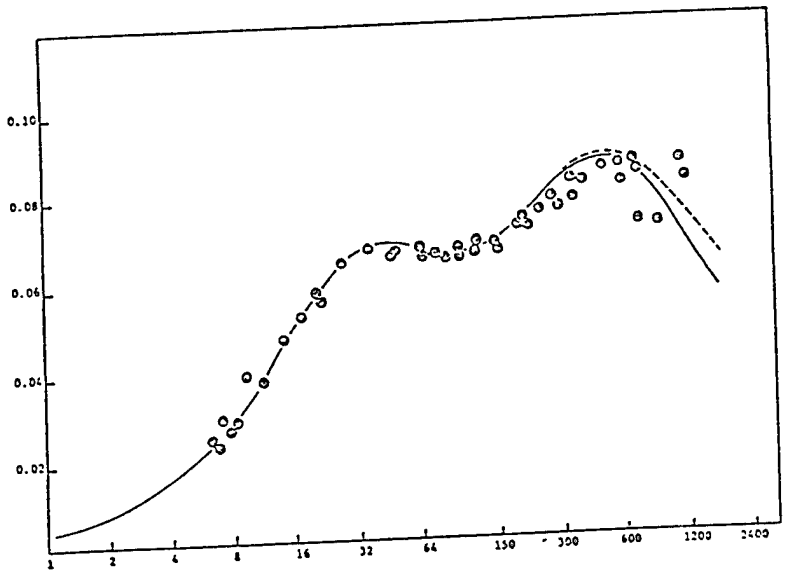


Fig. 8b. Relaxation absorption in 20% SO_2 80% O_2 at 500°K.

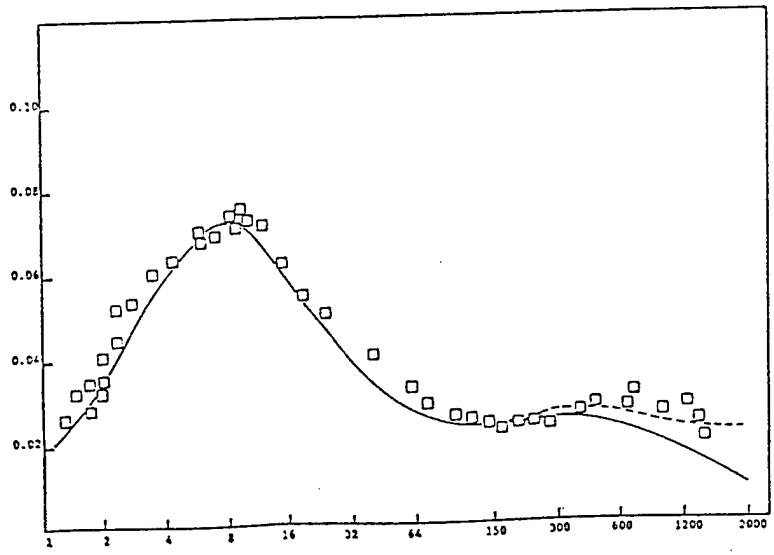


Fig. 8c. Relaxation absorption in 5% SO_2/95% O_2 at 500°K.

As an illustration of the kind of information that can be obtained from such studies, the following conclusions from the SO_2/O_2 and SO_2/Ar measurements are listed:

1. The relaxation in SO_2 is primarily a series process where translational energy flows into the bending mode and from there is shared with the symmetrical stretch with a two quantum for one exchange and from there with the asymmetrical stretch with a one for one exchange.
2. Ar is only about 1/10 and O_2 about 1/4 as effective as another SO_2 molecule in de-exciting the bending vibration in SO_2.
3. The first excited level of the O_2 vibration exchanges energy in vibration-to-vibration exchange primarily with the first excited level of the asymmetric stretch mode of SO_2.
4. The transition rate for the SO_2 vibration to vibration transfer increases with temperature faster than the vibration-to-translation rates. This is a peculiar property of SO_2 and has been attributed to its dipole moment.

2.4 Relaxation in CS_2 Gas and Liquid

Gaseous CS_2 has an unusually high vibrational specific heat at room temperature. From what has been said above, it is expected therefore, to have a large relaxation absorption at moderate f/p values. Such was found early to be the case.[37,38] Only a single relaxation absorption peak was found, indicating a series relaxation process with energy passing from translation first into the low energy bending vibration and from there shared quickly with the stretching modes. The frequency of maximum absorption increased from about 370 to 560 kHz/Atm as the temperature increased from 0 °C to 160 °C. This

corresponds to a transition probability that decreases relatively slowly (by about 1/3) in this temperature interval.[38]

More interesting is the absorption of CS_2 as a liquid. It is an example of a "Kneser" liquid where the vibrational relaxation can be identified in the liquid state just shifted to a higher frequency due to the higher molecular collision rate in the liquid state.

The concept of a collision loses meaning when dealing with liquids since all molecules are interacting with one or more other molecules simultaneously. If one does assume that the liquid is a very dense gas, a collision frequency in the range of 10^{13}/sec would be appropriate. This immediately suggests that relaxation processes in liquids will be fast. In fact, for most liquids up to a frequency of 10 MHz, no relaxation peak is observed so we must assume that $\omega\tau<<1$. Invoking this low frequency limit, Eq. 42 gives

$$\alpha = bf^2 \tag{51}$$

where b depends upon the relaxation time and relaxation strength. This is the same frequency dependence exhibited by viscous losses. As a result, relaxation, viscous and thermal conduction losses are difficult to sort out for most liquids.

Carbon Disulfide has a maximum in the quantity α/f^2 in the range of 100 MHz.[39] Takagi has shown that the experimental data is best explained in terms of two relaxation processes. The strongest process is associated with relaxation of the two lowest vibrational modes (ν_2 and ν_1) and has a relaxation time of 2.2 ns. The weaker relaxation is associated with relaxation of ν_3 with a relaxation time of 30 ns.

The observation that relaxation in CS_2 behaves like thermal relaxation in a dense gas makes it a very interesting liquid. CS_2 also strongly absorbs 337 nm UV which make it amenable to study with a pulsed N_2 laser.[40] Optoacoustic studies using a pulsed N_2 laser give relaxation times for the slow process of 90 ns. Given the difficulties in extracting the relaxation time for this weak process, a factor of three difference should not be surprising.

2.5 Example–Absorption of Sound in Air

Combining Eqs. (17) and (24).

$$\alpha = [\omega^2/(2\gamma Pc)][4\mu/3 + (\gamma - 1)\kappa/(\gamma c_v)] \tag{52}$$

in Np m^{-1}. Note that we now denote viscosity by μ in place of γ.

The Eucken expression[41]

$$\kappa = (15R\mu/4)[4c_v/(15R) + 3/5] \tag{53}$$

in J $(kg\ mol)^{-1}\ K^{-1}\ kg\ m^{-1}\ sec^{-1}$, can be used to relate κ and μ for real air. For dry air over the temperature range 225–370 K, the ratio c_v/R has the value $5/2 \pm 0.4\%$ at the most. For moist air, over the same temperature range, the c_v/R ratio becomes increasingly larger than the ideal gas value as the concentration of water vapor increases, but only by a small amount. For example, at a temperature of 370 K and 100% relative humidity, $c_v/R = 2.565$ or just 2.6% more than the ideal gas value[44]. Therefore, if we assume $c_v/R = 5/2$ in Eq. (12), the computed value for κ cannot be more than 1% different from the value calculated using the correct value of c_v/R. Using this assumption yields $\kappa = 19R\mu/4$. For an ideal gas, $c_p = c_v + R$. Thus, with $c_v/R = 5/2$, the ratio $c_p/R = 7/2$. Substituting $\kappa = 19R\mu/4$ into Eq. (11) and using $(\gamma - 1)R/(\gamma c_v) = 4/35$ with $\gamma = c_p/c_v = 1.4$ gives

$$\alpha = [2\pi^2 f^2/(\gamma Pc)](1.88\mu) \tag{54}$$

in Np m^{-1}.

For mixtures of gases, there is, in general, an additional term due to internal diffusion. It has been shown, however, that for air, this term is small and will be ignored here.

The classical absorption given by Eq. (54) changes with temperature due to changes in the speed of sound c, and the coefficient of viscosity μ. From Eq. (10), c varies as $T^{1/2}$. The coefficient of viscosity for dry air can be written in the form of Sutherland's equation [41]

$$\mu = \frac{\beta T^{1/2}}{1 + (S/T)} \tag{55}$$

in kg m^{-1} sec^{-1}, where B and S are empirical parameters. For air, the parameters B = 1.458×10^{-6} kg m^{-1} sec^{-1} $K^{-1/2}$ and S = 110.4 K are standard values[42] that provide a satisfactory fit to experimental data and permit reasonably accurate calculations of μ for the range of temperatures between 273.15 and 313.15 K (0° and 40°C).

Using the standard values of γ, R, and M for air from the Society of Automotive Engineers[43] and defining a reference temperature T_0 to be 293.15 K gives

$$c = 343.23(T/T_0)^{1/2} \tag{56}$$

in m sec^{-1}, and

$$\mu = 7.318 \times 10^{-3} \frac{(T/T_0)^{3/2}}{T + 110.4} \tag{57}$$

in kg m^{-1} sec^{-1}, so Eq. (54) becomes

$$\alpha_{cl} = 5.578 \times 10^{-9} \frac{T/T_0}{T + 110.4} f^2/(P/P_0) \tag{58}$$

in Np m^{-1}, where P_0 is the reference pressure of 1.01325×10^5 N m^{-2} or 1 atm, the pressure P is in N m^{-2}, and the temperature T is in degrees kelvin.

It was shown earlier that for a single relaxing degree of freedom

$$\alpha = -\frac{\delta K_s/K_s^\infty}{2c} \frac{\omega^2 \tau'_{vs}}{1 + (\omega\tau'_{vs})^2}, \tag{59}$$

in Np m^{-1}, where K_s^∞ is the instantaneous adiabatic compressibility (in units of N m^{-2} and equal to Pc_p^∞/c_v^∞); δK_s is the relaxing compressibility; and τ'_{vs} is the relaxation time at the partial pressure of the reactants in the mixture, at constant volume and under adiabatic conditions. For a single relaxing degree of freedom, $(\delta K_s/K_s^\infty) = -Rc'/[c_p^\infty(c_v^\infty + c')]$, where c' is the specific heat of the relaxing mode in J $(kg\ mol)^{-1}$ K^{-1}; c_p^∞ is the specific heat at constant pressure at frequencies $\gg 1/(2\pi\tau'_{vs})$; and c_v^∞ is the specific heat at constant volume under the same conditions. For frequencies $< 1/(2\pi\tau'_{vs})$, the specific heat at constant volume c_v equals $c_v^\infty + c'$.

Oxygen, nitrogen, and carbon dioxide all have rotational degrees of freedom and hence a rotational specific heat c' for rotational relaxation equal to the universal gas constant R. Water vapor has three rotational degrees of freedom and a rotational specific heat of 3R/2; however, since the mole fraction of water vapor in air is at most 8% in the temperature range 273–313 K[44], the rotational specific heat of air can be closely approximated by neglecting the contribution of water vapor.

Each rotational energy level relaxes with a different relaxation time, and the resultant

absorption as a function of frequency can be quite complex. However, the principal constituents of air (oxygen, nitrogen, and water vapor) have rotational energy levels that are closer together than the average thermal energy. As a result, the rotational relaxation process for air behaves as though the rotational energy levels were continuous and can be described by a single relaxation time for isometric and adiabatic conditions $\tau'_{vs,\,rot}$. At all frequencies less than 10 MHz, $\tau'_{vs\ rot}$ is $\ll \omega^{-1}$. For this case, and the value of $\delta K_s/K_s^\infty$ discussed above, with $c_v = c_v^\infty + c' = c_v^\infty + R$ and $R/c_v = \gamma - 1$, Eq (59) becomes

$$\alpha_{rot} = \{[\pi R(\gamma - 1)]/(cc_p^\infty)\}(f^2/f_{r,rot}) \tag{60}$$

in Np m^{-1}, where

$$f_{r,rot} = 1/(2\pi\tau'_{vs,rot}) \tag{61}$$

in Hertz in the rotational relaxation frequency.

Now consider again, Eq. (59). Write

$$\alpha = f^2 \left(\frac{c}{co^2}\right)\left(\frac{c_\infty^2 - co^2}{c_\infty^2}\right)\frac{2\pi^2\tau_{ps}}{1 + w^2\tau_{ps}^2}. \tag{62}$$

Experimentally, one finds that τ for rotational relaxation is very small so that below 10 MHz, $\omega^2\tau_{ps}^2 \approx 0$. Then

$$\alpha = \text{const} \cdot f^2 \cdot \tau_{ps}. \tag{63}$$

If we write $\tau_{ps} = \text{const} \cdot Z_{rot}/P$, where Z_{rot} is the rotational collision number, we find that

$$\alpha_{rot} = [2\pi^2 f^2/(\gamma Pc)]\mu[\gamma(\gamma - 1)R/(1.25c_p^0)]Z_{rot} \tag{64}$$

in Np m^{-1}. Again, assuming that $\gamma = 1.4$ and $c_p^0/R = 7/2$, the term in brackets becomes 0.128, a value which is independent of temperature since the specific heat of rotation does not change with temperature. Comparing Eqs. (34) and (64), we can now write

$$\alpha_{rot}/\alpha_{cl} = 0.128Z_{rot}/1.88 = 0.0681Z_{rot} \tag{65}$$

For dry air, the rotational collision number Z_{rot} has been measured near room temperature by Greenspan[45] and at higher temperatures by Bass and Keeton[46]. A summary of these experimental results can be written in the form

$$Z_{rot} = 61.1 \exp(-16.8T^{1/3}) \tag{66}$$

over the temperature range 293–690 K.

Acknowledging the presence of water vapor but ignoring the relaxation of the small water–vapor rotational specific heat, the rotational collision number can be written as

$$Z_{rot} = \{[X(N_2 + O_2)/Z_{rot}(N_2 + O_2) + [X(H_2O)/Z_{rot}(N_2 + O_2 + H_2O)]\}^{-1} \tag{67}$$

where $X(N_2 + 0_2)$ is the mole fraction of nitrogen plus the mole fraction of oxygen; Z_{rot} $(N_2 + 0_2)$ is the rotational collision number for dry air; $X(H_20)$ is the mole fraction of

water; and Z_{rot} ($N_2 + O_2 + H_20$) is the number of H_20 collisions required for N_2 and O_2 to establish rotational equilibrium. This latter quantity can take on values from ∞ to 1 (probability 0–1). The resultant Z_{rot} will be most dependent on $X(H_20)$ if $Z_{rot}(N_2 + O_2 + H_20) = 1$. Since Z_{rot} ($N_2 + O_2$) $\cong$ 5, when $X(H_20) < 0.02$, the rotational collision number for the mixture can change by no more than 2%, so $Z_{rot} \cong (N_2 + O_2)$; hence, we are justified in ignoring the effect of water vapor on the rotational collision number.

Combining Eqs. (65) and (66) with Eq. (58) gives the combined absorption due to classical absorption factors and rotational relaxation α_{cr} as

$$\alpha_{cr} = 5.578 \times 10^{-9} \frac{T/T_0}{T + 110.4} \frac{[1 + 4.16 \exp(-16.8T^{-1/3})] f^2}{P/P_0} \tag{68}$$

Evaluating Eq. (68) for various temperatures indicates that a simplified empirical equation of the form

$$\alpha_{cr} = 1.83 \times 10^{-11} \frac{(T/T_0)^{1/2} f^2}{P/P_0} \tag{69}$$

in Np m^{-1}, is within 2% of Eq. (33) for temperatures between 213 and 373 K.

The relaxation strengths of the two processes important in air are near to those one would expect for a simple single relaxation of N_2 and O_2. With $\lambda = c/f$ and $\omega = 2\pi f$, Eq. (59) can be written as

$$\alpha_{vib,j} = \frac{\pi s_j}{c} \frac{f^2/f_{r,j}}{1 + (f/f_{r,j})^2} \tag{70}$$

in Np m^{-1}, where j = O or N; $s_j = c_j' R/[c_p^\infty(c_p^\infty + c_j')]$; and $f_{r,j} = 1/(2\pi\tau_{vs,j})$. The relaxation strength s_j can be related to the particular atmospheric constituent and the temperature by using the Planck–Einstein relation[47],

$$c_j'/R = \frac{X_j(\theta_j/T)^2 e^{-\theta_j/T}}{(1 - e^{-\theta_j/T})^2} \tag{71}$$

where X_j is the mole fraction of the component considered; i.e., 0.20948 for oxygen and 0.78084 for nitrogen[42]; θ_j is the characteristic vibrational temperature (2239.1 K for oxygen and 3352.0 K for nitrogen). For the temperature range 0–40°C, c_j' is small with respect to c_p^∞, and thus $c_p^\infty \cong c_p$ and $c_v^\infty + c_j' = c_v$. Therefore, $s_j \cong (c_j'/R)(R^2/c_p c_v)$. Using the same reasoning that led to the approximations described in developing Eq. (13), the quantity $(R^2/c_p c_v)$ can be set equal to $(\gamma - 1)R/c_p$, which, with $\gamma c_v = c_p$, was shown to be equal to 4/35. Thus, with Eq. (71), Eq. (70) can be written as

$$\alpha_{vib,j} = \frac{4\pi X_j}{35c} \frac{(\theta_j/T)^2 e^{-\theta_j/T}}{(1 - e^{-\theta_j/T})^2} \frac{f^2/f_{r,j}}{1 + (f/f_{r,j})^2} \tag{72}$$

in Np m^{-1}. Using Eq. (56), the absorption due to vibrational relaxation can be computed if the $f_{r,j}$ frequencies for oxygen and nitrogen are known.

The frequencies of maximum absorption for oxygen and nitrogen vibrational relaxation have been computed by using the general theory described earlier [48] (Evans et al., 1972). To a good approximation,

$$f_{r,O} = (P/P_0)\{24 + 4.41 \times 10^4 h[(0.05 + h)/(0.391 + h)]\} \tag{73}$$

in Hertz, where h is the mole fraction of water vapor in percent. The various constants in Eq. (73) can be determined from the general theory or from experimental measurements in air.

The frequency of maximum absorption for nitrogen is theoretically less difficult to determine than that for oxygen, since it is dominated by direct vibration–translation (V–T) deexcitation of the excited nitrogen molecules by water vapor (or a one–step V–V transfer to H_2O). Carbon dioxide provides an alternate relaxation path at very low water–vapor concentrations. The form of the relaxation frequency is

$$f_{r,N} = (P/P_0)(T/T_0)^{-1/2}[9 + 350h \exp\{-6.142[T/T_0)^{-1/3} - 1]\}] \quad (74)$$

in Hertz.

An alternative model for nitrogen relaxation gives

$$f_{r,N} = (P/P_0)(9 + 200h) \quad \text{(Hz)} \quad (75)$$

The contribution to atmospheric absorption due to vibration relaxation processes can now be determined from Eqs. (72)–(75) if the percent mole fraction of water vapor is known. By Avagadro's law, the percent mole fraction is equal to the ratio of the partial pressure of water vapor P_w to the atmospheric pressure P of the sample volume of moist air, in percent. Thus,

$$h = 100P_w/P \quad (76)$$

in percent. Introducing the saturation vapor pressure of pure water over liquid water P_{sat}, Eq. (76) can be written as

$$h = (100P_w/P_{sat})(P_{sat}/P) \tag{77}$$

or

$$h = h_r(P_{sat}/P) \tag{78}$$

in percent, where h_r is by definition the relative humidity for a given sample of moist air under pressure P and at a temperature T, in percent, a quantity that is usually available from experiments. For convenience, Eq. (78) is written in terms of the standard atmospheric pressure P_0 as

$$h = \frac{h_r(P_{sat}/P_0)}{P_0} \tag{79}$$

in percent. The quality P_{sat} can be determined from standard reference or computed.

If we now combine forms, we find that

$$\begin{aligned}\log_{10}(P_{sat}/P_0) = {} & 10.79586[1 - (T_{01}/T)] \\ & - 5.02808 \log_{10}(T/T_{01}) \\ & + 1.50474 \times 10^{-4}(1 - 10^{-8.29692[(T/T_{01}) - 1]}) \\ & + 0.42873 \times 10^{-3}(10^{4.76955[1 - T_{01}/T)]} - 1) \\ & - 2.2195983\end{aligned} \tag{80}$$

where f is the acoustic frequency in Hz, p_s is the atmospheric pressure, p_{so} is the reference atmospheric pressure (1 atm), T is the atmospheric temperature in K, T_0 is the

reference atmospheric temperature (293.15 K), $f_{r,O}$ is the relaxation frequency of molecular oxygen and $f_{r,N}$ is the relaxation frequency of molecular nitrogen.

Figure 9a. shows the relative contributions of the different relaxation mechanisms. Figure 9b. shows comparison between theory and experiment.

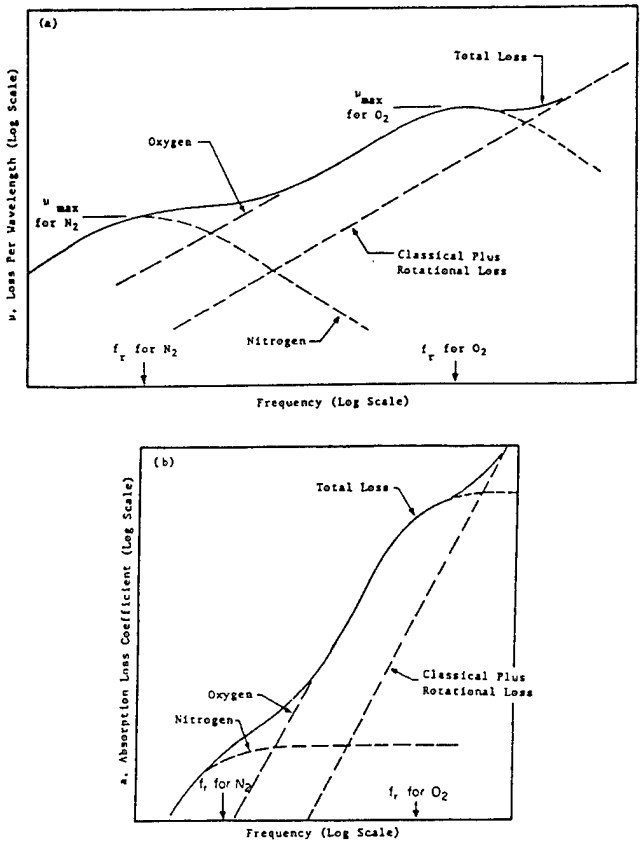


Fig. 9a. Components and general behavior of total air absorption in air in terms of (a) loss per wavelength and (b) loss per unit distance: (a) $\mu \sim \frac{2(f/f_r)}{1+(f/f_r)^2}$, (b) $\alpha \sim \frac{(f/f_r)^2}{1+(f/f_r)^2}$

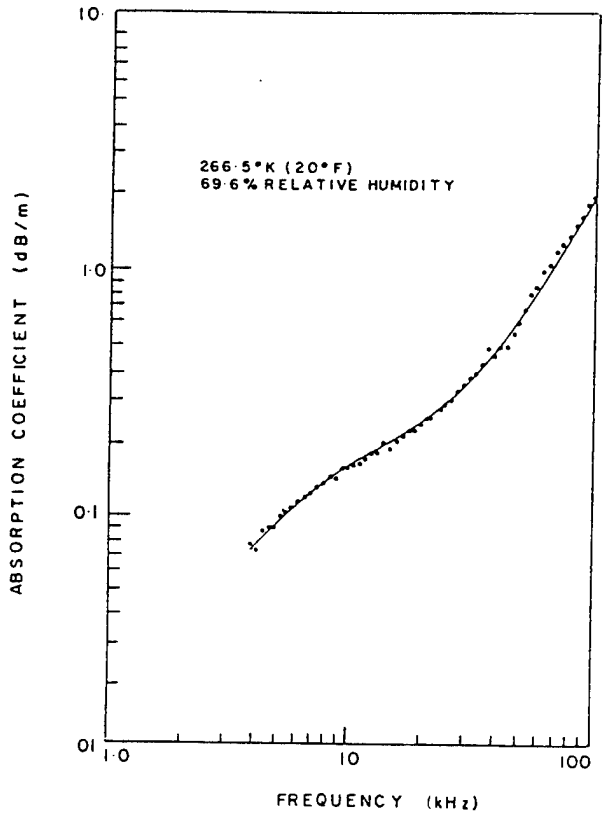


Figure 9b. Total free–field sound absorption in air at 266.5° K and 69.9% relative humidity; points represent experimental data; solid line calculated using the computational technique of Ref. 12.

Hopefully you will agree that the results are pretty impressive.

To this point, I have been emphasizing acoustic studies of the transfer of energy between molecules. We have learned a lot about molecular collision this way and will learn more in the future. During our final session, I would like to touch on some less developed aspects of the subject.

3.0 Classical Absorption at Low Pressure

3.1 Diffusion

There is yet another source of sound absorption which must be considered in gas mixtures. When there is local pressure or temperature gradient, less massive molecules with their higher thermal speeds move toward a condition of equilibrium more rapidly than heavier molecules. The diffusion due to the pressure gradient is accompanied by preferential diffusion of the lighter molecules due to the thermal gradient. The result is an additional absorption due to this diffusion given by[49]

$$\alpha = \frac{2\pi^2 f^2 \gamma X_1 X_2 P D_{12}}{P c^3}\left[\frac{M_2 - M_1}{M} + \frac{(\gamma - 1)k_T}{\gamma D_{12} X_1 X_2}\right]^2 \tag{81}$$

in Np m^{-1}, where X_1, X_2 and M_1, M_2 are the mole fractions and molecular weights of gases 1 and 2, respectively; M is the molecular weight of the mixture of gases 1 and 2; D_{12} is the concentration diffusion coefficient and the product of pressure P and is in units of N sec^{-1}; and k_T is the thermal diffusion coefficient in m^2 sec^{-1}.

The calculation of absorption due to the combination of pressure and thermal diffusion terms requires that PD_{12} and the ratio k_T/D_{12} be known. These quantities can be calculated from kinetic theory[41], but experimental values are sparse. For air, the term PD_{12} is largest when considering O_2/H_2O collisions; however, in this case, not only is the product X_1X_2 small, but the ratio $k_T/(X_1X_2)$ is also small. For O_2/N_2 collisions, X_1 and X_2 are both relatively large in air, but the differences in molecular weights ($M_2 - M_1$) and kT are both small. Ignoring relaxation, Bauer[50] has shown that for air, 99.5% of the total classical absorption can be attributed to the viscosity and thermal

conduction. Hence, additional absorption due to diffusion need not be considered for air. For mixtures where $M_2 - M_1$ is large, however, this situation could change dramatically.

3.2 Sound Propagation in Gases with Large Mass Disparity

When M_2 and M_1 differ by a large amount, there arises the possibility for added affects. This is an unexplored area and I will draw upon analogies to make my argument.

Perhaps you are familiar with work explaining the interaction of sound waves with the surface of the earth. Sound causes air to pulsate in the pores of the ground. There is viscous drag at the pore walls as well as direct momentum transfer if the pores are not vertical. The result is energy transfer to the matrix or frame. Acoustic waves in the frame travel with much different speed than that in the pores so there are effectively two propagation speeds allowed. In practice, a buried sensor will be affected to some extent by both waves. This has been observed. The first description of this type was developed by Biot.[51]

Now suppose that we fill the soil with water and conduct our experiment at the bottom of the ocean. Same principle applies but now the fluid is water. The Biot model has also been successfully applied to this case.

Now suppose that the ocean floor is not consolidated. A frame exists only to the extent that gravity causes the particles to be in contact. Even then, one can envision a frame wave transmitted through the contact points. So our picture should still be valid– two modes are still allowed.

But how far can we carry this model? What if the ocean floor is a suspension? This corresponds to what is often referred to as dusty gases. These are very massive

molecules but transfer of energy during collisions with other large particles is not very effective. This regime is common for aerosols in air or real ocean floors but a good description and experimental data remains elusive.

3.3 Boltzmann's Equations for Dilute Gases

This lecture started with a mathematical description of the wave equation which has as a fundamental basis the Navier–Stokes Equation. The Navier–Stokes Equation provides a good link between stress and strain in the limit of high gas density. More specifically, the results described in the first lecture are adequate when the mean distance between collision is very small compared to a wavelength. When this condition is no longer met, treating the gas as a continuous media is no longer accurate. One must resort to a microscopic description of molecular energy and momentum transfer even for the transfer of translational energy. In this case, the governing equation is the Boltzmann Equation.

3.3.1 Experimental Results

Sound propagation in dilute gases has been investigated by Greenspan[52] (Helium at 1 MHz; He, Ne, Ar, Kr, X at 11 MHz), Meyer and Sessler[53] (Ar at 100 and 200 kHz), and Hassler[54] (Ar at 100 and 200 kHz) (cf. the review by Greenspan). These data are plotted in Fig. 10 on the universal abscissa $\eta f/p = \omega\tau_c/2\pi$. One sees that the rare gases obey the Kirchhoff–Stokes theory for $\omega\tau_c \langle 0.2$.

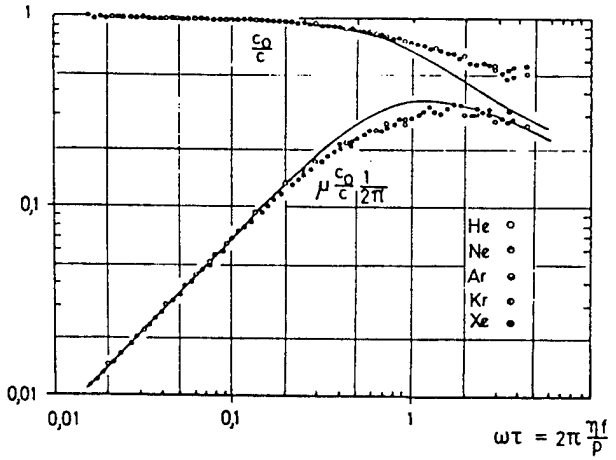


Fig. 10 Sound propagation in monatomic gases (Greenspan).–Kirchhoff–Stokes theory. By courtesy of the author and Academic Press, Inc., New York.

In Fig. 10 only those measurements which correspond to the conditions "inside" the gas are included. In dilute gases the experimental conditions are very unfavorable: the small density and the large absorption reduces the signal so much that measurements are taken more and more in the very neighborhood of the sound transmitter, i.e. at the "margin" of the gas. Here, where the molecules coming from the sound transmitter have not suffered a collision with other molecules, special phenomena show up.

3.3.2 Boltzmann's Transport Equation (From Ref. 50.)

The most detailed description of a system containing a single monatomic ideal gas is given by the distribution function f(c, r, t), f represents the number of molecules in the differential space dr at r which possess at t a velocity within dc of c. f can be understood as the density of particles in μ space. The thermodynamic variables and the irreversible fluxes are given by the moments of the distribution function. The zero

moment yields the number density

$$n = \int f dc \tag{82}$$

the first moment yields the mean (macroscopic) velocity

$$v = \int c(f/n)dc \tag{83}$$

The kinetic energy relative to a frame moving with v defines the kinetic temperature,

$$\frac{3}{2}kT = \frac{m}{2}\int (c - v)^2(f/n)dc\,. \tag{84}$$

Internal friction means transport of a component of the linear momentum $m(\underline{c}_\mu - \underline{v}_\mu)$with the speed $(c_v - v_v)$ (u, v = x, y, z). In tensor notation that establishes the pressure tensor

$$\underline{\underline{P}} = m\int (\underline{c} - \underline{v})(\underline{c} - \underline{v})fdc \tag{85}$$

The trace of Eq. (84) is the threefold static $3p = 3nkT$. The heat conduction stems from the transport of kinetic energy (relative to v) $m(c - v)^2/2$ in the direction $(c - v)$. Therefore the transport of kinetic energy is represented by q

$$\underline{q} = \frac{m}{2}\int (\underline{c} - \underline{v})(\underline{c} - \underline{v})^2 fdc \tag{86}$$

where q is the trace of the third tensorial moment. Higher moments do not possess a

particular thermodynamic meaning and therefore have no names. If we succeed in finding an equation for f(c, r, t), by insertion of the solution into Eqs. (85) and (86), the friction tensor and the heat current can be calculated. Together with the conservation equation, which also represents equations between the moments, the dispersion equation for sound propagation is obtained. For that purpose the moments may be linearized: since f contains members proportional to the sound amplitude, we can substitute (c – v) by c in Eqs. (84)–(85). In the case of Eq. (86), the same procedure would include convective energy transport.

$$q = \frac{m}{2}\int c^2\, cf\, \mathrm{d}c - \tfrac{5}{2}nkTv = \frac{m}{2}\int \left(c^2 - \frac{5}{2\beta}\right) c f\, \mathrm{d}c \tag{86a}$$

Where 5/2nkT denotes the enthalpy per unit volume.

A complete description of momentum and energy transport requires a knowledge of the distribution f. The equation for F is the Boltzmann equation. We report in the following some of its properties; more details can be gathered from the literature (e.g. Ref. 55, 56).

The local variation of the distribution function with time is effected (neglecting external fields and walls)

(a) by the flight of the molecules–when the time dt has elapsed, dr now contains molecules formerly–cdt apart;

(b) by collisions between two molecules which throw these out of or into the velocity range dc at c.

(c) by the action of external fields, which we assume to be absent.

If one neglects the duration of a collision against the mean free time (or the range of the intermolecular potential against the mean free path), the changes (a) and (b) are independent of each other.

$$\frac{\partial f}{\partial t} = -c\ \mathrm{grad} f + \left.\frac{\partial f}{\partial t}\right|_{\text{collision}} \tag{87}$$

In elastic collisions of identical particles (velocities c and c_1), the motion of the center of mass and the amount of the velocity difference is maintained, thus only that velocity difference $c - c_1 = g$ is rotated by angle X into a new direction (unit vector e'). The velocities of both particles c' and c'_1 after the collision are, therefore,

$$\begin{matrix} c' \\ c'_1 \end{matrix} = \tfrac{1}{2}(c + c_1) \pm \tfrac{1}{2}|c - c_1|e' \tag{88}$$

The probability of such rotation can be expressed in a differential cross–section $\sigma(g, X)$ which depends on the intermolecular potential. Now we can formulate the loss of molecules out of dc.

$$dc \left.\frac{\partial f}{\partial t}\right|_{\text{loss}} = dc \int g\sigma f(c) f(c_1) de'\, dc_1 \tag{89}$$

g stems from the proportionality between collision rate and velocity difference, $f(c)f(c_1)dcdc_1$ is the density of the colliding and the target molecule; the integration over the angle (de') takes into account all products of the collision, the integration over dc_1 considers all velocities of the target molecules. To calculate the gain we have to look for

collisions which force the colliding molecules into dc at c. If, in addition, the target molecules changes its velocity into dc_1 at c_1, then, for molecules without spin, we look for the "inverse" collisions, which we obtain from eqn. (88) by time reversal. Since the equations of motion of the molecules are reversible, those "inverse" collisions possess the same probability σ. By multiplication with the distribution function of colliding and target molecule at those c' and c'_1 values compatible with eqn. (88), and integration overall possibilities (de') and final velocities of the target molecules (dc_1), we obtain the production rate of molecules at c.

$$dc \left.\frac{\partial f}{\partial t}\right|_{\text{gain}} = dc \int g\sigma f(c')f(c'_1)de'\,dc_1 \tag{90}$$

The loss–gain balance is the Boltzmann equation in the absence of external fields

$$\frac{\partial f}{\partial t} + c\,\text{grad}\,f = \int g\sigma(f'f'_1 - ff_1)de'\,dc_1 \tag{91}$$

As usual, the argument of f is abbreviated, i.e., $f_1 = f(c_1)$, $f' = f(c')$. $f'f'_1$ therefore corresponds to eqn. (88). The non–linear integro–differential equation for f is highly complex, and for our purpose linearization is allowed.

Equation (91) possesses one simple (and the only exactly known) solution, the Maxwell distribution

$$f_M(c) = n\left(\frac{\beta}{\pi}\right)^{\frac{3}{2}} \exp\left(-\beta(c-v)^2\right), \qquad \beta = m/2kT \tag{92}$$

with number density n, mean velocity v, and temperature T constant in time and space. For the Maxwell distribution ($f'f'_1 - ff_1$) vanishes identically. However, there are systems with n, v, T = F(r, t) for which eqn. (92) is a good zero–order approximation. That is the case if he left–hand side of eqn. (91) is so small that a minute deviation of the distribution function from the Maxwellian, if inserted into the then non–vanishing collision integral, is sufficient to satisfy eqn. (91). Since the collision integral implies the division of the distribution function by a mean free time (in the sense of a dimensions analysis), we can think of two cases.

(a) The variations of n, T, and v in time and space are small within a mean free time or a mean free path, respectively, while the total variation may be considerable. Then we suppose the solution to be a small deviation from a local Maxwellian Eq. (92) with n, v, β = n, v, β (r, t) and have

$$f = f_M(r, t)[1 + \Phi(c, r, t)] \tag{93}$$

(b) The amplitudes of the variables n, v, T are small. Then even with "steep" gradients and "fast" variations the deviation from an absolute Maxwellian (constant in time and space) will be small.

$$f = f_0[1 + \Phi(c, r, t)] \tag{94}$$

That is the case in the propagation of small–amplitude sound. We therefore introduce Eq. (94) into Eq. (91) and obtain by neglection of terms quadratic in Φ, because $f'_0 f'_{01} = f_0 f_{01}$, the linearized Boltzmann equation

$$f_0\left\{\frac{\partial \Phi}{\partial t} + c\ \mathrm{grad}\ \Phi\right\} = \int g\sigma f_0 f_{01}[\Phi(c'_1) + \Phi(c') - \Phi(c_1) - \Phi(c)]\mathrm{d}e'\,\mathrm{d}c_1 \tag{95}$$

The collision integral on the right–hand side of eqn. (95) represents the rate of change of a deviation from the Maxwellian distribution $f_0\Phi(c)$ at the velocity c. It depends on the shape of that deviation, and consists of four parts: encounters with Maxwellian–distributed target molecules diminish an excess of the colliding molecules $f_0\Phi(c)$ at c (last term of eqn. (95); an excess in other velocity ranges consistent with eqn. (88); $\Phi(c')$; increases in the excess at c (second term). The collisions of the Maxwellian–distributed part of the colliding molecules against target molecules distributed according to $f_0(1 + \Phi)$ produce a gain (first term) and a loss (third term) at c. In general, the rate of change of the relative deviation from the Maxwellian, $\Phi(c)$, will be different at different velocities. Such relative deviations which change at every velocity with a rate proportional to $\Phi(c)$, conform to the eigenfunctions of the collision operator. They decay under retention of their shape.

1. Free–molecule propagation

We consider a region near the sound transmitter in which most of the molecules come directly from the oscillating surface without a binary collision in between. The the Boltzmann equation is

$$\frac{\partial f}{\partial t} + c_x \frac{\partial f}{\partial x} = 0 \tag{96}$$

A solution is any distribution function which contains the combination $c_x - x/t$. Special solutions are gained by the boundary conditions at the transmitter. The simplest assumption is the emission of a Maxwellian from the transmitter surface oscillating according to

$$w = \hat{w} \sin \omega t' \tag{97}$$

Then the molecules with direction to the right side (coming from the transmitter) have a distribution in the laboratory frame[58]

$$f_+(x = 0) = n_0 \left(\frac{\beta}{\pi}\right)^{\frac{3}{2}} \exp\left[-\beta(c_x - \hat{w}\sin\omega t')^2 - \beta c_y^2 - \beta c_z^2\right] \tag{98}$$

while the molecules flying to the left do not have any harmonic history, they shall be Maxwellian

$$f_-(x = 0) = n_0 \left(\frac{\beta}{\pi}\right)^{\frac{3}{2}} e^{-\beta c^2} \tag{99}$$

Half–range expansions of the form of eqs. (98) and (99), but with more complicated f_+ and other forms of accommodation at the oscillating surface have been introduced by Maidanik et al.[57] For small amplitudes, $\hat{w}$, eqn. (98) may be linearized

$$f_+(x = 0) = n_0 \left(\frac{\beta}{\pi}\right)^{\frac{3}{2}} \exp\left[-\beta c_x^2 - 2\beta\hat{w}c_x \sin\omega t' - \beta c_y^2 - \beta c_z^2\right] \tag{100}$$

or

$$f(x = 0) = f_0[1 + U(c_x)2\beta\hat{w}c_x \sin\omega t'] \tag{101}$$

U(x) is the unit step. The propagator $\delta(t' - t + x/c_x)$ moves the distribution function

(100) from the point x = 0 to x.

$$f = f_0[1 + U(c_x)2\beta\hat{w}c_x e^{j\omega(t-x/c_x)}] \tag{102}$$

The received pressure is the flux of linear momentum per unit time and area, if we assume full accommodation again. It contains a part oscillating with ω, the amplitude of which is

$$\Delta P_{xx} = 2\beta\hat{w}\, e^{j\omega t}\int f_0 c_x^3 e^{-j\omega x/c_x} dc = 2\beta\hat{w}\, e^{j\omega t} n_0 \left(\frac{\beta}{\pi}\right)^{\frac{3}{2}} \int_0^\infty e^{-\beta c_x^2} c_x^3 e^{-j\omega x/c_x} dc_x \tag{103}$$

Higher harmonics, expected by Kahn and Mintzer[58], are not present. The integral (103) cannot be solved in closed form. First we discuss the situation close to the transmitter.

(a) Short distance to the transmitter

For small values of x, the factor $\exp(-j\omega x/c_x)$ can be expanded for almost all molecules except the very slow ones. The integration can then be performed. If we force the result into an exponential representation again, we obtain (with the mean velocity $c_m = (4/\pi\beta)^{\frac{1}{2}}$)

$$\frac{\Delta P_{xx}}{p} \approx \frac{\sqrt{2}}{c_m}\hat{w}\, e^{j\omega t} \exp\left[-j\frac{\omega x}{c_m} + \left(\frac{1}{2} - \frac{2}{\pi}\right)\frac{\omega^2 x^2}{c_m^2} + j\left(\frac{1}{3} - \frac{2}{3\pi}\right)\frac{\omega^3 x^3}{c_m^3} + \cdots\right] \tag{104}$$

The phase is no longer a linear function of the distance, from the phase term

$$\exp\left(-j\int_0^x k(x')dx'\right) = \exp\left[-j\frac{\omega x}{c_m} + \left(\frac{1}{2} - \frac{2}{\pi}\right)\frac{\omega^3 x^2}{c_m^2} + j\left(\frac{1}{3} - \frac{2}{3\pi}\right)\frac{\omega^2 x^3}{c_m^3} + \cdots\right] \tag{105}$$

we gain differentiation as

$$k(x) = \frac{\omega}{c_m}\left[1 - j\left(\frac{4}{\pi} - 1\right)\frac{\omega x}{c_m} - \left(1 - \frac{2}{\pi}\right)\frac{\omega^2 x^2}{c_m^2} + \cdots\right] \tag{106}$$

This result is obvious. The phase velocity is at first the mean velocity of the molecules. It increases with increasing distance, and we notice an absorption proportional to the distance. That behavior results from the fact that in the region $\omega x/c_x \ll 1$ almost all molecules contribute to the propagation, while for increasing x the phase factor $\exp(-j\omega x/c_x)$ oscillates already for the slow molecules. Fewer molecules support the propagation, and these are faster in the mean.

(b) General solution

Because

$$\Delta P_{xx} = \Delta \hat{P}_{xx} e^{j\omega t} \exp\left(-j\int_0^x k(x')dx'\right) \tag{107}$$

we obtain k(x) from Eq. (104) by

$$k(x) = j\frac{\partial}{\partial x} \ln P_{xx} = \int_0^\infty c_x^2 e^{-\beta c_x^2} e^{-j\omega x/c_x} dc_x \Big/ \int_0^\infty c_x^3 e^{-\beta c_x^2} e^{-j\omega x/c_x} dc_x \tag{108}$$

A relation similar to Eq. (108) but with modified accommodation[53] and without the factor c_X^2 resulting from the weighting of the momentum transport, has been given by Meyer and Sessler[53] and solved by expansion techniques. The real part has also been plotted in Fig. 9; it deviates from eqn. (108) only for small distances. Apparently

variation of the accommodation plays a role only for small distances.

There are two standard methods of solution of the linearized Boltzmann equation and its application to the problem of sound propagation in rarefied gases. These various methods differ considerably in the path of the calculation, but they all yield the same class of solutions, the so–called normal solutions, and the same expressions for the pressure tensor $P_{\mu\nu}$ and the heat current q_ν Bauer discusses them in the Chapman–Enskog version, and applies them to the Boltzmann equation linearized about an absolute Maxwellian, eqn. (96). Then the calculation gains in simplicity, and the method becomes clear (the original method was based on the nonlinear Boltzmann equation itself or a linearization about a local Maxwellian). In principle, the method is an iteration: a lower–order approximation $\phi^{(i-1)}$is inserted into the differential part; in doing that the integro–differential equation is converted into an ordinary integral equation for a higher–order solution $\phi^{(i)}$. The dispersion relations obtained by that method and also the pertinent literature have been listed by Greenspan.[52]

4.0 Discussion

4.1 Negative Absorption

To this point, I have restricted this lecture to propagation in systems which are in equilibrium, on a scale large compared to a mean free path. Modern research is much more involved with propagation in systems far removed from equilibrium. You have heard Professor Atchley describe thermoacoustics where sound propagates through a region where there is a temperature gradient. Let's consider that case just a moment.

One way of looking at a thermoacoustic prime mover is in terms of counterpropagating plane waves. The effect of the stack in this picture is to amplify the incident acoustic wave. The wave propagates through a region of negative absorption. Rather than extracting energy from the wave, energy is added to the wave. A system in equilibrium could not do this without violating the third law of thermodynamics but when a temperature gradient is present, energy flow occurs, and acoustic gain can result. There are other cases where negative absorption has been observed. Doug Shields[59] has observed gain for a wave propagating through a gas vibrationally excited by an electric discharge. Fig. 11 shows amplitude as a function of time following an electric discharge.

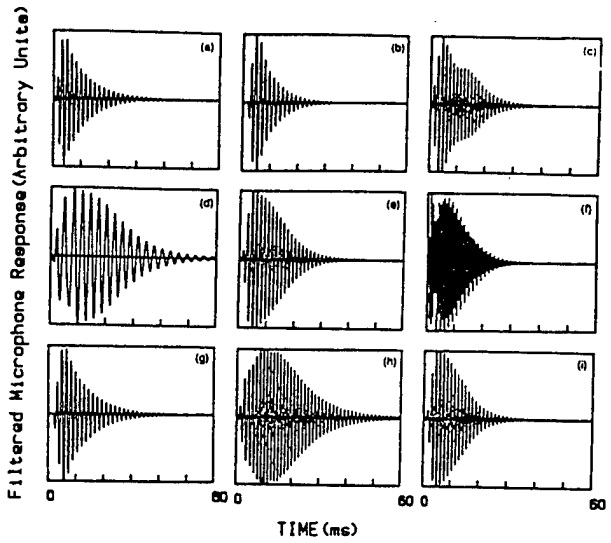


Fig. 11. Filtered microphone response. In moving from the central graph (e), only one of the four variables changes. For (c), (e), and (f) the sound frequency increases; (d) is the fundamental of the 60–cm tube; (e) is the fundamental for the 30–cm tube. Corresponding frequencies are approximately 340, 680, and 1360 Hz. For (b), (e), and (h) the pressure increases from 20 to 40 to 80 Torr. For (a), (e), and (i) and H2 concentration increases from 0 to 10% to 20%. For (g), (e), and (c) the energy per mole deposited by the discharge in the gas increases from 3700 to 7000 to 9000 J/mol. If all of this energy were to go into vibrations, the corresponding initial vibrational temperatures would be 1600, 2100, and 2400 K.

At this point in time, we can predict and observe negative gain from different mechanisms:

(a) Can you think of other mechanisms which might provide negative absorption?

(b) What can we use this for?

4.2 Electrical Discharge

One of the most dramatic sources of sound is lightning. This huge electrical discharge gives rise to thunder which can be heard for miles. Electrical discharges can also be good laboratory sources of N waves. The fundamental processes involved in electrical discharges are quite complicated and interesting.

To treat this problem, ignore coronal discharge and space discharge, and make some assumptions about the lighting being nice and symmetrical, et cetera. Then what one does is take the driving term as electrical heating of the discharge channel and numerically integrate the equations of gas dynamics. From that point of view it sounds fairly straightforward and, conceptually, it is pretty straightforward.

Energy Conservation

$$\frac{\partial \varepsilon}{\partial t} = -(p+q)\frac{\partial \frac{1}{\rho}}{\partial t} + \frac{\partial Q}{\partial t}$$

q = pseudoviscous term

$\frac{\partial Q}{\partial t}$ = rate at which energy is added to or taken from element

Conductivity of plasma

$$\sigma = e^2 n_e \lambda_e (3m_e hT)^{-1/2}$$

λ_e = electron mean free path

$$\sigma = \frac{4.173 \times 10^{-10} (A_1 + A_2) T^{-1/2}}{2 \times 10^{-15} (1 - A_1) + A_1 \langle a_i \rangle_{av}}$$

A_1 = fraction of atoms which have been ionized
A_2 = fraction of atoms which have lost second electron
$\langle a_i \rangle_{av}$ = average electron–ion cross section

$$= 2.8 \times 10^{-16} T^{-2} [(A_1 + A_2) / (A_1 + A_2)]^2 \times \log \{1.727 \times 10^{-5} [(A_1+A_2) / (A_1 +3A_2)] \Gamma (A_1\rho)^{-1/2} \}$$

Fig. 12. A few equations.

The problem comes computationally in trying to describe the conductivity of the channel as a function of time as the electric discharge progresses. There are successively greater states of ionization as the discharge progresses, and so the conductivity is a very strong function of time throughout the discharge process.

Basically one takes into account energy conservation. The p there is acoustic pressure, or pressure, whatever. q is a pseudoviscous term which hydrodynamicists

love to put in to make everything work. The desired quantity is the conductivity of the plasma that results from the discharge as a function of time.

In order to get the conductivity we need an accurate equation of state. One which involves all the ionized species that are available in the atmosphere as the temperature and current ramp up.

For air, the ionization cross sections for the different species are known. Those have been measured individually in the laboratory.

Knowing all the cross sections, one can get rate equations and from those derive an equation of state. The types of information needed are things like fractions of atoms which have been ionized, the secondly ionized fractions (of course, you start out with dissociation first, but we know dissociation rates and the energies required for that). Most of these, in the case of air, can be measured individually in a laboratory setting. That made the calculation, do–able, because those measurements are independent of the prediction of lightning and thunder itself.

Input

$I = I_0 (e^{-\alpha t} - e^{-\beta t})$ $\quad I_0 = 4 \times 10^4$ A

$\alpha = 4 \times 10^4 \text{ sec}^{-1}$

$\beta = 4 \times 10^5 \text{ sec}^{-1}$

$\left(\frac{\partial Q_j}{\partial t} \right)_{el} = \frac{E^2 \sigma_j}{\rho_j}$ where j is a zone outward from center of discharge

Losses

Thermal conduction

Radiative Energy Transport

Bremsstrahlung – Increases as $T^{1/2}$

Black–body type radiation – Increases as T^4

Fig. 13. Energy balance.

In order to drive the whole thing, one assumes some type of current input. The current input waveform that is often assumed is two exponentials. There are some logical reasons for this but the primary reason is the need to express the current mathematically and numerically in the integration process.

The equation of state depends upon the temperature and the state of ionization, which also depends on temperature. We treat a cylindrically symmetric problem, letting current flow down through the center with a radius of the current flow on the order of a fraction of a centimeter.

The differential equations are solved moving outward from the center of the discharge. Then we get the temperature and pressure change. We solve for heat generated as a function of time in each of J zones, where J is measured outward from the discharge itself.

It was found early on that there are some important loss mechanisms that must be accounted for to get decent results. These are radiation loss mechanisms. Specifically, black–body radiation from the channel can become very important when the temperature gets very high. It can become the term that limits the magnitude of the pressure pulse.

The assumed channel radius, determined photographically, is somewhat less than a centimeter, maybe a half–centimeter or so.

The conductivity as a function of time comes out of the calculation based on the equation of state. Where the conductivity is increasing very rapidly as the temperature goes up, resistance is changing very rapidly with time. So resistance and time are strongly related.

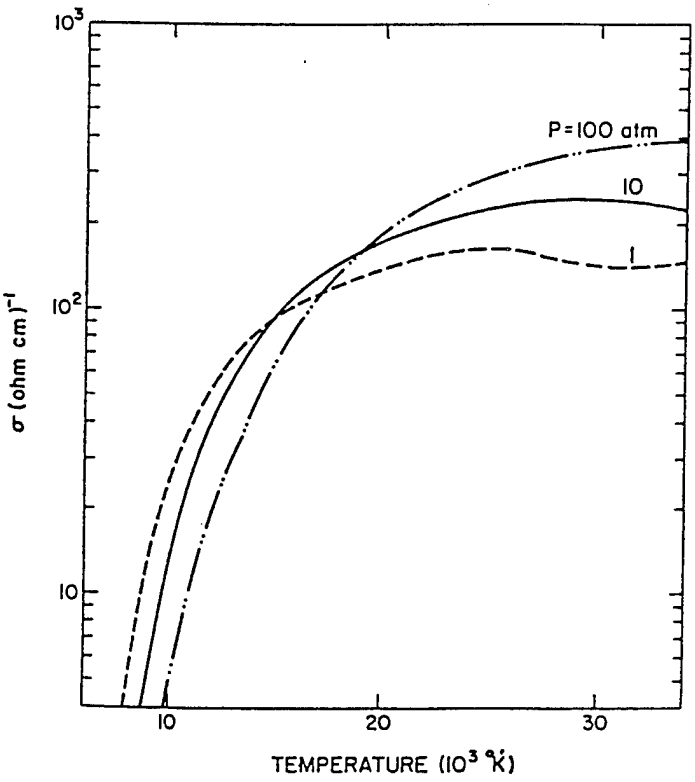


Fig. 14. Computed conductivity.

Some sampling of results. Energy for lightning is probably about 27 J/cm. Prior to these calculations by Plooster, that number was assumed to be about two orders of magnitude larger. That orders of magnitude larger assumption was based on two bad errors. One was that the thermal conduction was basically linear with temperature without limit. The other mistake was that people did not read Dave Blackstock's papers on nonlinearity. It turns out that when you start generating pressure pulses that have overpresures on the order of 10 to 15 atm, you can no longer assume that it is a linear acoustic wave as it propagates away.

As a matter interest, the acoustic energy generated goes approximately as the current to the 1.2 power.

Typical Lightning Energy ~ 27 J/cm = 2.7 kJ/m

$$E_0 \sim I_0^{1.2}$$

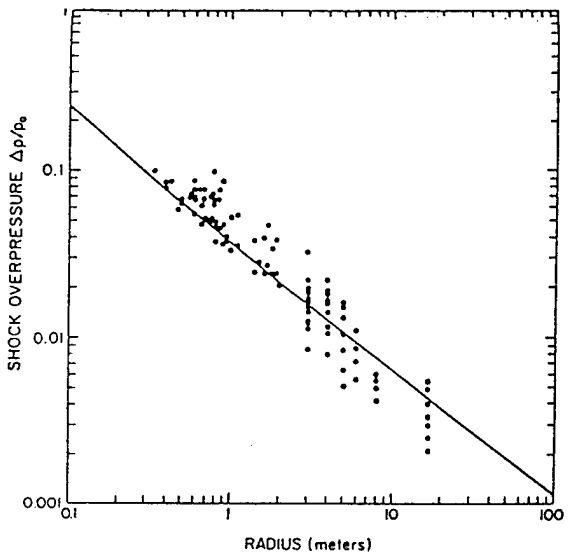


Fig. 15. Sampling of results.

That is an empirical result. The point is that you do not get a square–type relationship between acoustic energy and current.

Based upon the microscopic model, one can predict waveforms, pressure, density, et cetera, as a function of radius moving outward from the discharge channel.

References

1. G. G. Stokes, Trans. Cambridge Philos. Soc. 8, 287 (1845).

2. G. Kirchoff, Ann. Phys. Chem. 134, 177 (1868).

3. C. Kittel, Rept. Pogr. Phys. 11, 205 (1948).

4. S. P. Timon Shenko, "Theory of Elasticity," p. 9 (McGraw–Hill, New York, 1934).

5. K. F. Herzfeld and T. A. Litovitz, "Absorption and Dispersion of Ultrasonic Waves," (Academic Press, New York, 1959).

6. T. G. Winter and G. L. Hill, "High–Temperature Ultrasonic Measurements of Rotational Relaxation in Hydrogen, Deuterium, Nitrogen, and Oxygen," J. Acoust. Soc. Am. 42 (4), 848 (1967).

7. A. J. Zuckerwar and W. A. Griffin, "Resonant Tube for Measurement of Sound Absorption in Gases at Low Frequency/Pressure Ratios," J. Acoust. Soc. Am. 68 (1), 218 (1980).

8. G. L. Hill and T. G. Winter, "The Effect of Temperature on the Rotational and Vibrational Relaxation Times of some Hydrocarbons," J. Chem Phys. 49, 440 (1968).

9. L. Landau and E. Teller, "Zur Theorie der Schall Dispersion," Phy. Z. Soviet–Union 10, 34 (1936).

10. Robert T. Beyer and Stephen V. Letcher, "Physical Acoustics," Chapter 5 (Academic Press, New York, 1969).

11. F. D. Shields, "Thermal Relaxation in Fluorine," J. Acoust. Soc. Am 34 (3), 271 (1962).

12. H. E. Bass, L. C. Sutherland, Joe Piercy, and Landon Evans, "Absorption of Sound by the Atmosphere," Physical Acoustics VXII, 161 (Academic, 1984).

13. J. C. Hubbard, Phys. Rev. 38, 1011 (1931).

14. J. V. Counter, J. Acoust. Soc. Am. 30, 297 (1958).

15. E. S. Stewart and J. L. Stewart, J. Acoust. Soc. Am. 24, 194 (1952).

16. R. T. Lagemann, J. Acoust. Soc. Am. 24, 86 (1952).

17. Bradley Jacobs, Richard Carain, John R., Olson, and R. C. Aunne, J. Acoust. Soc. Am. 88, 2812–2815, (1990).

18. F. A. Angova, J. Acoust. Soc. Am. 25, 336 (1953).

19. F. D. Shields, J. Acoust. Soc. Am. 31, 248 (1959).

20. H. E. Bass, T. G. Winter, and L. B. Evans, J. Chem. Phys. 54, 644 (1971).

21. F. D. Shields, H. E. Bass, and L. N. Bolen, J. Acoust. Soc. Am. 62, 246–253 (1977).

22. Dennis Cravens, F. D. Shields, H. E. Bass and W. D. Breshears, "Vibrational Relaxation of UF_6: Ultrasonic Measurements in Mixtures with Argon and N_2," J. Chem. Phys. 71, 2797–2802 (1979).

23. F. D. Shields and R. T. Lagemann, "Tube Corrections in the Study of Sound Absorption," J. Acoust. Soc. Am. 29 (4), 470 (1957).

24. F. D. Shields and J. Faughn, "Sound Velocity and Absorption in Low–Pressure Gases Confined to Tubes of Circular Cross Section," J. Acoust. Soc. Am. 46 (1), 158 (1969). F. D. Shields, "An Acoustical Method for Determining the Thermal and Momentum Accommodation Coefficients of Gases on Solids," J. Chem. Phys. 62 (4), 1248 (1975).

25. F. D. Shields, H. E. Bass and L. N. Bolen, "Tube Method of Sound Absorption Measurements Extended to Frequencies for Above Cutoff," J. Acoust. Soc. Am. 62 (2), 346 (1977).

26. H. E. Bass and H. X. Yan, "Pulsed Spectrophone Measurements of Vibrational Energy Transfer in CO_2," J. Acoust. Soc. Am. 74 (6), 1817 (1983).

27. F. Douglas Shields, "Sound Absorption in the Halogen Gases," J. Acoust. Soc. Am 32 (2), 271 (1960).

28. J. W. L. Lewis and K. P. Lee, "Vibrational Relaxation in Carbon Dioxide/Water Vapor Mixtures," J. Acoust. Soc. Am. 38 (5), 813 (1965).

29. F. D. Shields and J. A. Burks, "Vibrational Relaxation in CO_2/D_2O Mixtures," J. Acoust. Soc. Am. 43 (3), 510 (1968).

30. F. D. Shields, "Sound Absorption and Velocity in H_2S and CO_2/H_2S Mixtures," J. Acoust. Soc. Am. 45 (2), 481 (1969).

31. F. D. Shields and G. P. Carney, "Sound Absorption in D_2S and CO_2/D_2S Mixtures," J. Acoust. Soc. Am. 47 (5) 1269 (1970).

32. H. E. Bass and F. D. Shields, "Vibrational Relaxation and Sound Absorption in O_2/H_2O Mixtures," J. Acoust. Soc. Am. 56 (3), 856 (1974).

33. B. Anderson, F. D. Shields, and H. E. Bass, "Vibrational Relaxation in SO_2/O_2 Mixtures,: J. Chem. Phys. 56 (3) 1147 (1972).

34. F. D. Shields, "Vibrational Relaxation in SO_2 and SO_2/Ar Mixtures," J. Chem. Phys. 46 (3), 1063 (1967).

35. F. D. Shields and B. Anderson, "More on Vibrational Relaxation in SO_2/Ar Mixtures," J. Chem. Phys. 55 (6) 2636 (1971).

36. F. D. Shields and K. P. Lee, "Sound Absorption and Velocity Measurements in Oxygen," J. Acoust. Soc. Am. 35 (2), 251 (1963).

37. W. T. Richards and J. A. Reid, "
," J. Chem. Phys. 2 ( ), 193 (1934).

38. J. C. Gravitt, "Sound Absorption in Carbon Disulfide Vapor as a Function of Temperature," J. Acoust. Soc. Am. 32 (5), 560 (1960).

39. K. Takagi, "Vibrational Relaxation in Liquid Carbon Disulfide," J. Acoust. Soc. Am 71, 74–77 (1982).

40. Charles H. Thompson, Stanley A. Cheyne, Henry E. Bass, and Richard Raspet, "Optoacoustic Observation of Internal Relaxation in Liquid CS_2," J. Acoust. Soc. Am. 85 (6), 2405–2409 (1989).

41. J. O. Hirschfedlder, C. F. Curtiss, and R. B. Bird, "Molecular Theory of Gases and Liquids," (Wiley, New York, 1958).

42. "U. S. Standard Atmosphere, 1962," U. S. Gov. Print. Off. Washington, D. C.

43. Society of Automotive Engineers, Committee A–21, "Standard Values of Atmospheric Absorption as a Function of Temperature and Humidity for Use in Evaluating Aircraft Flyover Noise," SAE Aerosp. Recomm. Pract. ARP 86 6 (August, 1964).

44. National Bureau of Standards, "Talks of Thermodynamic and Transport Properties of Air, Argon, Carbon Dioxide, Carbon Monoxide, Hydrogen, Nitrogen, Oxygen and Steam," (Pergamon, New York, 1960).

45. M. J. Greenspan, "Rotational Relaxation in Nitrogen, Oxygen and Air," J. Acoust. Soc. Am. 31, 155 (1959).

46. H. E. Bass and R. G. Keeton, "Ultrasonic Absorption in Air at Elevated Temperatures," J. Acoust. Soc. Am. 58, 110 (1975).

47. J. P. Holman, "Thermodynamics," 169 (McGraw–Hill, New York, 1969).

48. L. B. Evans, H. E. Bass and L. C. Sutherland, "Atmospheric Absorption of Sound: Theoretical Predictions," J. Acoust. Soc. Am. 51, 1565 (1972).

49. T. L. Cottrell and J. C. McCoubrey, "Molecular Energy Transfer in Gases," (Butterworth, London, 1961).

50. H. J. Bauer, "Influences of Transport Mechanisms on Sound Propagation in Gases," Adv. Mol. Relaxation Processes 2, 319 (1972).

51. M. A. Biot, "Theory of Propagation of Elastic in a Fluid–Saturated Porous Solid. II. Higher Frequency Range," J. Acoust. Soc. Am. 28, 179–191 (1956).

52. M. Greenspan in "Physical Acoustics IIA", pp. 1–45 (Academic Press, New York, 1965).

53. E. Meyer and G. Sessler, Z. Phys. 149, 15 (1957).

54. H. Hassler, Acoustica 20, 271 (1968).

55. L. Walsman, Handbach der Physik XII, pp. 345) Springer, Heidelberg, 1958).

56. H. Grad, Handbach der Physik XII, pp. 205–295 (Springer, Heidelberg, 1958).

57. G. Maidanik and M. Heckl, Phys. Fluids 8, 266 (1965).

58. D. Kahn and D. Mintzer, Phys. Fluids 8, 1090 (1965).

59. F. Douglas Shields, "Propagation of Sound in Vibrationally Excited N_2/ Mixtures," J. Acoust. Soc. Am. 81, 87–92 (1987).

60. M. N. Plooster, Phys. Fluids 13, 2665–2675 (1970) and M. N. Plooster, Phys. Fluids, 2124–2135 (1971).

SPEAKER

Robert Mitchel Keolian received his B.S., M.S., and Ph.D. in physics from UCLA in 1976, 1978, and 1985, respectively. His doctoral training was in acoustics, low temperature physics, and nonlinear dynamics. He then did two post-docs at Cornell University. In the first, during 1985-1987, he used ultrasonics at very low temperatures in liquid He-3. In the second, during 1987-1990, he studied the biophysics of the inner ear. Presently, he is an Assistant Professor of Physics at the U.S. Naval Postgraduate School, where he has worked in nonlinear physics, fiberoptic transducers, magnetic resonance imaging, sonar, helicopter noise, and thermoacoustics.

ACOUSTICS DEMONSTRATIONS

Abstract

Bent Tuning Curves. At small amplitudes, the resonant frequency of a mass on a spring, a pendulum, or other such oscillator, is independent of the amplitude of motion. If we sweep the drive frequency we get the familiar Lorentzian tuning curve. But at larger drives, nonlinear contributions to the restoring force change the resonant frequency as a function of the response amplitude, and the tuning curve bends to the right or the left, for hardening or softening nonlinearities, respectively. This leads to hysteresis, which will be demonstrated with the aid of a loudspeaker and a rubber band.

Parametrically Driven Oscillator. If we oscillate the stiffness or mass of an oscillator instead of directly driving it with a force, we can pump energy into the system when the oscillations are strong enough and occur close enough to double the system's resonant frequency. The response will grow exponentially with time until a nonlinearity limits the response. The tuning curves have an unusual shape and they display hysteresis as well. This will be demonstrated with a parametrically driven rigid rod pendulum.

Doubly Bent Tuning Curves. The pendulum has a softening nonlinearity and its tuning curve initially bends to the left. But if the pendulum should hit a hard stop at larger amplitudes, giving a hardening nonlinearity, the tuning curve bends back on itself and goes off to the right.

Parametric Stabilization. With really large drive levels, with accelerations large compared to g, it's possible to make a pendulum defy gravity and stand upside down.

Velocity Feedback and Self Maintained Oscillators. Many of the properties of the simple harmonic oscillator as well as self maintained oscillations (negative damping) and the active damping of vibrations (positive damping) can be shown with a tuning fork instrumented with an electromechanical drive and pickup. The drive and pickup consist of fixed coils of wire taken from relays and small but powerful rare earth magnets on the tuning fork tines. An eraser between the tines can be used to vary the passive Q. By amplifying the signal from the pickup coil, compensating for inductively induced phase shifts, and applying this signal back to the drive coil, a force proportional to the velocity of the tines can be used to subtract from or add to the passive damping, resulting in a self excited tuning fork, a tuning fork with infinite Q, or one with very low Q.

Phase Locking. A self maintained oscillator, such as a beating heart, influenced by another oscillator, such as a pacemaker, may adjust its frequency so that the phase between two oscillators is constant, thus making their frequencies identical. In general, the two oscillators will synchronize if the coupling between them is strong enough and their natural

frequencies are close enough. This phase locking or entrainment phenomenon is fairly common. Two clocks on the same wall can force each other to tick in unison, the moon shows us only one face because its spin and orbital frequencies are locked, two people walking together unconsciously adjust their pace until their feet hit the ground simultaneously, and our sleep-wake cycle is synchronized with the rotations of the earth. Here, phase locking will be demonstrated with an organ pipe and a loudspeaker.

Shock Waves and Sound Eating Sound. With large enough amplitude, the propagation of sound becomes nonlinear; superposition breaks down and sound interacts with sound. Shock waves, N waves, and the interaction of sound with noise will be shown in a long nonlinear acoustic wave tube.

REFERENCES

I find A.B. Pippard's book, *The Physics of Vibration,* (Cambridge University Press, Vol. 1, 1st ed., 1978; or the omnibus edition, 1989), chapters 9-12, to be a great reference for nonlinear oscillators, parametrics, and phase locking.

Landau and Lifshitz's *Mechanics* (Pergamon Press, 3rd ed., 1976) is also useful.

Resonant Ultrasound Spectroscopy

Albert Migliori
Los Alamos National Laboratory, MS K764, Los Alamos, New Mexico 87545

The use of mechanical resonances to test properties of materials is perhaps older than the industrial revolution. Early documented cases of British railroad engineers tapping the wheels of a train and using the sound to detect cracks perhaps mark the first real use of resonances to test the integrity of high-performance alloys. Attempts were made in the following years to understand the resonances of solids mathematically, based on the shape and composition. But Nobel Laureate Lord Rayleigh best summarized the state of affairs in 1894, stating "the problem has, for the most part, resisted attack". More recently, modern computers and electronics have enabled Anderson and co-workers with their work on minerals, and our work at Los Alamos on new materials and manufactured components to advance the use of resonances to a precision non-destructive testing tool that makes anisotropic modulus measurements, defect detection and geometry error detection routine. The result is that resonances can achieve the highest absolute accuracy for any dynamic modulus measurement technique, can be used on the smallest samples, and can also enable detection of errors in certain classes of precision manufactured components faster and more accurately than any other technique.

The mechanical resonances of a freely suspended solid object are special solutions to the equations of motion in the absence of energy loss mechanisms that depend only on the density, elastic moduli and shape. It is these resonances that will be discussed here. These modes include all the possible frequencies at which such an object would "ring" at if struck. Because a solid with N atoms in it has 6N degrees of freedom, there are 6N-6 resonances (we remove 6 frequencies that correspond to 3 rigid rotations and three rigid translations). Most of these resonances cannot be detected as individual modes because dissipation in the solid broadens the higher-frequency resonances so that they overlap to form a continuum response. For a typical solid object, of the 10^{24} modes possible, perhaps 10^{4} or so are very special because they are isolated from other modes and hence can be individually studied. These special modes or resonances are not uniquely determined in the sense that there are many different solid objects that can produce identical resonance spectra. However, the information content remains important and extensive. For example, if the lowest 50 or so resonances of a single crystal solid of known shape and density are measured, even if the solid is orthorhombic with nine separate elastic moduli, all the moduli can be determined uniquely with unprecedented absolute accuracy in samples as small as 0.5mm on a side. Or, consider a solid object with nearly perfect cylindrical symmetry and constructed of an isotropic material such as a cylindrical roller bearing element. The cylindrical symmetry produces many groups of measurable modes that should be degenerate. Deviations from perfect cylindrical symmetry of as little as 1 part in 10^{6} break the degeneracy to produce multiple resonance peaks where only one should be. The measurement of one such set of modes can detect this tiny cylindricity error in less than 1 second in a 1 cm diameter bearing. The means to perform such powerful measurements is possible today because recent advances in electronic instrumentation,

transducers, and computational techniques have replaced the previous century's railroad engineer and his practiced ear for a dull ring produced by a cracked train wheel with precision, quantitative and reliable measurement systems. We describe here the present state-of-the-art for modern mechanical resonance measurement processes, loosely called resonant ultrasound spectroscopy or RUS.

NORTH-HOLLAND
PHYSICS
PUBLISHING
N·H

Resonant ultrasound spectroscopic techniques for measurement of the elastic moduli of solids

A. Migliori, J.L. Sarrao, William M. Visscher, T.M. Bell, Ming Lei, Z. Fisk[1] and R.G. Leisure[2]

Los Alamos National Laboratory, Los Alamos, NM, USA

Received 20 October 1992

The mechanical resonant response of a solid depends on its shape, density, elastic moduli and dissipation. We describe here instrumentation and computational methods for acquiring and analyzing the resonant ultrasound spectrum of very small (0.001 cm^3) samples as a function of temperature, and provide examples to demonstrate the power of the technique. The information acquired is in some cases comparable to that obtained from other more conventional ultrasonic measurement techniques, but one unique feature of resonant ultrasound spectroscopy (RUS) is that all moduli are determined simultaneously to very high accuracy. Thus in circumstances where high relative or absolute accuracy is required for very small crystalline or other anisotropic samples RUS can provide unique information. RUS is also sensitive to the fundamental symmetry of the object under test so that certain symmetry breaking effects are uniquely observable, and because transducers require neither couplant nor a flat surface, broken fragments of a material can be quickly screened for phase transitions and other temperature-dependent responses.

Reprinted from PHYSICA B

Physica B 183 (1993) 1–24
North-Holland

PHYSICA B

Resonant ultrasound spectroscopic techniques for measurement of the elastic moduli of solids

A. Migliori, J.L. Sarrao, William M. Visscher, T.M. Bell, Ming Lei, Z. Fisk[1] and R.G. Leisure[2]

Los Alamos National Laboratory, Los Alamos, NM, USA

Received 20 October 1992

The mechanical resonant response of a solid depends on its shape, density, elastic moduli and dissipation. We describe here instrumentation and computational methods for acquiring and analyzing the resonant ultrasound spectrum of very small (0.001 cm^3) samples as a function of temperature, and provide examples to demonstrate the power of the technique. The information acquired is in some cases comparable to that obtained from other more conventional ultrasonic measurement techniques, but one unique feature of resonant ultrasound spectroscopy (RUS) is that all moduli are determined simultaneously to very high accuracy. Thus in circumstances where high relative or absolute accuracy is required for very small crystalline or other anisotropic samples RUS can provide unique information. RUS is also sensitive to the fundamental symmetry of the object under test so that certain symmetry breaking effects are uniquely observable. and because transducers require neither couplant nor a flat surface, broken fragments of a material can be quickly screened for phase transitions and other temperature-dependent responses.

1. Introduction

Large single crystals are always highly prized, in part because of their appearance, but also because usually they are the result of considerable effort on the part of the grower. Such effort is justified because the usual implementation of many measurement techniques, for various complex and often mundane reasons, requires samples with dimensions in the centimeter range. Ultrasound measurements, traditionally of great importance because of their connection to thermodynamics, transport properties and microstructural effects, are typically subject to this size constraint. When only small samples are available, it is possible to perform pulse-echo ultrasound measurements at GHz frequencies [1] or to use the vibrating reed and related methods [2], but the cost to the scientist is either instrument complexity or loss of information. There are, of course, other nonacoustic techniques for obtaining sound velocity and attenuation data such as Brillouin scattering [3], inelastic neutron scattering [4], X-ray based methods [5] and others. Each of these nonacoustic techniques has advantages and disadvantages. Among the disadvantages common to all of them is the lack of high precision. Only the acoustic techniques can achieve 10^{-6} or better reproducibility. Because the speed of sound may vary only a percent or less at a phase transition [6] or a few percent from 300 K to 4 K, this lack of precision can be a serious failing. On the other hand, neutron scattering can provide the entire dispersion curve from Brillouin-zone center to edge, but with worse than percent accuracy, Brillouin scattering can obtain data at frequencies in the tens of GHz range with 1% accuracy on very small samples but suffers at cryogenic temperatures or with

Correspondence to: A. Migliori, Los Alamos National Laboratory, MSK 764 Los Alamos, NM 87545, USA.

[1] Also at Department of Physics and Institute for Pure and Applied Physical Sciences, University of California, San Diego, La Jolla, CA 92093, USA.

[2] Permanent address: Department of Physics, Colorado State University, Fort Collins, CO 80523, USA.

0921-4526/93/$06.00 © 1993 – Elsevier Science Publishers B.V. All rights reserved

opaque samples, and X-ray techniques achieve moderate accuracy on very small samples. Thus each of these techniques retains an important place in elasticity studies (or it would not be used, of course).

Of the acoustic techniques, such methods as the vibrating reed and torsion pendulum can measure only some of the elastic moduli, while pulse-echo ultrasound and most of the nonacoustic techniques can measure all of them. This is not a particularly important constraint for isotropic materials such as glasses, polycrystalline metals and ceramics. However, for single crystals, textured alloys and the like, the value of ultrasound measurements is often critically dependent on measurement of all moduli with both precision and accuracy.

This requirement for accuracy and precision is exemplified by ultrasonic studies of the physics surrounding second-order phase transitions. At such transitions, there are no microscopic discontinuities in the material. No atoms suddenly change position, magnetism and ferroelectricity do not suddenly appear, and electrical conductivity in a superconductor becomes infinite only for infinitesimal currents carried at zero magnetic field. However, several thermodynamic second derivatives do exhibit discontinuities. For a liquid, with only one elastic modulus, it is simple to write down the important relations, using pressure (P), volume (V) and temperature (T) instead of stress σ_{ij}, strain ε_{ij} and T. They are

$$\partial^2 \Delta G/\partial P^2 = \partial \Delta V/\partial P = -1/B\,, \tag{1}$$

$$\partial^2 \Delta G/\partial T^2 = -\partial \Delta S/\partial T = -C_p/T\,, \tag{2}$$

$$\partial^2 \Delta G/\partial P\,\partial T = \partial \Delta V/\partial T = \alpha\,, \tag{3}$$

where ΔG is the Gibb's free energy difference per unit volume across the phase boundary and is continuous, ΔV is the fractional volume discontinuity across the phase boundary, equal to zero, ΔS the entropy discontinuity, also zero, C_p is the specific heat, α is the volume thermal expansion coefficient and B the bulk modulus. Each of the quantities C_p, α and B can exhibit discontinuities at T_s, the second-order phase transition temperature at which the high-temperature (usually the so-called symmetric) phase transforms to the low-temperature (usually the unsymmetric) phase.

Discontinuities are of great importance to the experimentalist because they are often the most unambiguous of measured quantities. Moreover, in general C_p is a scalar, α a vector and, if we were to write eq. (1) for stress and strain rather than pressure and volume, we would find that the right side of eq. (1) would be a tensor. That the modulus is a tensor and is discontinuous at a second-order phase transition is a key motivating factor for the development of RUS. To see why, consider a simple soft-mode structural phase transition such as occurs in La_2CuO_4. This transition, described in more detail below, arises from a zone-edge double-well [7] potential V for one phonon branch. As the material in its tetragonal phase is cooled, the thermal excitation level drops through the point where the double well becomes important. The free energy G exhibits a single-well behavior from thermal smearing at high temperature and a double-well behavior cold. At the temperature T_s where the behavior just switches over, the phonon's frequency decreases to zero resulting in a static displacement. This static displacement increases from zero as the material is cooled further, doubling the unit cell to an orthorhombic structure. What is of most interest here is that the zone-edge static displacement, coupled with group theoretical considerations including phonon and crystal symmetry, Ginsburg–Landau theory [8] and fluctuation theory [9], forces very specific predictions about which moduli exhibit discontinuities and how big, and the temperature dependence of the moduli as T_s is approached from either direction. Such an analysis can be made for any second-order phase transition, and, if the full response of the modulus tensor is available, one can work backwards to extract much of the physics driving the transition. Without a discontinuous tensor to work from, such an analysis would be more subject to interpretational errors.

RUS can determine the full elastic tensor in a single measurement with unprecedented absolute

and relative accuracy from cryogenic to very high temperatures. The basic principle behind this simple, inexpensive bench-top measurement technique is that the mechanical resonances of a solid depend on its shape and moduli in a way sufficiently complex such that a measurement of the resonant frequencies of a carefully made sample can be used to determine the full elastic tensor. To illustrate this we show in table 1 the results of such a measurement on a Si_3N_4 ball bearing. These data represent our current state-of-the-art for accuracy, primarily because this object, a nominal 5/16 inch diameter sphere, is spherical to a few parts per million and is made from a carefully controlled ceramic with very isotropic properties. Thus only two moduli, the density and the diameter are required to fully characterize its resonances. As can be seen from the columns labelled f_m and f_r, agreement between experiment and computation is of order 0.01% after a best fit to the moduli (in this case, we use the shear modulus μ and Poisson's ratio σ as the independent parameters) is found. Density and diameter are measured independently. For a larger 1/2 inch nominal-diameter sphere of the same material, we obtain 0.004% agreement, primarily because the larger object is less perturbed by air and transducer contact and because its temperature cannot change as quickly.

2. Measurement techniques

Current practices in the design of the hardware, data analysis and sample preparation systems required to make and interpret RUS measurements have not been described in detail anywhere. Because of the novelty of the technique, and because of its utility, it is important to understand the measurement system in order to appreciate the data produced by it. Because an example is often the best focus, a state-of-the-art apparatus for making low-temperature RUS measurements on rectangular parallelepiped (RP), spherical and cylindrical samples with smallest dimension of about 0.05 cm, from 20 K to 400 K, will be described schematically. Using data on $SrTiO_3$, La_2CuO_4 and $La_{2-x}Sr_xCuO_4$ single crystals, we will illustrate what can be learned with RUS including certain effects relating to crystal symmetries not accessible by any other measurement method.

2.1. Data analysis

The key to the successful application of RUS is the ability to compute mechanical resonances from a body's shape, density and moduli. For solids, such as a sphere or RP, having a shape sufficiently simple to enable description by a few

Table 1
Resonant ultrasound measurement of a 0.63500 cm diameter Si_3N_4 ceramic sphere with a density of 3.2325 g/cm^3. f_m are measured frequencies, f_r are fitted, n is the mode number. k is our designator (to be discussed below) for the symmetry of the mode and i is in essence the harmonic number of each symmetry type. Multiple entries indicate the mode degeneracy. The fit for $\mu = 1.2374 \times 10^{12}$ dyne/cm^2 and $\sigma = 0.2703$ has a χ^2 (%) = 0.0124. This is sufficient to determine μ to about 0.01% and σ to about 0.05%. There are no corrections so these values are absolute.

n	f_r (MHz)	f_m (MHz)	% error	(k, i)
1	0.775706	0.775707	−0.000138	(6, 1), (1, 1), (4, 1), (4, 2), (7, 1)
6	0.819567	0.819983	−0.050778	(5, 1), (3, 1), (5, 2), (8, 1), (2, 1)
11	1.075664	1.075399	0.024614	(1, 2), (7, 2), (6, 2)
14	1.198616	1.198505	0.009239	(5, 3), (2, 2), (3, 2), (8, 2), (3, 3), (8, 3), (2, 3)
21	1.217375	1.217850	−0.039042	(1, 3), (6, 3), (7, 3), (1, 4), (6, 4), (7, 4), (4, 3)
28	1.440760	1.440750	0.000712	(5, 4)
29	1.527080	1.526474	0.039695	(5, 5), (8, 4), (3, 4), (5, 6), (2, 4)
34	1.558358	1.558848	−0.031448	(5, 7), (5, 8), (5, 9), (3, 5), (8, 5), (2, 5), (3, 6), (8, 6), (2, 6)
43	1.580067	1.579871	0.012426	(6, 5), (7, 5), (7, 6), (1, 5), (4, 4), (1, 6), (6, 6), (4, 5), (4, 6)

mathematical functions, an approach based on work by Holland [10], Demarest [11], Anderson et al. [12] and Ohno [13] is used. This procedure does not use finite-element methods but instead seeks stationary points of the Lagrangian for a solid with free surfaces [14], and with a nondissipative symmetric elastic tensor with no far-field effects such as those associated with ferromagnets or ferroelectrics. Far-field systems [15] and the dissipative problem [16] have been analyzed successfully, but will not be discussed here. Once the mechanical resonances are computed (the direct problem), carefully constructed fitting procedures can be used to work backwards to find moduli from resonant frequencies (the inverse problem).

The procedure for solving the direct problem for an arbitrarily shaped elastic solid with volume V, elastic tensor c_{ijkl}, density ρ, and with a free surface S begins with the Lagrangian

$$L = \int_V (\mathrm{KE} - \mathrm{PE})\, \mathrm{d}V \tag{4}$$

where the kinetic energy, KE, is given by

$$\mathrm{KE} = \tfrac{1}{2}\rho\omega^2 u_i^2 , \tag{5}$$

and the potential energy, PE, by

$$\mathrm{PE} = \tfrac{1}{2} c_{ijkl} u_{i,j} u_{k,l} . \tag{6}$$

Here u_i is the ith component of the displacement vector, the usual summation convention applies, indices following a comma denote differentiation with respect to that coordinate and the time dependence of the displacements is assumed to be $\mathrm{e}^{\mathrm{i}\omega t}$ where ω is the angular frequency and t is time.

Following Hamilton, we allow u_i to vary arbitrarily in the volume V and on the surface S ($u_i \rightarrow u_i + \delta u_i$) and calculate the variation δL in L. The result is

$$\delta L = \int_V (\text{left side of eq. (8)})_i \delta u_i \, \mathrm{d}V + \int_S (\text{left side of eq. (9)})_i \delta u_i \, \mathrm{d}S \tag{7}$$

plus higher-order terms in δu_i. The elastic wave equation is

$$\rho\omega^2 u_i + c_{ijkl} u_{k,lj} = 0 , \tag{8}$$

and the vanishing of the ith component of the surface traction vector is expressed by

$$n_j c_{ijkl} u_{k,l} = 0 \tag{9}$$

where $\{n_i\}$ is the unit outer normal to S.

Because of the arbitrariness of δu_i in V and on S, the u_i's which correspond to stationary points of L (i.e. $\delta L = 0$) must satisfy eq. (8) in V and eq. (9) on S. There are no such u_i's, of course, unless ω^2 is one of a discrete set of eigenvalues, the normal mode frequencies of free vibration of the system. This simple result makes possible the following powerful procedure for obtaining the free vibrations of an object.

Following the Rayleigh–Ritz prescription, we expand the displacement vector in a complete set of functions $\{\Phi_\lambda\}$,

$$u_i = a_{\lambda i} \Phi_\lambda , \tag{10}$$

and choose as our basis functions powers of cartesian coordinates:

$$\Phi_\lambda = x^l y^m z^n , \tag{11}$$

where $\lambda = (l, m, n)$ is the function label, a set of three nonnegative integers. After substituting eq. (10) into eq. (4), we obtain ($\boldsymbol{a}$ becomes a column vector)

$$L = \tfrac{1}{2}\omega^2 \boldsymbol{a}^{\mathrm{T}} \boldsymbol{E} \boldsymbol{a} - \tfrac{1}{2} \boldsymbol{a}^{\mathrm{T}} \boldsymbol{\Gamma} \boldsymbol{a} \tag{12}$$

where $\boldsymbol{E}$ and $\boldsymbol{\Gamma}$ are matrices whose order R is determined by the truncation condition

$$l + m + n \leq N , \tag{13}$$

with $R = 3(N+1)(N+2)(N+3)/6$. We have found that $N = 10$ gives a good compromise between computational accuracy, computing time and typical sample preparation errors consistent with data spanning the first 50 or so modes.

The matrix E has elements

$$E_{\lambda i\lambda' i'} = \delta_{ii'} \int_V \Phi_\lambda \rho \Phi_{\lambda'} \, dV \, . \tag{14}$$

If we had chosen Φ_λ to be an orthonormal set with respect to the density ρ (for example, normalized Legendre polynomials [13]), E would have been the unit matrix, simplifying subsequent manipulations. Our choice of Φ_λ, although extracting a moderate computational penalty, is more easily applied to complex shapes than an orthonormal set.

The matrix Γ has elements

$$\Gamma_{\lambda i\lambda' i'} = c_{iji'j'} \int_V \Phi_{\lambda,j} \Phi_{\lambda',j'} \, dV \, . \tag{15}$$

The volume integrals which appear here are quite tractable for many shapes if the choice (11) is made.

The expression (12) for the Lagrangian is stationary if the displacements u_i are solutions of the free-vibration problem. These solutions may be obtained by setting the derivatives of eq. (12) with respect to each of the R amplitudes $a_{i\lambda}$ equal to zero. This yields the following eigenvalue equation:

$$\omega^2 \boldsymbol{E}\boldsymbol{a} = \boldsymbol{\Gamma}\boldsymbol{a} \, . \tag{16}$$

The matrix E is symmetric and positive definite and Γ is symmetric, so a standard eigenvalue–eigenvector subroutine package (RSG in EISPACK-[17]) can be used to solve (16).

For our choice of Φ_λ the matrix elements of E and Γ are all of the form

$$f(p, q, r) = \int_V x^p y^q z^r \, dV \, , \tag{17}$$

where p, q and r are nonnegative integers. This integral can be evaluated analytically for a variety of shapes [14]. For the RP with sides $2d_1$, $2d_2$, $2d_3$, it is

$$f(p, q, r) = \frac{8 d_1^{p+1} d_2^{q+1} d_3^{r+1}}{(p+1)(q+1)(r+1)} \, . \tag{18}$$

To solve the inverse problem, the derivatives of the eigenfrequencies $f = \omega/2\pi$ (where ω^2 is an eigenvalue of eq. (16)) with respect to parameters of the sample are required. These can be obtained easily in the following way. First, differentiate eq. (16) with respect to one of the sample parameters p (an elastic constant, dimension or angle specifying the orientation of the crystallographic axes with respect to the parallelepipedal axes) to obtain

$$\partial\omega^2/\partial p \, \boldsymbol{E}\boldsymbol{a} + \omega^2 \boldsymbol{E} \, \partial\boldsymbol{a}/\partial p + a\omega^2 \, \partial\boldsymbol{E}/\partial p \, \boldsymbol{a} = \partial\boldsymbol{\Gamma}/\partial p \, \boldsymbol{a} + \boldsymbol{\Gamma} \, \partial\boldsymbol{a}/\partial p \, . \tag{19}$$

Then multiply this from the left with $\boldsymbol{a}^{\mathrm{T}}$ and compare with the transpose of eq. (16) to get

$$\partial\omega^2/\partial p = (\boldsymbol{a}^{\mathrm{T}}[\partial\boldsymbol{\Gamma}/\partial p - \omega^2 \, \partial\boldsymbol{E}/\partial p]\boldsymbol{a}) \, . \tag{20}$$

Because we have already computed the eigenvectors $\boldsymbol{a}$ and the volume integrals occurring in $\partial\boldsymbol{\Gamma}/\partial p$ and $\partial\boldsymbol{E}/\partial p$ are trivial, the computation of the derivatives represents only a minor increase in computational time.

We can speed up the calculation immensely by exploiting the symmetries $x \to -x$, $y \to -y$, $z \to -z$ that occur if the crystal is of orthorhombic or higher symmetry and the crystallographic axes are aligned with those of the sample. Then by inspection of the PE in eq. (6) we see that if u_x is characterized by a parity triplet $(-\xi, \mu, \nu)$ where

$$-\xi = (-1)^l, \quad \mu = (-1)^m, \quad \nu = (-1)^n \, ,$$

the matrix Γ only connects this u_x with u_y and u_z having the following parities:

$$\begin{aligned} u_x &: (-\xi, \mu, \nu) \, , \\ u_y &: (\xi, -\mu, \nu) \, , \\ u_z &: (\xi, \mu, -\nu) \, . \end{aligned} \tag{21}$$

Thus the matrix Γ degenerates into a block-diagonal matrix with eight blocks, each characterized by one parity triplet, say the parity of u_x. We label this parity as follows:

$$k = \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{matrix}$$

$$\begin{pmatrix} \xi \\ \mu \\ \nu \end{pmatrix} = \begin{pmatrix} + \\ + \\ + \end{pmatrix}\begin{pmatrix} + \\ + \\ - \end{pmatrix}\begin{pmatrix} + \\ - \\ + \end{pmatrix}\begin{pmatrix} + \\ - \\ - \end{pmatrix}\begin{pmatrix} - \\ + \\ + \end{pmatrix}\begin{pmatrix} - \\ + \\ - \end{pmatrix}\begin{pmatrix} - \\ - \\ + \end{pmatrix}\begin{pmatrix} - \\ - \\ - \end{pmatrix}. \tag{22}$$

The modes for each k-value (22) are uncoupled, so that the maximum order for which we need to solve the eigenvalue problem (16) and the total computational time are much reduced even though we now have 8 eigenvalue problems to solve.

Each of the 8 k-values represents a different symmetry for the displacement of the material in the object. For example a uniform translation in the x-direction will have u_x with $(l, m, n) = (0, 0, 0)$ or $k = 1$. A translation in the y-direction has $k = 7$, and a translation in the z-direction $k = 6$. Similarly, a rotation about the z-axis will have $k = 3$, one about the y-axis has $k = 2$, and one about the x-axis $k = 4$. These six special eigenvectors all have an eigenvalue of zero. One other special case occurs for $k = 5$ which has div $\boldsymbol{u} \neq 0$ after averaging over V, so it is the only k-value for which the volume oscillates.

This very fast and accurate solution to the direct problem is the key tool for solution of the inverse problem. However, the inverse problem is not at all straightforward. First note that there is no unique solution to the inverse problem because all the frequencies scale inversely with the linear dimensions of the sample and with the square root of the elastic constants. This simple scaling problem is easily dealt with, but for real data on imperfect objects, other uniqueness problems arise that are difficult to circumvent. The best procedure we have found is to begin with a 'figure of merit':

$$F = \sum_i w_i (f_i - g_i)^2 . \tag{23}$$

Here the sum is over a sufficient number of measured frequencies, w_i is a weighting factor chosen (usually either 0 or $1/g_i^2$, so that F is a measure of fractional deviation) to reflect one's degree of confidence in the measured frequency g_i (a function of signal strength and resonance width) and $f_i = \omega_i/2\pi$ is the ith calculated frequency. Note that the derivatives computed in eq. (20) are such that several resonant frequencies depend in almost exactly the same way on certain weighted sums of the c_{ij}. Thus many more than M resonances (where M is the number of parameters to be fit) must be measured for a meaningful fit to be achieved.

A systematic scheme is used to locate the minimum of F in the space of chosen parameters. This can be a 1-dimensional space if all we need to find is the compressibility of a fluid, or a 24-dimensional space if we need the 21 independent elastic constants of a triclinic crystal plus its three dimensions. The method we present here works in both these cases as well as many intermediate ones. In the process of searching for a viable minimization recipe we have tried several and settled on the Levenberg–Marquardt scheme [18] because it is relatively flexible, controllable, stable and reliable.

First we expand F in a Taylor series

$$F(\boldsymbol{x}) = F(\boldsymbol{x}_0) + (\boldsymbol{x} - \boldsymbol{x}_0)_\alpha F_{,\alpha}(\boldsymbol{x}_0) + \tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)_\alpha F_{,\alpha\beta}(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0)_\beta + \cdots . \tag{24}$$

$\boldsymbol{x}$ is the vector whose components $\{\boldsymbol{x}_a\}$, $\alpha = 1, \ldots, M$, are the parameters we need to estimate (elastic constants, dimensions and Euler angles relating crystallographic axes to sample surfaces in a misaligned sample have all been successfully determined). This expansion is valid only in a limited domain such that $\boldsymbol{x} - \boldsymbol{x}_0$ is in some sense small, thus it is important to use any available information to guess accurately $\boldsymbol{x}_0$ at the start.

If F is a minimum at $\boldsymbol{x}$, then

$$F_{,\alpha}(\boldsymbol{x}) = 0 , \quad \alpha = 1, \ldots, M . \tag{25}$$

Using, eq. (24),

$$F_{,\alpha}(\boldsymbol{x}_0) + F_{,\alpha\beta}(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0)_\beta = 0 \tag{26}$$

which, when solved iteratively for $\boldsymbol{x}$, is just Newton's method in M dimensions. The derivatives are

$$F_{,\alpha} = 2w_i(f_i - g_i)f_{i,\alpha}\,, \tag{27}$$

$$F_{,\alpha\beta} = 2w_i f_{i,\alpha} f_{i,\beta} + 2w_i(f_i - g_i)f_{i,\alpha\beta}\,. \tag{28}$$

The first derivatives $f_{i,\alpha}$ are given by eq. (20); we drop the second-derivative term $f_{i,\alpha\beta}$. One may make four arguments to justify this. First, the second term in eq. (28) is a sum over the measured frequencies; this sum will probably include about as many positive terms as negative ones, and consequently should be small. Secondly, dropping the second term will never affect the position of the minimum, only the route and speed of getting to it. Third, although $f_{i,\alpha\beta}$ can be expressed in terms of already computed eigenvectors and eigenvalues, actually evaluating it requires considerably more computer time than the evaluation of $f_{i,\alpha}$. Finally, to implement the minimization scheme one must obviously solve eq. (26) for $\boldsymbol{x}$, which may involve finding the inverse of $F_{,\alpha\beta}$, often a difficult thing to do if $F_{,\alpha\beta}$ is not positive definite (the first term in eq. (28) is positive-definite, but not the second). Following ref. [18], let

$$B_\alpha = w_i(f_i - g_i)f_{i,\alpha}\,, \tag{29}$$

$$A_{\alpha\beta} = w_i f_{i,\alpha} f_{i,\beta}\,, \tag{30}$$

and the solution of eq. (26) is

$$x_\alpha = x_{0\alpha} - A^{-1}_{\alpha\beta} B_\beta\,. \tag{31}$$

This equation is valid whenever eq. (24) is a good approximation, i.e. when x_α is close to the minimum. If not, a best guess is to move in a direction opposite to the gradient (downhill), i.e.

$$x_\alpha = x_{0\alpha} - \text{constant} * B_\alpha\,, \tag{32}$$

where the positive constant has dimensions x^2/F. $A_{\alpha\alpha}$ (no summation) has dimensions F/x^2 and is a measure of the $\alpha\alpha$th element of the F-surface curvature tensor. It therefore may be reasonably used to limit the distance moved in the αth direction in parameter space (this is important because there are many shallow local minima available to trap the solution. Such minima appear if a mode is too weak to be detected and no allowance is made for a missing mode in the group of measured frequencies, or if large steps are taken in following the gradient 'downhill'). Following Marquardt, introduce a dimensionless positive quantity Ω and replace eq. (31) with

$$x_\alpha = x_{0\alpha} - G_{\alpha\beta} B_\beta \tag{33}$$

where

$$G^{-1}_{\alpha\beta} = A_{\alpha\beta}(1 + \Omega\delta_{\alpha\beta})\,. \tag{34}$$

without a sum in eq. (34). Equation (33) is identical to eq. (31) if $\Omega = 0$ and is very much like (32) for large Ω, when $\boldsymbol{G}$ becomes nearly diagonal. By choosing a large Ω we can proceed as cautiously as we like along the M-dimensional surface F, only decreasing Ω to zero when in the neighborhood of the minimum.

If by iterating eq. (33) a number of times convergence is achieved at a point $x_{\min}$ in M-dimensional parameter space where the gradients $B_\alpha = 0$, $\alpha = 1, \ldots, M$, then F may be expanded about that point:

$$F(\boldsymbol{x}) = F(x_{\min}) + \delta x_\alpha A_{\alpha\beta}(x_{\min})\,\delta x_\beta + \cdots \tag{35}$$

where $\delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_{\min}$.

Because the curvature of F in different directions varies over as much as two orders of magnitude at the minimum, the accuracy for determining $x_{\min}$ is very different for different parameters. Recognizing that $2\boldsymbol{A}$ is just the inverse of the covariance matrix for this problem, diagonalizing it (or equivalently, $\boldsymbol{G}$) yields M eigenvectors y^μ and eigenvalues σ^2_μ. Specifically,

$$G_{\alpha\beta} y^\mu_\beta = 2\sigma^2_\mu y^\mu_\alpha\,. \tag{36}$$

In terms of these variances and eigenvectors eq. (35) becomes

$$F(\boldsymbol{x}) = F(\boldsymbol{x}_{\min}) + (\delta\boldsymbol{x}, y^\mu)^2/2\sigma^2_\mu\,. \tag{37}$$

Here $(\delta\boldsymbol{x}, y^\mu)$ is the inner product of two vectors, and because y^μ is a unit vector it is just the projection of $\delta\boldsymbol{x}$ in the y^μ direction. So eq. (37)

tells us the shape of the surface F near the minimum. The surfaces of constant F are ellipsoids in M-dimensional parameter space with semi-major axes in the $\{y^\mu\}$ directions. The lengths of the semi-major axes are given by

$$s_\mu = \sigma_\mu (2\delta F)^{1/2} \tag{38}$$

where δF is the amount by which F exceeds the minimum. In practice it is often true that one or more of the σ_μ's is quite large, meaning that a large uncertainty is attached to the corresponding linear combination of the x_i's given by $(\delta x, y^\mu)$. Thus probable errors cannot be attached easily to individual elastic constants (and/or dimensions), but only to these linear combinations of them. We estimate the error for a particular parameter x_i by examining several of these linear combinations. In this way the sharpness of the minimum for a particular parameter and thus an error estimate for that parameter may be determined. The error estimate is very sensitive to sample geometry errors including chipped corners for a RP and inhomogeneities. Such errors may make the absolute minimum for F shallow and introduce other local minima that may trap the solution in the wrong place. The effect is compounded if a mode is missed. For a measurement with one or two missing modes and a 5 μm parallelism error in 2 mm it is essentially impossible to obtain an accurate (20% errors are easy to get) value in a cubic material for c_{11} or c_{12}; however c_{44} is always reasonably accurately obtained. Typically with a 'good' fit and where not more than two modes are missed out of 30, the RMS error between fitted and measured frequencies is less than 0.1%, the solution does not get trapped in a local minimum, and a change in this error of 2% is larger than all reproducibility and other error sources occurring in the measurement. Thus an M-dimensional ellipsoid in parameter space surrounding the minimum in F with a surface corresponding to a 2% increase in χ^2 provides a realistic error estimate for determination of parameters. Using this criterion, the compressional moduli (c_{ii}, $i = 1, 3$) are determined to better than 1%, shear moduli (c_{ii}, $i = 4, 6$) to 0.02% and off-diagonal moduli to better than 3%. This way of determining errors can be tested directly by making the dimensions of the sample free parameters. To circumvent the ambiguity mentioned above, we add to eq. (23) a term $\Delta(d_1 d_2 d_3 - V/8)^2$, which, for large Δ, has the effect of fixing the volume of the RP. Using measured values as the initial guess, for good data on a sample with good geometry, the fitted dimensions are typically within 2 μm (0.1%) of the measured value, the limit of our accuracy for length measurement.

2.2. Sample preparation

As described above, one can easily fit to a local and incorrect minimum in modulus space with very large concommitant errors in parameter determination if (1) the sample geometry and properties are not consistent with the mechanical model used for the computation, (2) some resonances are missed (i.e. f_i is paired with the wrong g_i in eq. (23)) or (3) the resonant frequencies are incorrectly measured. To achieved an accurate fit the faces of a millimeter-sized RP sample must be accurate to 2 μm or better. This is easily accomplished using ground steel shims and a glass plate as shown in fig. 1. The shims, surface-ground to be 10 to 50 μm thinner than the distance between sample faces to be polished, and with edges squared up in an ordinary milling machine, are arranged as shown on a flat glass plate coated with molten wax [19] and held down with a large magnet. The X-ray oriented sample is trapped by the shims and polished on 15 μm and then 3 μm optical lapping paper [20] using an appropriate lubricant such as kerosene. That the shims can force a sample face to be either parallel or perpendicular to the glass depending on how pressure is applied to the sample as the wax cools is crucial. In addition, as the sample nears completion, the shims support the sample edges, ensuring that sharp edges and corners are produced, especially for brittle materials such as La_2CuO_4. This appears to be important both for accuracy and to minimize the number of missed modes. There is no definite way that we know of for the quantifying requisite corner and edge sharpness.

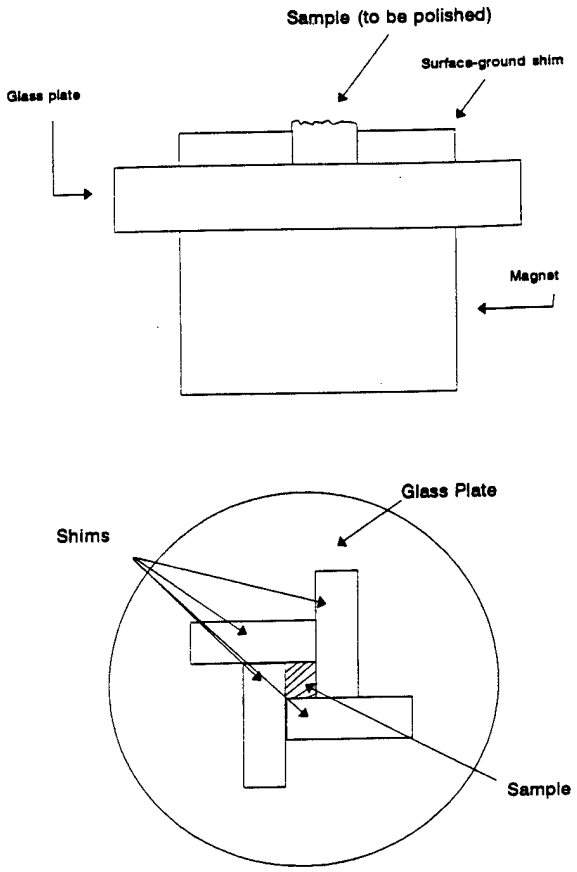


Fig. 1. Shown is the arrangement of ground steel shims on a glass plate used for obtaining an accurate rectangular parallelipiped from an as-grown oriented single crystal.

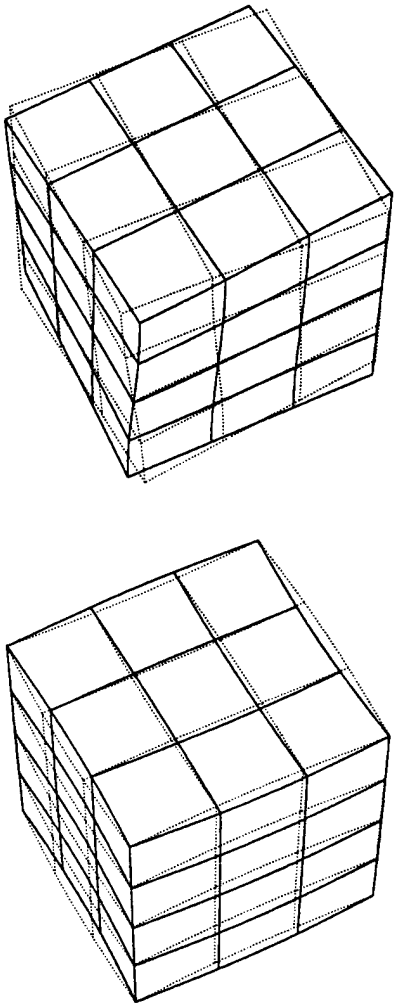


Fig. 2. Eigenvectors (local instantaneous peak displacements) for $k = 4$, a pure shear mode and $k = 5$, the only mode type for which the volume oscillates are displayed. The $k = 5$ mode illustrates how a node can occur near a corner, making it very difficult to observe that mode.

2.3. Hardware

Even with a well-prepared sample, certain modes, especially those having $k = 5$ (the only mode type where the volume oscillates) may have nodes near the corners of the sample. Computation of the mode shape is a simple addition to the codes used to find resonant frequencies, and the result for two mode types is shown in fig. 2. Because the task is to excite resonances, it is important to drive the sample at a low-symmetry location to excite as many modes as possible. The lowest symmetry point on a RP sample is the corner, thus this is the most desirable point to drive and detect, an important principle discovered by Demarest [11] and Anderson et al. [12], and derived, group-theoretically, by Mochizuki [21]. Moreover, the corners have a low mechanical impedance so that touching them with a transducer has minimal (less than a 10^{-5} fractional frequency shift) effect on the free-surface boundary conditions if the contact force is low (10^3 dynes or less). Other excitation schemes have also been used, such as electromagnetic [22] and polyvinylidene fluoride strips [23], with mixed sucess, in the sense that

the ultimate accuracy of determination of the free-sample resonances is not as good as for point contact.

Surprisingly, the lowest modes, even for an RP sample, are usually pure shear modes. Thus even if moduli and shape conspire to force a node to be near a corner, thereby making it so weak that it may be missed, the first several modes will generally determine the shear moduli to 1% or so, providing a good guess for them in the inversion calculation. The inversion code now has somewhat less work to do, making it harder to fall into a false minimum. More important, though, is that anything that helps ensure a good initial guess for the moduli should be used, including published values, because with a good starting point and the use of only the first five or ten resonances, the first pass of the code may help identify places where a mode is missing. Then a more careful scan or a remount of the sample may reveal the mode. There are other ways of finding missing modes using simple modifications to the apparatus described below [24], or by simply inserting, measuring and removing the sample several times. We cannot overstress the importance of finding nearly (95% or better) all modes before relying on the analysis of the data.

The hardware and electronics used to obtain accurate resonance data include a cryostat or furnace, transducers, a preamplifier, amplifier and mixer/filter or some equivalent. The electronic components and transducers used in our and Anderson's systems [25] are now commercially available from Quatro Corp. [26]. We will begin the hardware discussion with the transducers and work our way back to the PC-AT compatible computer used as a controller. Although other approaches will work, the one described here has been demonstrated to produce sufficiently high relative and absolute accuracy that overall errors are determined by temperature shifts and intrinsic sample preparation problems.

To measure resonances of the sample it is important to eliminate the resonant response of the apparatus, or extra modes may be observed. Because most samples in the size range of 1 mm have resonances above about 0.4 MHz, and 30 or more resonances below 3 MHz, the transducers used to excite and detect must be either damped or nonresonant in this frequency range. Damping does not work well over such a large range of frequencies or at 20 K, thus the nonresonant approach is best. However, all transducer materials have sound velocities comparable to the samples measured. One cannot get around this by using electrostatic, magnetic or optical detection schemes. The optical ones are much too noisy. Magnetic schemes, occasionally used by others [22], suffer from two serious problems. The first is that the sample must be either ferromagnetic or conducting or coated with a ferromagnetic or thick conducting layer. Even 1 μm of ferromagnetic layer can be a 0.2% perturbation on small samples, but worse still, the ferromagnet does not have a symmetric elastic tensor. Second, the coil used to drive or detect interacts mechanically with the sample via a magnetic field. Thus the usually numerous coil resonances shift and degrade the sample modes as well as perhaps introducing some new ones. This is known as 'coil disease' in NMR measurements [27]. Electrostatic systems have similar problems. The solution, applicable to direct contact (and electrostatic and magnetic drive systems as is obvious after a moment's reflection) is not to make very small, and therefore very weak transducers but to construct the transducer mostly out of single-crystal diamond. Our system [25] uses commercial 30 MHz compressional mode $LiNbO_3$ discs 1.5 mm in diameter and approximately 0.1 mm thick [28]. Such discs have a thickness mode of 30 MHz but bending modes near 180 kHz. However, using our cylinder code we know that a diamond cylinder 1.5 mm in diameter and 1.0 mm long has a lowest mode of 4.47 MHz. Thus if we bond the diamond to the transducer, the assembly has a lowest mode near 4 MHz. The diamond also acts as an inertial load, so that the response of the $LiNbO_3$ in direct contact with the sample is enhanced by the diamond behind it. We also use a Ag-coated polyimide film 25 μm thick with 1 μm of evaporated Ag as the ground plane, and a strip of this material 1.5 mm wide as a low inductance, low mechanical Q electrical contact by inserting the

strip between diamond and $LiNbO_3$. This strip also helps damp the 4 MHz and above assembly resonances. All joints are made with a nonconducting epoxy [29] because at these frequencies no direct electrical contact with the transducer is required, capacitive coupling is sufficient. The transducer assembly is shown in fig. 3.

In order to minimize transducer loading effects on the sample, we mount the transducers in a Mg machined 'tone arm' assembly, shown in fig. 4 schematically. This assembly provides of order 1 g of unbalanced mass above the transducer to provide a low contact force. The tone arm is suspended on 1 mm wide strips of Ag-coated polyimide film, the same used in the transducer. The width of the film provides a low-inductance electrical ground with excellent low-temperature

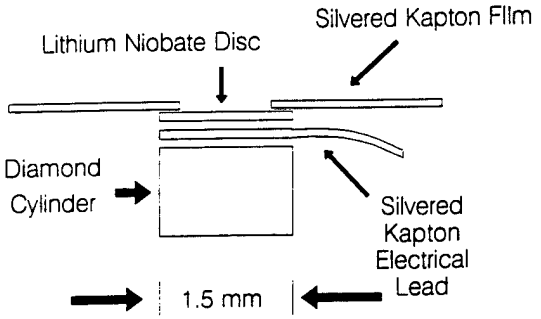


Fig. 3. Shown is a schematic of the diamond/polyimide/$LiNbO_3$ composite transducer used for all the measurements.

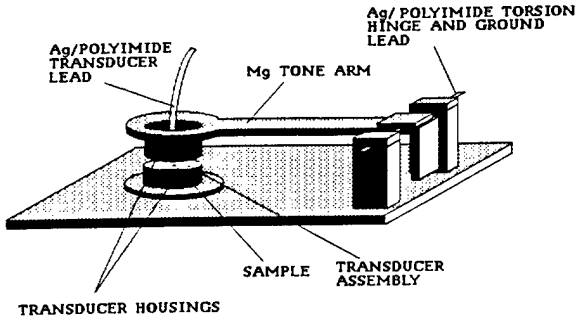


Fig. 4. The sketch here roughly illustrates how mechanical contact is made to the sample while preserving electrical shielding and maintaining a low contact force. This arrangement works well from 2 K, the lowest temperature we are set up to reach, up to the temperature at which the epoxy bonds in the transducers fail.

properties. The combination of low contact force and small, nonresonant transducers produces signals that are much weaker than those used by Ohno, Sumino and colleagues [15,30]. Their measurements were always made using a force balance that enabled resonant frequencies to be measured at successively lower contact forces. The frequencies shifted substantially (0.1%–0.5%) as force decreased, and the extrapolated asymptote was used as the zero-force frequency. With our system, even at comparable contact force, we observe less than 20 ppm frequency shifts for changes in loading from 2 g to 0.5 g. The shifts observed in refs. [15] and [30] appear to be primarily associated with high drive levels, and are absent for us. This is important because it greatly reduces both the amount of data required and the possibility of shifting the sample accidentally during a run, a problem that can cause artificial discontinuities in both frequencies and Q. Another effect appearing in Anderson's system is associated with alumina buffer rods [31]. These rods were necessary to isolate the transducers from temperatures exceeding an astonishing 2000 K, the highest temperatures ever used in a conventional ultrasound measurement system. Because the rods were long (i.e. several orders of magnitude longer than the largest sample dimension) they operated in the reverberation limit. That is, at the frequencies of interest for sample resonances, the rods themselves had such a higher mode density that the modes overlap strongly. As frequency is swept, the response of the buffer rods is convoluted with the sample response producing essentially random amplitude and phase mechanical motion, but with resonances still clearly detectable. The result is the observation of non-Lorentzian line shapes for the sample resonances, making it difficult to determine accurately either the center frequency or the Q. However, considering the temperatures reached, and with no obvious cure, Anderson was forced to use such an approach. For most other system designs, buffer rods should be avoided. A similar effect in our system is associated with the gas surrounding the sample which provides an undesirable ultrasound path between transducers in the reverberation limit. A disk of

ordinary filter paper with a small hole punched in the center, and split in half was constructed and inserted between transducer assemblies and surrounding the sample. This completely eliminates the gas path for ultrasound and thus minimizes gas resonances.

To access temperatures from 20 K to 400 K, our measurement cell is inserted into a vacuum-insulated cylinder. One end of the cylinder (or flow cryostat) is open, the other end is connected via vacuum insulated tubing (a conventional liquid-He transfer line) to the gas space above a liquid-He storage dewar. Inside the dewar, below the liquid level is a 1 kΩ, 2 W carbon resistor. This resistor is heated using an ordinary power-line-type variable autotransformer, with applied voltages up to about 30 VAC. The cold boil-off gas passes through the transfer line into the insulated cylinder and around the measurement cell. A second 50 Ω heater, constructed of 10 Ω/m cotton-insulated resistance wire (the wire is simply wadded up into a rough ball) is inserted in the gas flow path at the flow cryostat/transfer tube joint and is controlled using any commercial cryogenic temperature controller. Temperature sensing is via a silicon diode thermometer mounted inside the RUS cell within a few millimeters of the sample. This arrangement is shown schematically in fig. 5, and is capable of 20 mK temperature control.

The weak signals produced by our transducers and low contact force require the best possible signal/noise ratio (s/n) for the receiver electronics to ensure detection of as many modes as possible. The electronics design is centered around the electrical equivalent circuit for our transducer assembly over the frequencies of interest, essentially a pure 10 pF capacitor. Signals produced are in the tens of microvolt range and up. To detect such signals, two basic approaches can be taken. The one we reject is to use a broadband excitation pulse and Fourier transform the result. This is the best approach if large signals and overlapping modes are present [32]. However, to ensure that we do not miss even the weakest modes, and because the most usable samples have a mechanical Q in excess of 500, mode overlap is not a problem but s/n is. The broadband approach must have an electronic bandwidth exceeding that of the group of resonances to be measured, and must also signal average a measurement having a low duty cycle. That is, to average the signal (and noise over the receiver bandwidth) for some amount of measurement time, many excitation pulses must be generated, digitized time series taken, data

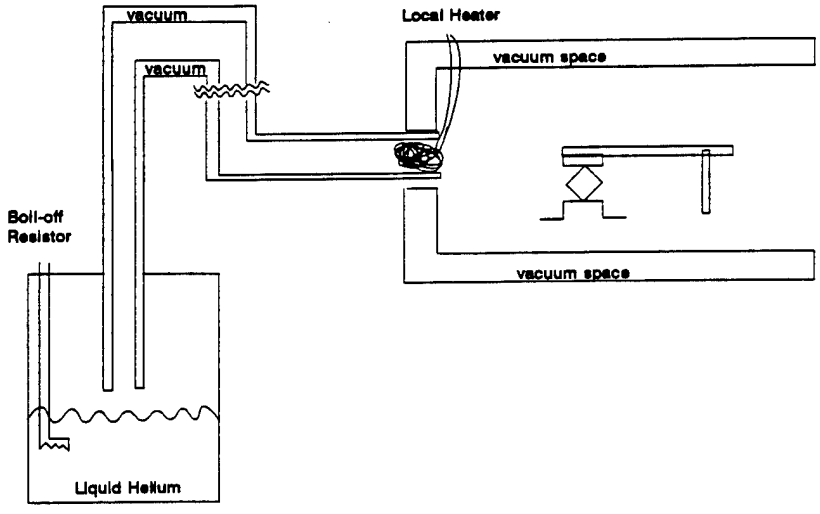


Fig. 5. Shown is the arrangement of the He storage dewar, transfer line and simple vacuum-insulated chamber that make up the flow cryostat that is so convenient for RUS.

transferred between pulses etc. Thus most of real time is spent with no signal present. Moreover, because measurements are made only in regions of frequency space where the sample exhibits mostly well-separated very sharp modes, any broadband system is acquiring much data that contains no information. Finally, for sharp, Lorentzian, well-separated modes, phase information is unnecessary. We have, therefore, chosen to use a swept sine approach based on a heterodyne receiver.

A heterodyne/swept-sine receiver (HSSR) can have an arbitrarily narrow bandwidth, has a duty cycle of unity, need only measure where resonances exist, and can signal average for arbitrarily long times a signal from which most of the noise is already eliminated using digital detection of the final, high-amplitude information-containing output of the analog section of the receiver. As with any receiver, the unavoidable noise is controlled primarily by the preamplification stage. The best approach for preamplification is to locate a JFET preamp very close to the receive transducer. In this way, the shunt effect of the capacitance (easily 100 pF) of cables connecting the 10 pF transducer to the preamp is eliminated. However, this would require a warm preamp to be located inside the flow cryostat, a complication we chose to forego. Instead, we use an accurately unity gain preamp at the end of a

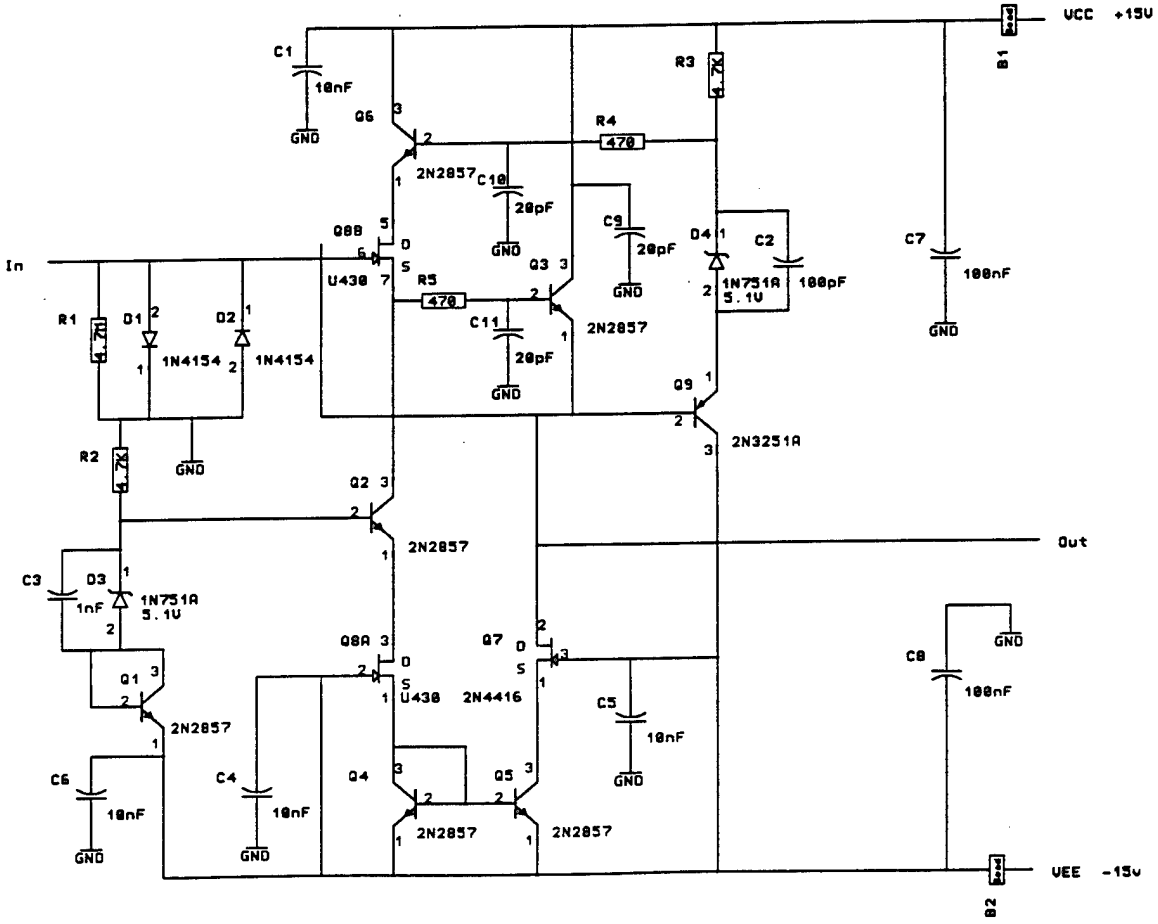


Fig. 6. We show here the schematic diagram of the low-noise unity-gain preamplifier used to bootstrap the cable capacitance of the transducer connection. The construction of this device requires very careful layout to prevent instabilities.

triaxial cable connecting preamp to transducer. By connecting the preamp output to the inner shield of the triax, we eliminate cable capacitance effects (this is known as a bootstrap or guard). This is only important because the preamp is voltage sensitive. For a current sensitive preamp or an op-amp type preamp, cable shunt capacitance has no effect on overall s/n. But for any type of preamp used with so much cable capacitance, much care must be taken to eliminate high-frequency phase shifts that could cause oscillation. This is why we have used low-inductance leads on the transducers and also why we find it easier to design a good JFET front end. The JFET unity-gain preamp shown schematically in fig. 6, modified from one described in Horowitz and Hill [33], is designed to have a bandwidth of 50 MHz, ensuring no unwanted phase shifts below 4 MHz, and has an input noise figure of a few $nV/Hz^{1/2}$, controlled primarily by the dual JFET. Because the triax effectively connects the output directly to the input, the design and layout of the circuitry is crucial. The inherent response of a HSSR is such that it is not the full bandwidth of the preamp, but instead the noise per unit bandwidth that is important for overall s/n.

The preamp is followed by a simple but quiet RF amplifier with a bandwidth of about 20 MHz based on an application note for the MAR-6 amplifier made by Mini-Circuits [34]. This stage, shown in fig. 7, has a voltage gain of 100.

Following the preamp is a mixer/filter shown in fig. 8 adapted from the data sheets from Motorola [35] and National Semiconductor [36]. This mixer performs an instantaneous multiplication of the RF transducer signal at frequency f, including whatever noise is present, with a (noise free) local oscillator (LO) signal at $f + \Delta f$, where Δf is the intermediate frequency or IF. The output of the mixer is, then, the amplifier noise upshifted in frequency by $f + \Delta f$, and information-containing signals at $2f + \Delta f$, and Δf. Because the noise is basically white, noise per unit bandwidth is unaffected. After the mixer is a state-variable analog filter tuned to Δf. Because the LO source is of constant amplitude, the amplitude of the mixer/filter IF output at Δf is

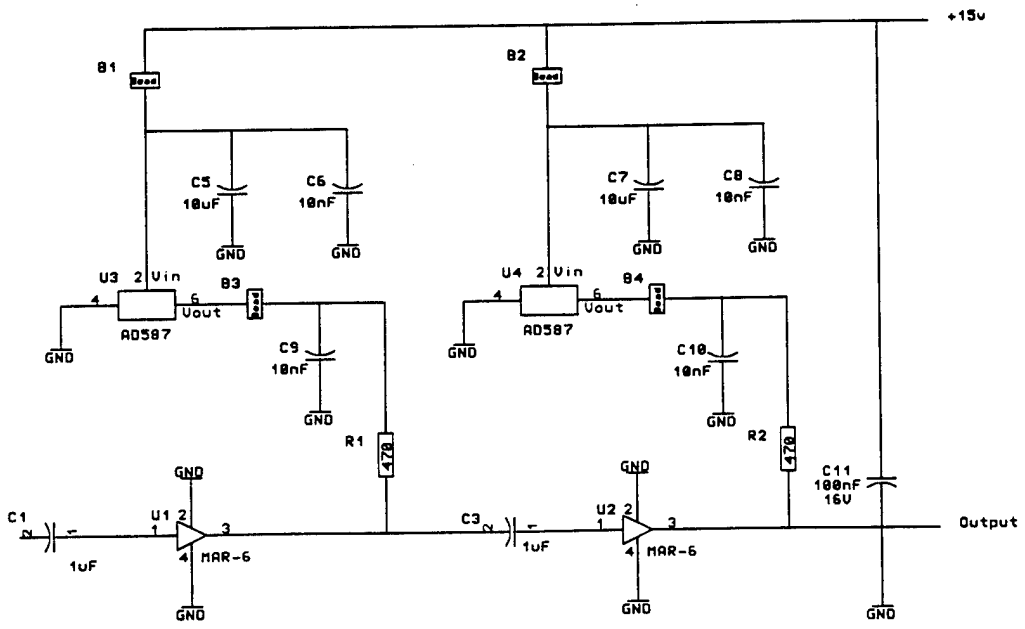


Fig. 7. Shown is the schematic diagram of the low-noise RF amplifier used for our RUS measurements, based on the MAR-6 integrated circuit. The layout of this circuit is extremely important.

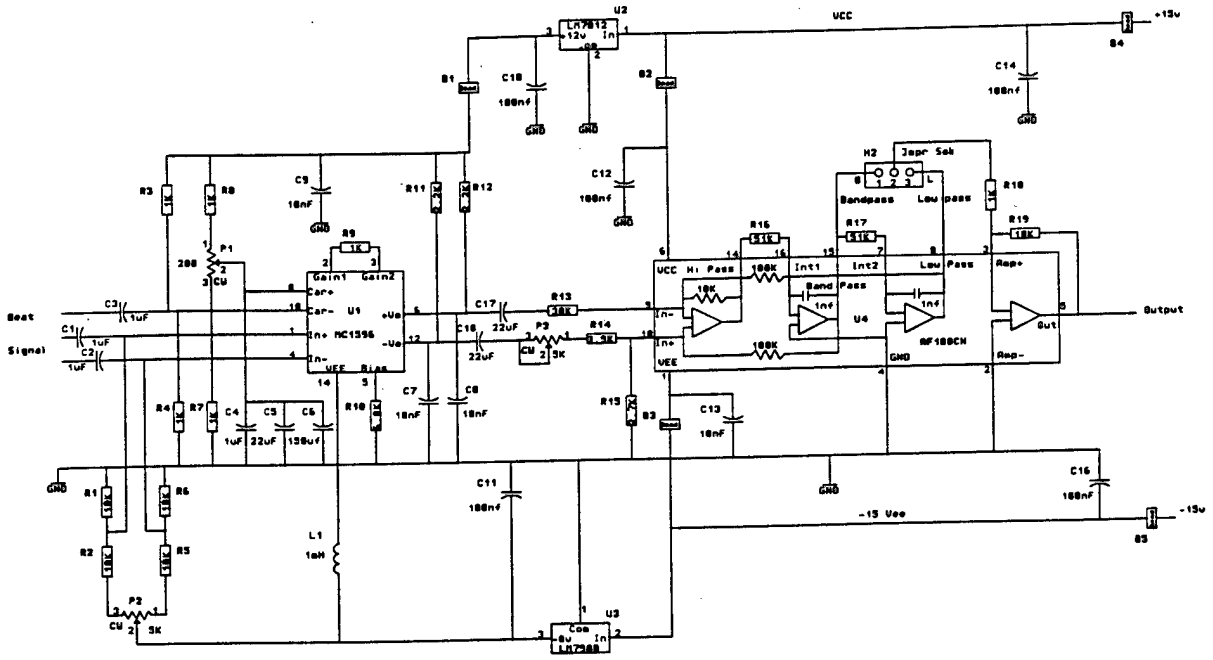


Fig. 8. The mixer/filter based on the widely used Motorola 1496 mixer IC and the National Semiconductor AF100 state-variable active filter is shown schematically. The layout of this circuit is extremely important.

proportional to the resonance signal at f. We use an IF frequency of 1 kHz and a filter bandwidth of 100 Hz. This yields a noise bandwidth 100 Hz wide at the filter output, but also limits the receiver response time to of order 10 ms. This is not a real limitation on the data acquisition rate because each sample resonance is stepped through using steps much narrower than the resonance width. Thus the receiver need only respond to incremental signal amplitude changes at sample resonance (i.e. if n steps are taken through one resonance then the receiver response time required is reduced by a factor of n).

At this point the resonance information wanted appears at a fixed IF frequency of 1 kHz combined with a 100 Hz wide slice of noise. To get the amplitude of the IF signal into a PC-AT type computer requires that it be 'detected'. The best approach is to use an analog to digital converter (ADC) to digitize the filter output such as an Analogic LSDAS-16, a 16-bit, 16 Channel, 50 kHz ADC [37]. By running the converter at 32 kHz and taking 320 readings, we acquire 10 cycles of the IF signal. We also acquire almost exactly 320 cycles of the most important interference source, the electric field associated with the VGA monitor of the computer. The *absolute value* of the 320 readings is averaged with software to generate the resonance amplitude at the frequency step chosen. The effect of this is (1) to produce a constant background offset arising from VGA interference, and (2) to signal average (equivalent to another 100 Hz bandpass filter) the IF signal over 10 cycles with no analog time constants to generate glitches if the measurement is started with uncontrolled phase of the IF signal (which it is).

A complete sweep through each individual mode is made using the above system and a dual digital synthesizer card to generate the LO and RF signals. The card we use was designed by us, is capable of 32 bit frequency and phase control and 15 bit amplitude control of two separate outputs, and is commercially available [26]. The card fits in a standard PC-AT slot, is controllable by Microsoft QuickBASIC or other languages and has a maximum output of 1 VAC (sufficient to drive the transducers directly) up to about

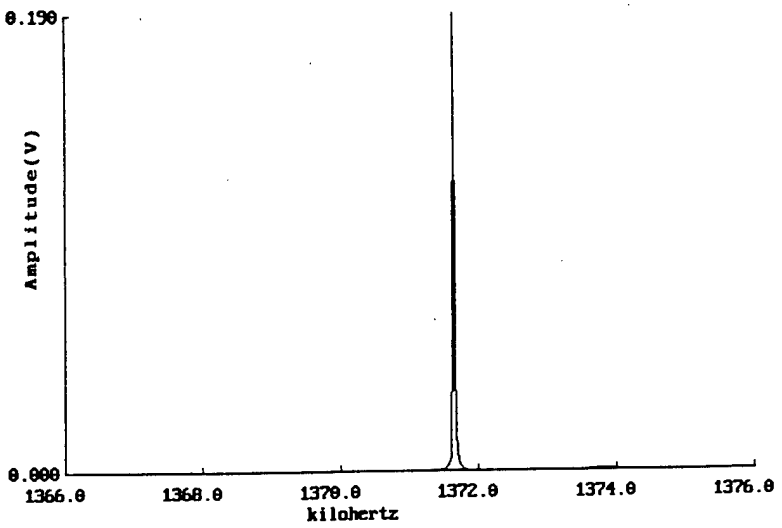


Fig. 9. A typical (and definitely not the best) resonance in the $La_{1.86}Sr_{0.14}CuO_4$ single crystal is shown to illustrate the quality of the data obtainable from a RUS measurement.

8 MHz. Other computer-controlled frequency synthesizers can be used successfully. A typical resonance is shown in fig. 9.

The resonance data acquired is processed to subtract background and then, using all the data acquired, a first moment of the usually Lorentzian resonances is computed. In this way, further noise reduction occurs so that we can determine reliably the center frequency of a resonance to about 3% or less of the linewidth. A complete experimental run would consist of (1) a search for all modes at room temperature, (2) acquisition of a data file produced by a narrow sweep through each resonance found, (3) a room-temperature fit to the moduli (and iteration of (1) and (2) if the fit is inadequate or predicts missing modes, and (4) a repeat of (2) for each temperature desired using sufficiently small temperature steps so as not to lose track of mode identifications.

3. Typical results

Having described the principles, procedures and hardware for making RUS measurements and analyzing the data, we present here some examples of the application of RUS to the study of structural phase transitions (SPT). The examples reviewed here are the soft-mode-driven SPTs in the perovskite system La_2CuO_4, which includes several high-temperature superconductors, and in $SrTiO_3$.

$SrTiO_3$ is a perovskite that undergoes a SPT from cubic (the high-temperature or 'symmetric' phase) to tetragonal (the low-temperature or 'unsymmetric' phase) crystal symmetry at 105 K. This material is particularly interesting because its SPT is a canonical example of a soft-mode [38] phase transition and has been well studied both theoretically [39] and experimentally using conventional ultrasonic techniques [40,41,42]. Furthermore, the perovskite structure and underlying titanium–oxygen octahedra in $SrTiO_3$ are very similar to the structure of the high-temperature superconductors La_2CuO_4 and $La_{2-x}Sr_xCuO_4$ to be discussed below. On a more practical level, large high-quality single crystals are readily available commercially, and no macroscopic fields develop at the SPT to complicate data analysis. We obtained a large single crystal [43] and prepared several RP samples (1.9 mm × 1.5 mm × 1.0 mm) as was described above.

The sample geometry was accurate to about 2 μm and the intrinsic quality was high so that

we were able to measure and fit the first 33 resonant frequencies at room temperature with an rms error of 0.075 percent. The values of the elastic moduli determined by the fit, as well as several sets of values from measurements by others, are shown in table 2. The agreement is excellent.

Our real interest in this material is not so much a test of RUS but rather a study of the details of the temperature dependence of the elastic moduli through the SPT. Because of the unique capability of RUS to determine all moduli simultaneously, precise comparisons between moduli can be made and compared to Ginsburg–Landau (G–L) type predictions. The SPT in $SrTiO_3$ occurs as a result of the softening of a particular zone edge lattice vibration, the tilting of the titanium–oxygen octahedra around the (1 0 0), (0 1 0) or (0 0 1) axes. As the frequency of this phonon decreases to zero, these octahedra develop a static displacement, reducing the crystal's symmetry to tetragonal and doubling the size of the unit cell. Because of the three-fold degeneracy of this octahedral tilt (rotation about x, y or z) the material also twins as it goes through the SPT, making RUS measurements difficult in the low-temperature phase.

Theoretically, this transition is well understood. The soft-mode description coupled with a G–L free energy enables accurate quantitative predictions to be made for the elastic response of the material through the transition. Rather than work with the complete expression for the free energy including all possible strains and the full three-component order parameter, discussed in detail elsewhere [39], we will focus here on a qualitative understanding by considering a single-strain and one-component order parameter. This is more than casually justified because if one knows which way the octahedra tilt, then use of a single-component order parameter causes no loss of generality. Because in an unstrained single crystal the order parameter does develop in a single direction, one certainly knows its direction after the fact. It is only important that one is careful with the group theory. That is, the full symmetry and number of required components of the order parameter are use to determine what terms must be included in a single-order-parameter description. Thus an accurate single-component-order-parameter free energy can be written

$$F = \tfrac{1}{2}c_0\varepsilon^2 + \tfrac{1}{2}\alpha(T - T_c)q^2 + \tfrac{1}{4}\beta q^4 + \tfrac{1}{2}\gamma\varepsilon q^2 \tag{39}$$

where ε is the strain, q the order parameter and α, β, γ and c_0 temperature-independent constants. The first term in eq. (39) represents the usual elastic energy, the second and third an expansion in the order parameter, and the last term the strain–order parameter coupling. Because the order parameter in this transition is a tilt in a mirror plane, only even powers appear (positive or negative tilts are equivalent energetically). Given such a free energy, the change in elastic moduli can be calculated by minimizing eq. (39) with respect to strain [44]. Such a minimization gives

$$c = c_0\,, \qquad T > T_c\,, \tag{40}$$

$$c = c_0 - \gamma^2/2\beta\,, \quad T < T_c\,. \tag{41}$$

Thus, a Ginsburg–Landau analysis of this phase transition predicts a step decrease in elastic moduli at the transition and, using the full crystal symmetry, predicts the relative size of the step for each individual modulus. Our data for the elastic moduli of $SrTiO_3$ as a function of temperature are shown in fig. 10. Each modulus increases with decreasing temperature from thermal contraction and then exhibits a sharp (but not step-like) decrease in the region near the transition, in agreement with the above analysis

Table 2
Room temperature elastic moduli (all values are in units of 10^{12} dyne/cm^2) of $SrTiO_3$ determined by RUS as well as by conventional ultrasonic techniques. The percentages after our data are the error estimates for the individual moduli, determined as discussed above.

c_{11}	c_{12}	c_{44}	Source
3.17	1.02	1.23	Ref. [40]
3.31	1.05	1.26	Ref. [41]
3.15 (0.2%)	1.02 (0.7%)	1.22 (0.01%)	This work

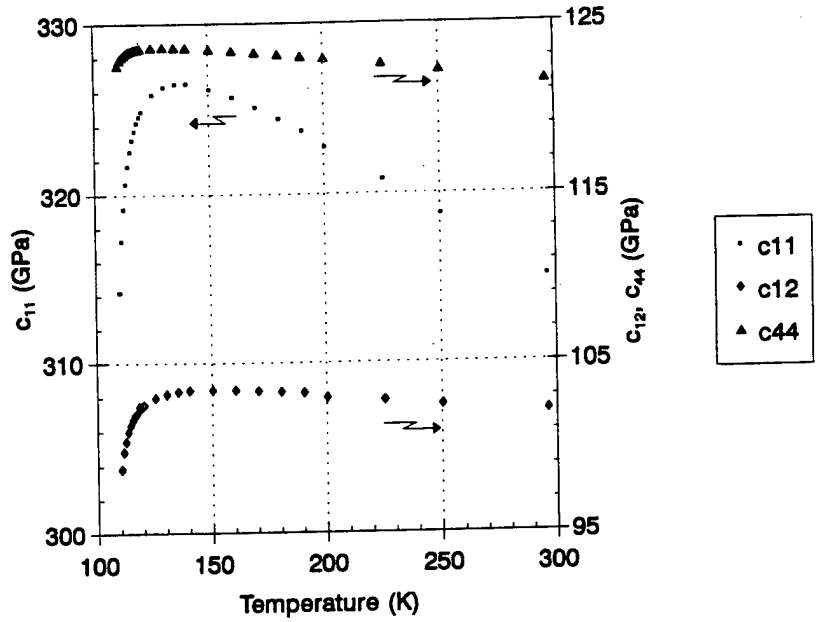


Fig. 10. The three elastic moduli of a single crystal of $SrTiO_3$ near the structural phase transition are shown as a function of temperature. These data were obtained using RUS.

as well as with the experimental work of others. Unfortunately, twinning of the crystal at the SPT prevents RUS from accessing the moduli in the unsymmetric phase so that some quantitative predictions of the Ginsburg–Landau theory cannot be tested by us. This has, however, been done by other groups using pulse-echo measurements and the results are in accord with theory [41]. Note that any ultrasound measurements in a twinned sample average over some set of moduli, and that to obtain useful information, some detailed knowledge of the twinning pattern and a model for backing out moduli are required. RUS is so sensitive to macroscopic twinning that usable data of any sort often cannot be obtained in a twinned sample because of substantial degradation of the resonance signals. This is typically not the case for pulse-echo, nor for RUS on microtwinned or polycrystal materials.

The departure from true step-like behavior can be attributed to both thermal fluctuations and to imperfections or defects in the crystal [45]. In order to verify this fact and to demonstrate the fundamental difference between this rounding and the effects observed in $La_{2-x}Sr_xCuO_4$, discussed below, we vacuum-annealed a RP of $SrTiO_3$ to create oxygen vacancies in an attempt to braden the transition. While a sharp decrease remains after annealing, the decrease is distinctly broader in the annealed sample, in agreement with previous work [42]. Having demonstrated that there are no surprises in a RUS study of $SrTiO_3$, we describe similar measurements on a more difficult system with results that prove to be not so accommodating.

Unlike $SrTiO_3$, La_2CuO_4 and also its high-temperature superconducting relative $La_{2-x}Sr_xCuO_4$ are not readily available as untwinned, morphologically perfect single crystals. The only such samples extant are in the 1 mm size range and require very considerable effort to produce. Thus conventional ultrasound techniques must be applied to either large, poor-quality twinned samples or not at all. In this system, the tetragonal to orthorhombic (TO) SPT occurs at 223 K for $x = 0.14$ and about 530 K for $x = 0.0$ [46]. In table 3 are the elastic moduli for an untwinned, orthorhombic crystal of La_2CuO_4 (1.735 mm × 1.536 mm × 1.108 mm, 7.026 g/cc) and also for a tetragonal crystal of $La_{1.86}Sr_{0.14}CuO_4$

Table 3
The full elastic moduli of $La_{1.6}Sr_{0.14}CuO_4$ and La_2CuO_4 in units of 10^{12} dyne/cm^2 determined using RUS. The first entry for $La_{1.86}Sr_{0.14}CuO_4$ is for the usual tetragonal basis. The second entry is simply the first one rotated by 45° such that a direct comparison to the orthorhombic stoichiometric compound can be made. The errors are about 1.3% for c_{ii}, $i = 1, 3$, 0.05% for c_{ii}, $i = 4, 6$, and 3.1% for the off-diagonal moduli for La_2CuO_4. The corresponding error estimates for $La_{1.86}Sr_{0.14}CuO_4$ are 0.25%, 0.01% and 1.0% respectively.

	c_{11}	c_{22}	c_{33}	c_{23}	c_{13}	c_{12}	c_{44}	c_{55}	c_{66}
$La_{1.86}Sr_{0.14}CuO_4$	2.666	–	2.571	–	0.992	0.649	0.677	–	0.587
$La_{1.86}Sr_{0.14}CuO_4$	2.245	–	2.571	–	0.992	1.071	0.677	–	1.009
La_2CuO_4	1.747	1.730	2.662	0.991	0.928	0.900	0.653	0.669	0.992

(2.470 mm × 2.267 mm × 2.192 mm, 6.946 g/cc) determined using RUS. The La_2CuO_4 crystal was grown by Canfield, Fisk and Kodali from a flux [47]. The Sr doped crystal was produced by Tanaka and Kojima using a travelling solvent/floating zone system [48]. We note that the axes of the La_2CuO_4 sample were not aligned with the crystallographic axes. Thus the fitting procedure had to determine not only the moduli and dimensions but also the crystal orientation. This takes far more computer time because the matrices to be inverted are no longer block diagonal. Nevertheless, the errors for the determination of moduli and angles are low because over 35 resonances were used in the fit.

We review here the microscopic deformations at the SPT in the La_2CuO_4 compounds, the Ginsburg–Landau Hamiltonian, and the expected effects on the sound velocities. Using RUS measurements on $La_{2-x}Sr_xCuO_4$ and a comparison with $SrTiO_3$, we present direct evidence for breakdown of tetragonal symmetry at the Brillouin zone center, show how this can explain the very strong temperature dependence of c_{66} above the TO SPT, and discuss some implications.

To understand what drives the TO transition in La_2CuO_4, consider first (fig. 11) the four Cu–O bonds that lie in the Cu–O plane and also form part of the O octahedra. Through the TO transition, these Cu–O bond lengths remain fixed [49]. What does change is the angle between the two O–Cu–O diagonals of the octahedra. In the tetragonal phase, the diagonals are perpendicular. In the orthorhombic phase they scissor slightly, doubling the unit cell. This Jahn–

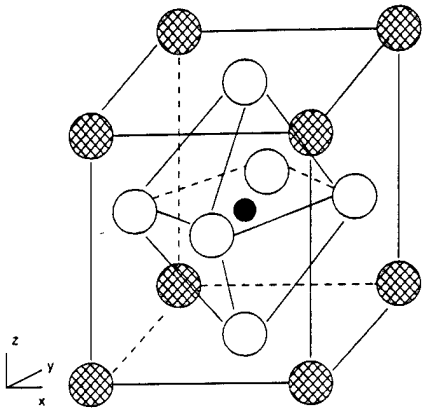


Fig. 11. We illustrate here the arrangement of ions (not to scale) of La_2CuO_4. Cu is the solid circle, the open circles are O and the shaded circles are La.

Teller-like distortion has the effect of increasing the length of either the (1 1 0) or the (1 $\underline{1}$ 0) axis. This is shown schematically in fig. 12. For the crystal to accomodate this, the Cu–O plane also buckles in the corresponding direction, taking the octahedra with it so that they tilt. This tilt is the x-point soft mode. Of course, a given octahedron could tilt in either of four possible directions, i.e. in the positive (1 1 0) direction, the negative (1 1 0) direction, the positive (1 $\underline{1}$ 0) direction, or the negative (1 $\underline{1}$ 0) direction. Thus both kinks (positive–negative tilt phase error) and twins (1 1 0) versus (1 $\underline{1}$ 0) tilt are possible, and the order parameter must have two components, q_1 and q_2.

Because any possible tilt preserves mirror symmetry, either sign of tilt is equivalent. Therefore any coupling of the Brillouin-zone-edge octahedron-tilt phonon mode to any zone-center

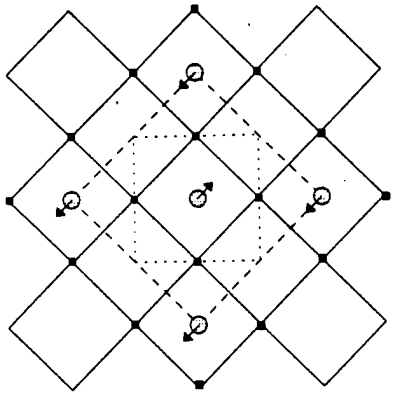


Fig. 12. Shown is a diagram showing the O atoms in the Cu–O plane (solid circles) and the O atoms at the apices of the O octahedra (shaded circles) projected onto the Cu–O plane. The Cu atoms are directly beneath the undisplaced O atoms and are not shown. Upon transition to the orthorhombic state, the apical O atoms displace as shown by the arrows (a twin would have displacements in the Cu–O plane perpendicular to those shown, i.e. turn the figure on its side). The tetragonal unit cell is the smaller dashed square, the orthorhombic one is the larger dashed square, although it is really a rectangle with the longer sides parallel to the arrows.

acoustic phonons must be quadratic in lowest order. Moreover, most of the effects of the phase transition are seen in c_{66}, the shear modulus for deformations of the Cu–O plane. This is also easily seen because when the O–Cu–O diagonals scissor, the square base of the tetragonal unit cell becomes a rhombus in the orthorhombic phase (fig. 12). Deforming the square into a rhombus is exactly equivalent to a c_{66} shear. We have, as yet, no measurements above the TO transition in La_2CuO_4 because our RUS cell cannot handle such temperatures. However, in $La_{1.86}Sr_{0.14}CuO_4$ the TO transition occurs at 223 K, a very convenient temperature.

In fig. 13 we show the resonant frequency of an eigenmode of the single crystal of $La_{1.86}Sr_{0.14}CuO_4$ as a function of temperature T and in fig. 14 we show $1/Q$ vs T where Q is the quality factor for the resonance. Numerical analysis of the motion establishes that the eigenmode of fig. 13 depends almost purely on c_{66}. Absent dynamical effects, we would treat the tempeature dependence of c_{66} with the same Ginsburg–Landau (G–L) Hamiltonian used for $SrTiO_3$. As with $SrTiO_3$, simple quadratic coupling and no dynamics produces only a step discontinuity in c_{66} at the SPT. This is not what the data show. The data fit a Curie–Weiss (C–W) softening of the form

$$c_{66}(T) = c_{66}(1 - T_0/(T - T_c)) \tag{42}$$

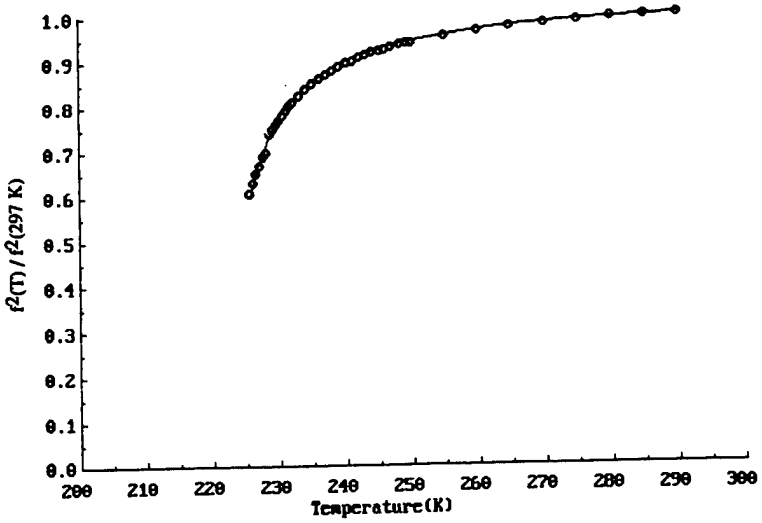


Fig. 13. Shown is the normalized resonant frequency squared (proportional to a modulus) for a $La_{1.86}Sr_{0.14}CuO_4$ mode that is nearly pure c_{66} as a function of temperature (circles). The solid line is a Curie–Weiss fit to the data. The gap in the data just below 230 K is evidence of symmetry-breaking effects.

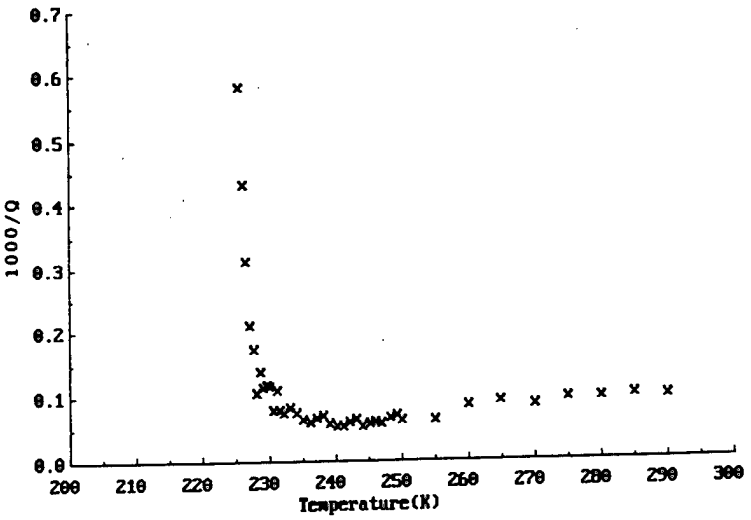


Fig. 14. The scaled inverse quality factor $1000/Q$ is proportional to the loss or ultrasonic attenuation in the single crystal of $La_{1.86}Sr_{0.14}CuO_4$. Very near the SPT at 223 K, the attenuation increases dramatically, but at temperatures well above the SPT but still in the Curie–Weiss region, no excess attenuation is observed.

where T_c is 223 K, c_{66} is a temperature-independent constant and the fit, shown as the solid line in fig. 13, is accurate to 0.2% over more than a decade in $T_0/(T - T_c)$.

Gaussian fluctuations of the order parameter [9], self-consistent phonons [7] and linear coupling between strain and order parameter [44] all yield a C–W behavior for c_{66}. For Gaussian fluctuations, the critical exponent for the specific heat and for the elastic moduli is $\Lambda = 2 - d/2$ where d is the dimension of the order parameter. In our system, the order parameter is two-dimensional, thus the critical exponent (the exponent of $1/(T - T_c)$) is unity, in agreement with the data. However, our C–W fit is over a temperature range of about 80 K ($T_0 = 1.47$ K). This is a very large range for fluctuations to be important, much larger than the range for the similar SPT in $SrTiO_3$. An upper-bound estimate for the fluctuations regime [9] is found by using a few lattice spacings for the coherence length, and by using a 1% ($SrTiO_3$ has about a 10% modulus discontinuity at its phase transition temperature) modulus discontinuity to obtain a fluctuation range of about 1 K, comparable to the region in fig. 14 where the ultrasonic attenuation increases sharply. Thus it appears very unlikely that 2-D Gaussian fluctuations can explain what we observe.

A self-consistent phonon treatment of the anharmonic potential associated with the zone-edge soft mode of the O octahedra can also produce C–W modulus softening [7]. For this sort of treatment to work, the zone-edge soft mode must be linearly coupled to the zone-center acoustic mode. The 1-D treatment in ref. [7] deals with this by inserting the anharmonic spring, used in the shell-model construct to develop the self-consistent phonon dispersion curve, in series with the ion cores. Thus this spring contributes to the potential energy for any value of k, the phonon wave vector.

Neutron scattering measurements [4] show that the soft mode is part of the phonon branch corresponding to c_{44}, not c_{66}. Without some linear coupling term to the c_{66} dispersion curve, it is not easy to see the applicability of self-consistent phonons. Were such a term to be added, the model would be forced to become explicitly 3-D, and because both the coupling and the energies would depend on the anharmonicity, the C–W exponent would likely be lost.

The third possibility we consider here is the

replacement of quadratic coupling with linear coupling (for $T > T_c$, the inclusion of the quadratic term has no effect with or without the linear term present) between order parameter and strain in eq. (39). Using the same justification for a single-component order parameter as in $SrTiO_3$ we find

$$F = \tfrac{1}{2}c_0\varepsilon^2 + \tfrac{1}{2}\alpha(T - T_c)q^2 + \tfrac{1}{4}\beta q^4 + \gamma\varepsilon q \qquad (43)$$

and

$$c_{66}(T) = c_{66} - \gamma^2/(T - T_c) \qquad \text{for } T > T_c\,, \qquad (44)$$

$$c_{66}(T) = c_{66} - \gamma^2/(2(T - T_c)) \quad \text{for } T < T_c\,, \qquad (45)$$

as required to fit the data of fig. 13. To justify a linear coupling term, the $La_{1.86}Sr_{0.14}CuO_4$ crystal must be either nonlinear or nontetragonal.

In fig. 15 we plot the lowest eigenfrequencies of the $La_{1.86}Sr_{0.14}CuO_4$ crystal vs T and in fig. 16 are plotted the lowest two eigenmodes on an expanded scale, showing an avoided crossing of 2%. Note that in fig. 15 avoided crossings are observable in several places. The mechanical Lagrangian for analysis of the resonances of this material is based on a linear tetragonal solid [13,14]. The model produces eight orthogonal symmetry classes for the modes, therefore none of the avoided crossings should occur. Their existence can be explained only if the crystal is nonlinear, not tetragonal, or has excessive preparation errors.

This sample has a TO transition width of much less than 1 K as determined by the RUS variation of both Q and c_{66}, it has a superconducting T_c of 38 K, its faces have been ground parallel to better than 1 micron, some resonances have $Q > 70\,000$, it has near theoretical bulk density and no visible defects. There are, therefore, no sample preparation errors even remotely close to the few percent required to produce the observed avoided crossings. To test for anharmonicity, the measurements of fig. 16 were repeated at resonance amplitudes varying over an order of magnitude. All the scans were identical to about 0.01%, the limit of our temperature control precision. Thus no anharmonic affects are present.

X-ray [46] and elastic neutron scattering [4,49] studies of this material all indicate unambiguously that the structure is tetragonal. However, both types of probe operate at an effective $k = 0$. That

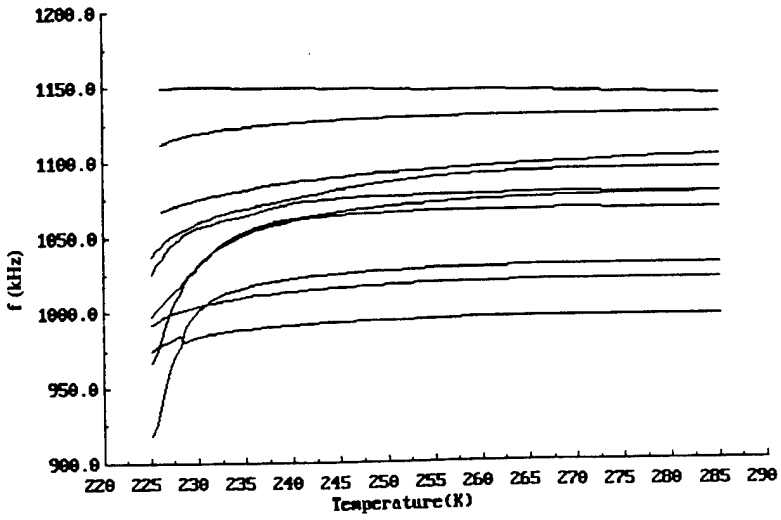


Fig. 15. This is a plot of many of the resonances of the $La_{1.86}Sr_{0.14}CuO_4$ single crystal as a function of temperature above the SPT. Avoided crossings are numerous.

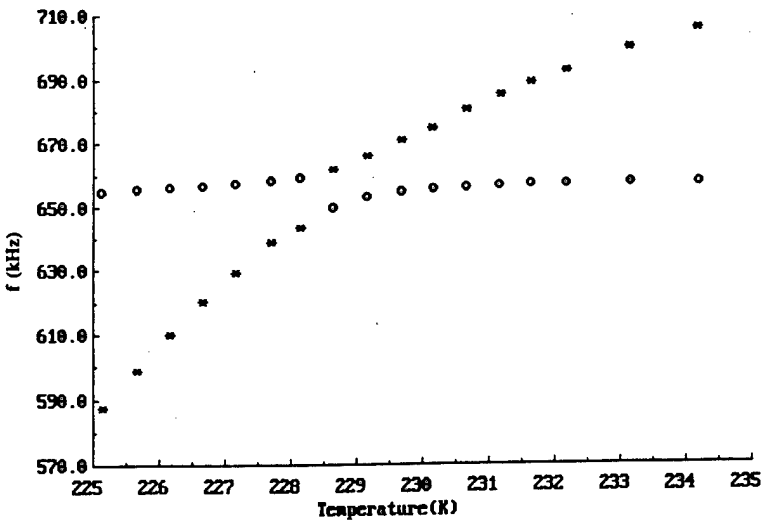


Fig. 16. Plotted are the data of fig. 13 for the c_{66} mode and also data for the c_{44} mode for the $La_{1.86}Sr_{0.14}CuO_4$ single crystal as a function of temperature. These modes have different symmetries so that no avoided crossing is expected. The surprisingly large effect may arise from symmetry-breaking at the Brillouin-zone-edge produced by Sr disorder combined with the soft-mode-driven SPT.

is, they average over many unit cells. Because RUS is also a $k = 0$ probe, it too should see a tetragonal structure. That it does not might be related to the destruction of zone-edge symmetry because of the disordered Sr doping. At small k, the Sr concentration fluctuations average out, and a tetragonal structure is observed. However, RUS is sensitive to the strain susceptibility. At temperatures well above (150 K or more) the TO transition, the x-point soft mode has negligible effect on c_{66} (see fig. 13). At temperatures somewhat above 300 K, all the sound velocities begin to decrease on cooling as a result of the coupling between the soft mode and the strains. This coupling might also carry with it a symmetry-breaking term at temperatures near the TO transition related to the total softening of c_{66}.

Such a symmetry-breaking effect is only observable if some effect, the SPT for this example, makes the moduli vary sufficiently rapidly that nominally orthogonal modes cross, and if the experimental probe can clearly separate the responses of the modes that cross. We know of no other ultrasound probe with this property. Thus RUS is capable of extracting new information as well as providing an alternative high-precision general ultrasound probe.

Acknowledgements

The authors wish to acknowledge Orson L. Anderson for many wonderful discussions, W. Gilbert Clark for his insights into electronics and instrumentation, Joe D. Thompson, Paul C. Canfield, Stuart E. Brown, Raymond M. Dixon and Selmer Wong for assistance with sample preparation and characterization, Philip B. Allen, Stuart A. Trugman and Lu J. Sham for tremendous patience and attention with the theoretical analyses and Tom Stockebrand for preparation of the printed circuits that make the electronics work so well. This work was performed under the auspices of the United States Department of Energy.

References

[1] B. Golding, W.H. Haemmerle, L.F. Schneemeyer and J.V. Waszczak, in: IEEE 1988 Ultrasonics Symposium (IEEE, Piscataway, 1988) p. 1079.

[2] T. Laegreid, Wu Ting, O.-M. Nes, M. Slaski, E. Eidem, E.J. Samuelsen, K. Fossheim and Y. Hidaka, in: Advances in Superconductivity II: Proceedings from the 2nd International Symposium on Superconductivity, November 14–17, 1989, Tsukaba, Ibaraki, Japan.

[3] P. Baumgart, S. Blumenroder, A. Erle, B. Hillebrands, G. Guntherodt and H. Schmidt, Solid State Commun. 69 (1989) 1135.
[4] P. Boni, J.D. Axe, G. Shirane, R.J. Birgeneau, D.R. Gabbe, H.P. Jenssen, M.A. Kastner, C.J. Peters, P.J. Picone and T.R. Thurston, Phys. Rev. B 38 (1988) 185.
[5] I.V. Aleksandrov, A.R. Goncharov and S.M. Stishov, JETP Lett. 47 (1988) 428.
[6] K. Kawasaki and A. Ikushima, Phys. Rev. B 1 (1970) 3143.
[7] A. Bussmann-Holder, A. Migliori, Z. Fisk, J.L. Sarrao, R.G. Leisure and S.-W. Cheong, Phys. Rev. Lett. 67 (1991) 512.
[8] L.D. Landau and E.M. Lifshitz, Statistical Physics, Part I (Pergamon Press, Oxford, 1980; 3rd ed.).
[9] S.K. Ma, Modern Theory of Critical Phenomena (Benjamin–Cummings, Reading, MA, 1976).
[10] R. Holland, J. Acoust. Soc. Am. 43 (1968) 988.
[11] H.H. Demarest, J. Acoust. Soc. Am. 49 (1971) 768.
[12] O.L. Anderson, E. Schreiber and N. Soga, Elastic Constants and Their Measurements (McGraw-Hill, New York, 1973).
[13] I. Ohno, J. Phys. Earth 24 (1976) 355.
[14] William M. Visscher, A. Migliori, T.M. Bell and R.A. Reinert, J. Acoust. Soc. Am. 90 (1991) 2154.
[15] I. Ohno, Phys. Chem. Minerals 17 (1990) 371.
[16] H. Oda, O.L. Anderson and Isao Suzuki, in: Proceedings of the 28th Annual Technical Meeting of the Society of Engineering Sciences, November 6–8, 1991, Gainesville, FL.
[17] B.T. Smith, J.M. Boyle, J.J. Dongarra, B.S. Garbow, Y. Ikebe, V.C. Klema and C.B. Moler, in: Matrix Eigensystem Routines – EISPACK Guide, Lecture Notes in Computer Science No. 6, eds. G. Goos and J. Hartmanis (Springer, New York, 1976); B.S. Garbow, J.M. Boyle, J.J. Dongarra and C.B. Moler, in: Matrix Eigensystem Routines – EISPACK Guide Extension, Lecture Notes in Computer Science No. 51, eds. G. Goos and J. Hartmanis (Springer, New York, 1977).
[18] W.H. Press, B.P. Lannery, S.A. Teukolsky and W.T. Vetterling, Numerical Recipes (Cambridge University Press, 1986).
[19] Crystal Bond; Aremco Products Inc., P.O. Box 429, Ossining, NY 10562.
[20] Optical Fiber Technologies, 2 Lyberty Way, Westford, MA 01886.
[21] E. Mochizuki, J. Phys. Earth 35 (1987) 159.
[22] W.L. Johnson, S.J. Norton, F. Benedec and R. Pless, J. Acoust. Soc. Am. 91 (1992) 2637.
[23] J.D. Maynard, J. Acoust. Soc. Am. 91 (1992) 1754.
[24] A. Stekel, J.L. Sarrao, T.M. Bell, Ming Lei, R.G. Leisure, William M. Visscher and A. Migliori, J. Acoust. Soc. Am. 92 (1992) 663.
[25] US patent # 4 976 148 Resonant Ultrasound Spectrometer, A. Migliori, William M. Visscher and Z. Fisk; US patent # 5 062 296 Resonant Ultrasound Spectroscopy, A. Migliori.
[26] Quatro Corp., 4300 San Mateo NE, Suite B-2890, Albuquerque, NM 87110.
[27] W.G. Clark, Rev. Sci. Instr. 35 (1964) 316.
[28] Valpey-Fisher, Ultrasound Division, 75-T South St., Hopkington, MA 01748.
[29] Any low-viscosity, clear epoxy, e.g. Stycast 1266; Emerson and Cuming Inc., 77 Dragon Ct., Woburn, MA 01888.
[30] Y. Sumino, I. Ohno, T. Goto and M. Kumazawa, J. Phys. Earth 24 (1976) 263.
[31] O. L. Anderson and T. Goto, Physics of the Earth and Planetary Interiors 55 (1989) 241.
[32] W.G. Clark, private communication.
[33] P. Horowitz and W. Hill, The Art of Electronics (Cambridge University Press, Cambridge, 1980; 1st ed.) p. 236.
[34] Mini-Circuits, P.O. Box 350166, Brooklyn, NY 11235–0003.
[35] MC1496 balanced Modulator–Demodulator, Motorola Semiconductor Products Inc., Box 20912, Phoenix, AZ 85036.
[36] AF100 Universal Active Filter, National Semiconductor Corp., 2900 Semiconductor Dr., P.O. Box 58090, Santa Clara, CA 95052–8090.
[37] Analogic Corp., 360 Audobon Rd., Wakefield, MA 01880.
[38] W. Cochran, Adv. Phys. 9 (1960) 387; P.W. Anderson, in: Fizika Dielektrov, ed. G.I. Sknavi (AN SSSR, Moscow, 1960) p. 290.
[39] J.C. Slonczewski and H. Thomas, Phys. Rev. B 1 (1970) 3599.
[40] R.O. Bell and G. Rupprecht, Phys. Rev. 129 (1963) 90.
[41] B. Luthi and T.J. Moran, Phys. Rev. B 2 (1970) 1211.
[42] D. Bauerle and W. Rehwald, Solid State Commun. 27 (1978) 1343.
[43] Sample supplied by I.D. Raistrick, Los Alamos National Laboratory.
[44] W. Rehwald, Adv. Phys. 22 (1973) 721.
[45] A.P. Levanyuk and A.S. Sigov, Defects and Structural Phase Transitions (Gordon and Breach, New York, 1989).
[46] R.M. Fleming, B. Batlogg, R.J. Cava and E.A. Rietman, Phys. Rev. B 35 (1987) 7191.
[47] S.-W. Cheong, J.D. Thompson and Z. Fisk, Physica C 158 (1989) 109.
[48] I. Tanaka and H. Kojima, Nature 337 (1989) 21.
[49] R.J. Cava, A. Santoro, D.W. Johnson Jr. and W.W. Rhodes, Phys. Rev. B 35 (1987) 6716.

On the normal modes of free vibration of inhomogeneous and anisotropic elastic objects

William M. Visscher, Albert Migliori, Thomas M. Bell, and Robert A. Reinert
Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 5 April 1991; accepted for publication 23 June 1991)

The Hamilton's principle approach to the calculation of vibrational modes of elastic objects with free boundaries is exploited to compute the resonance frequencies of a variety of anisotropic elastic objects, including spheres, hemispheres, spheroids, ellipsoids, cylinders, eggs, shells, bells, sandwiches, parallelepipeds, cones, pyramids, prisms, tetrahedra, octahedra, and potatoes. The paramount feature of this calculation, which distinguishes it from previous ones, is the choice of products of powers of the Cartesian coordinates as a basis for expansion of the displacement in a truncated complete set, enabling one to analytically evaluate the required matrix elements for these systems. Because these basis functions are products of powers of x, y, and z, this scheme is called the *xyz* algorithm. The *xyz* algorithm allows a general anisotropic elastic tensor with any position dependence and any shape with arbitrary density variation. A number of plots of resonance spectra of families of elastic objects are displayed as functions of relevant parameters, and, to illustrate the versatility of the method, the measured resonant frequencies of a precision machined but irregularly shaped sample of aluminum (called a potato) are compared with its computed normal modes. Applications to materials science and to seismology are mentioned.

PACS numbers: 43.20.Ks, 43.40.At

INTRODUCTION

Some years ago, it was noticed by Holland[1] and Demarest[2,3] that a very simple variational principle can be used to derive an eigenvalue equation for the normal-mode frequencies and eigenvectors for the vibrations of an elastic body with free boundaries. Subsequently Ohno[4] and others[5–8] have refined this method and its application to the resonant ultrasonic determination of elastic constants of materials, using small rectangular parallelepipedal single-crystal samples. The idea is illustrated in Fig. 1, which shows a crystal suspended between two transducers, one of which excites the sample, and the other measures its response. Others have used spheres,[9,10] again single-crystals with anisotropic elastic tensors (with orthorhombic or better symmetry). Spheres have the advantage that it becomes unnecessary (indeed, impossible) to align the faces of the sample with crystal axes. It is the purpose of the present paper to present a simple computational scheme with which the free vibrations of all the systems that have been analyzed up to now can be computed, and to give results for some new systems. We have fabricated a sample of one of these new systems, and have measured and computed its resonant frequencies.

I. THE METHOD

A. Motivation

It is reasonable to expect that, if one knew all the normal-mode frequencies of an elastic object, one could deduce all of the relevant properties of that object, including, up to a common scale factor, its elastic tensor and density (including their spacial variations), and its shape. Whether this is rigorously true, even for such a simple system as a two-dimensional elastic membrane, is still an unsolved problem in applied mathematics.[11] In practice, resonant ultrasonic methods have been developed, and are still being refined in significant ways,[12] which promise to become benchmark procedures for measuring these material properties. In order to exploit these ideas, it is necessary to develop a computational scheme to obtain the material parameters from measured frequencies. The procedures that have been proposed and implemented to do this generally involve defining a figure of merit F, such as the mean-square deviation of the measured from the computed resonant frequencies. Minimizing F always involves some kind of a search in the space of the relevant material parameters (elastic constants, density, dimensions, shape parameters), which requires repeated computations of the frequencies for different values of those parameters. Thus an essential part of a viable procedure is an accurate method to compute normal modes,

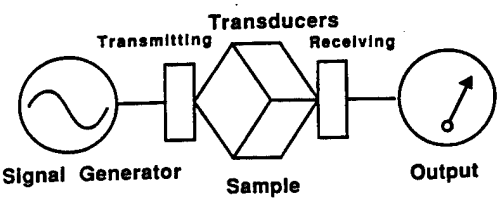


FIG. 1. Schematic resonant ultrasound setup. The sample, a rectangular parallepiped in this illustration, is suspended along a body-diagonal between transmitter and receiver transducers and the input frequency is swept through a given range.

 © 1991 Acoustical Society of America

which should be fast, since it must be performed many times. It is the main purpose of the present paper to present such a method, which works for a wide variety of systems, in a simple form which may easily be programmed for a computer.

B. Hamilton's principle

Consider an arbitrarily shaped object with (possibly position dependent) elastic tensor C_{ijkl} and density ρ with a free surface S surrounding a volume V. One forms the kinetic energy density $\mathrm{KE} = 1/2\rho\omega^2 u_i u_i$, the potential energy density $\mathrm{PE} = 1/2 C_{ijkl} u_{i,j} u_{k,l}$, takes their difference, and integrates to get the Lagrangian

$$L = \int_V (\mathrm{KE} - \mathrm{PE}) dV. \tag{1}$$

Here u_i is the ith component of the displacement vector, we use the usual summation convention, and indices following a comma denote differentiation with respect to that coordinate. As usual, we have assumed that the displacements' time dependence is $e^{i\omega t}$.

The following felicitous facts facilitate the computation. If one allows u_i to vary arbitrarily in the volume V and on the surface S, viz., $u_i \to u_i + \delta u_i$, and asks what the variation in L is, the answer is, as the interested reader can easily verify (one integration by parts is needed),

$$L \to L + \delta L, \tag{2}$$

where

$$\delta L = \int_V (\text{elastic wave equation})_i \, \delta u_i \, dV + \int_S (\text{surface traction})_i \, \delta u_i \, dS, \tag{3}$$

plus terms of higher order in δu_i. The elastic wave equation is

$$\rho\omega^2 u_i + C_{ijkl} u_{k,lj} = 0 \tag{4}$$

and the ith component of the surface traction vector is

$$n_j C_{ijkl} u_{k,l}, \tag{5}$$

where $\{n_i\}$ is the unit outer normal to S.

So, apparently due to a mathematical fortuity that may have occurred during a lapse in Murphy's vigilance, the displacement vectors u_i, which are solutions to the elastic wave equation with free boundary conditions on S, are just those points in function space at which L is stationary. This fact suggests the following procedure for obtaining the solutions.

C. Choice of basis

If the computation is to be implemented on a computer, and is to work for a variety of shapes, one should, for reasons soon to become apparent, expand the displacement vector in the simplest possible complete set of functions. There is none simpler than powers of the Cartesian coordinates, so we choose the following set of basis functions:

$$\Phi_\lambda = x^l y^m z^n, \tag{6}$$

where $\lambda = (l,m,n)$ is the function label, a set of three non-negative integers. In terms of this basis we expand the displacements on a truncated set Ω [specified below in (9)]

$$u_i = \sum_{\lambda\in\Omega} a_{i\lambda} \Phi_\lambda, \quad i = 1,2,3. \tag{7}$$

After substituting this into the above expression for the Lagrangian L, the latter can be written (a becomes a column vector)

$$L = 1/2\omega^2 \mathbf{a}^T E \mathbf{a} - 1/2 \mathbf{a}^T \Gamma \mathbf{a}, \tag{8}$$

where E and Γ are matrices whose order R is determined by Ω, which is specified by the condition

$$l + m + n \le N, \tag{9}$$

viz. $R = 3(N+1)(N+2)(N+3)/6$. [The initial factor of 3 here corresponds to the three coordinates x, y, and z. The other factors are the number of ways (9) can be realized for non-negative l, m, and n.] To keep the size of the matrices within manageable limits we must clearly be restrained in our choice of N.

The matrix E has elements

$$E_{\lambda\mu\cdot i} = \delta_{ii} \int_V \Phi_\lambda \rho \Phi_\lambda \cdot dV. \tag{10}$$

If we had chosen $\{\Phi_\lambda\}$ to be a set orthonormal with respect to the weight function ρ, then E would have been the unit matrix, which would have simplified some subsequent manipulations. But then we would have had to choose a different $\{\Phi_\lambda\}$ for each shape, and for each different density ρ, which can be a function with arbitrary position dependence, and we would not be able to express the required matrix elements which follow in closed form. The moderate penalty exacted by a nondiagonal E is more than compensated by the other advantages concomitant with (6).

The matrix Γ has elements

$$\Gamma_{\lambda\mu\cdot ij} = C_{ijij} \int_V \Phi_{\lambda,j} \Phi_{\lambda\cdot,j} \, dV. \tag{11}$$

Again, the volume integrals that appear here are quite tractable for many shapes V if the choice (6) is made, which is not the case for other basis sets.

D. Generalized eigenvalue equation

The expression (8) for the Lagrangian (1) is stationary, according to (3), for the displacements u_i that are solutions of the free-vibration problem. So the solutions may be obtained by setting the derivatives of (8) with respect to each of the R amplitudes $a_{i\mu}$ equal to zero. This yields the following matrix eigenvalue equation:

$$\omega^2 E \mathbf{a} = \Gamma \mathbf{a}. \tag{12}$$

The matrix E, although not diagonal, as it would be if we had chosen an orthonormal basis, is symmetric and positive definite, and Γ is symmetric, so that a standard eigenvalue–eigenvector subroutine package (RSG in EISPACK[13,14]) can be used to solve (12) as easily and nearly as quickly as it could if E had been diagonal.

II. EXAMPLES

A. The potato

For our choice of $\{\Phi_\lambda\}$ the matrix elements of E and of Γ are all of the form

$$f(p,q,r) = \int_V x^p y^q z^r \, dV, \tag{13}$$

where p, q, and r are non-negative integers. This integral can be evaluated analytically for a variety of V's, including one which we call the potato, for want of a better name. The potato is constructed by choosing six different semiaxes $(d_{1+}, d_{1-}, d_{2+}, d_{2-}, d_{3+}, d_{3-})$, and filling each octant of the Cartesian coordinate system (x,y,z) with a (different) ellipsoidal segment. Figure 2 shows a contour plot of the upper half of a typical potato with semiaxes given below.

The integral (13) for this shape is just the sum over the octants of terms of the form

$$f_8(p,q,r) = \{1,\pi/2\} d_{1+}{}^{p+1} d_{2+}{}^{q+1} d_{3+}{}^{r+1} \times (p-1)!!(q-1)!!(r-1)!!/(p+q+r+3)!!, \tag{14}$$

which is the integral (13) on the $(+++)$ octant. (It is an integral representation for the beta function.[15]) The curly bracket is unity if two or three of the integers p, q, and r are odd; otherwise it is equal to $\pi/2$. The double factorial is, as usual, the product of alternate positive integers up to the argument, with $(0)!! = (-1)!! = 1$. The integrals on the other octants have the same form, but have a sign that depends in an obvious way on whether p, q, and r are even or odd.

1. Comparison with measured potato frequencies

Table I lists some of the computed frequencies ($f = \omega/2\pi$) of the normal modes of an aluminum half-potato with measured density $\rho = 2.7065$ g/cm^3 and Lamé parameters $\lambda = 0.54\times10^{12}$, $\mu = 0.27\times10^{12}$ erg/cm^3 (Poisson's ratio $= 1/3$) appropriate to Al ($C_{11} = \lambda + 2\mu$, $C_{12} = \lambda$, $C_{44} = \mu$, using the Voight contracted notation for the elastic constants), and semiaxes $(d_{1+},\ldots,d_{3-})$ $= (2,5,4,1,3,0)$ in. We have adjusted λ, but not Poisson's ratio, so that the average of the lowest ten frequencies is equal to that measured. Omitted are the zero frequencies, of which there are always six, corresponding to three translations and three rotations. These frequencies were computed by solving the generalized eigenvalue problem (12) with $N = 6$, 7, 8, and 9.

TABLE I. Measured and calculated Al potato frequencies (Hz).

Mode	Measured	$N=6$	$N=7$	$N=8$	$N=9$
1	9914	9925	9923	9922	9922
2	10391	10413	10410	10409	10408
3	13513	13484	13481	13480	13480
4	15802	15860	15841	15837	15836
5	17064	17116	17091	17085	17083
6	17285	17270	17268	17268	17268
7	17892	17903	17885	17881	17880
8	19225	19244	19195	19184	19181
9	20999	21060	21045	21038	21037
10	21800	21936	21842	21815	21808
20	28933	29258	29038	28985	28968
30	34866	36230	35423	35077	34958
Cray-1 CPU time		3s	7s	15s	29s
		$R=252$	$R=360$	$R=495$	$R=660$

Comparison of the computed frequencies in the corresponding columns shows that $N = 6$ gives 1/2% or better accuracy for the lowest ten lines, and larger N's give much better accuracy. (Notice that the rows are monotonically decreasing sequences, as is required by Cauchy's inequality.[16]) Convergence is slower for this shape, which has sharp edges, than for smooth-surfaced objects. It should be emphasized that, although this potato is elastically isotropic, it is no more difficult and takes little more computer time to calculate the frequencies (and eigenfunctions) for the general elastic potato with 21, or even 81 independent C_{ij}'s.

Figure 3 is a photograph of this aluminum potato, which was fabricated using a numerically controlled ball mill. The reason that $d_{3-} = 0$ is that the half-potato has a plane surface for measurement references, which makes it easier to machine. The surface quality is 32 μin. rms, which is also a limit on the accuracy with which it reproduces the nominal composite-ellipsoidal shape.

Transducers were cemented to the flat side of this object, and its resonant frequencies were measured.[17] They are listed in the second column of Table I. The measurement errors, as gauged by the scatter in resonant frequencies when the suspension and transducer positions are changed, are comparable to the rms difference between measured and calculated frequencies, which is 0.16% for the lowest ten resonances. The above choice of λ minimizes the error, and probably amounts to as good a determination of the elastic constants of this Al alloy as is available.

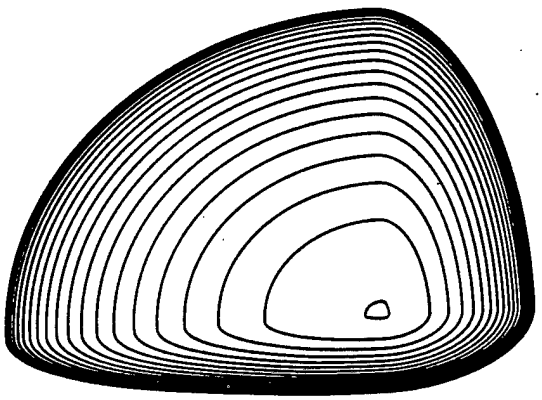


FIG. 2. Elevation contours of the half-potato described in the text. The bottom of this object is flat. It comprises four ellipsoidal segments.

B. The tetrahedron, the octahedron, and the prism

Another object of low symmetry is a tetrahedron whose faces comprise the yz, xz, xy planes and the plane

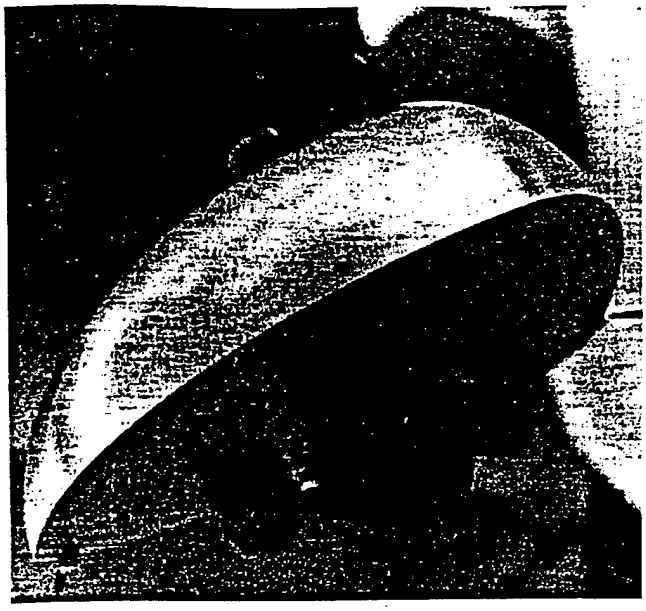


FIG. 3. Photograph of the machined aluminum half-potato that was made by a numerically controlled ball mill and whose ultrasonic resonant frequencies were measured. This object weighs 2438.4 g, and its density is 2.7065 g/cm^3. Two 1/4 in.-diam transducers cemented to the flat side are visible here.

$x/a + y/b + z/c = 1$. This is the so-called corner prism, and one can easily[18] work out (13) for this shape. It is

$$f(p,q,r) = a^{p+1}b^{q+1}c^{r+1}p!q!r!/(p+q+r+3)!. \quad (15)$$

This prism occupies the $(+++)$ octant and, just as the potato was formed from eight ellipsoidal segments, one can make an octahedron by piecing together eight of these prisms.

The cylindrical prism bounded by the planes $x=0$, $y=0$, $z=0$, $z=c$, and $x/a + y/b = 1$ has

$$f(p,q,r) = a^{p+1}b^{q+1}c^{r+1}p!q!/(p+q+2)!(r+1), \quad (16)$$

and it is possible, albeit sometimes tedious and not always easy, to work out many other polyhedric shapes.

III. SYMMETRIES

The potato is an object of low symmetry. If we go to the other extreme, the ellipsoid (which is the potato for $d_{i+} = d_{i-}$, $i = 1,2,3$), we can speed up the calculation immensely by exploiting the threefold inversion symmetry $x \to -x$, $y \to -y$, $z \to -z$, which then obtains. Now we require that the symmetry of the crystal be orthorhombic or better (i.e., that off-diagonal elements $C_{ij} = 0$ if either i or j is 4, 5, or 6). Then, by inspection of the PE in (1) (see Ohno[4]) we see that, if u_x is characterized by a parity triplet $(-\lambda,\mu,\nu)$ [where $\lambda = (-1)^l$, $\mu = (-1)^m$, $\nu = (-1)^n$], then the Γ matrix only connects this u_x with u_y and u_z having the following parities:

$$u_x:(-\lambda,\mu,\nu), \quad u_y:(\lambda,-\mu,\nu), \quad u_z:(\lambda,\mu,-\nu). \quad (17)$$

Therefore, the Γ matrix degenerates into a block-diagonal matrix with eight blocks, each characterized by one parity triplet, say the parity of u_x. We label this parity as follows:

$$\begin{array}{lccccccccc} k= & 1 & & 2 & & 3 & & 4 & & \\ (-\lambda,\mu,\nu)= & (+,+,+) & & (+,+,-) & & (+,-,+) & & (+,-,-) & & \\ & & 5 & & 6 & & 7 & & 8, & \\ & & (-,+,+) & & (-,+,-) & & (-,-,+) & & (-,-,-). & \end{array} \quad (18)$$

Because of the high symmetry the modes for each k value (18) are uncoupled (i.e., k is a "good quantum number"), and with $N = 7$, as in the potato, the largest order for which we have to solve the eigenvalue problem (12) is $R = 60$, in contrast to $R = 360$ to get the same accuracy in the potato, although it must be added that now we need to solve an eigenvalue problem for each of the eight k values.

The layered sphere (or ellipsoid), which might be used as a model for oscillations of the earth, may also be studied using linear combinations of (14) (one term for each layer in the model) to represent $f_s(p,q,r)$. We have computed frequencies by this method for a model consisting of a core plus a crust (both elastically isotropic) and have found, by comparing the results with an exact calculation using spherical Bessel functions and spherical harmonics (eigenfunctions of the elastic wave equation in spherical coordinates), that the accuracy of the present method is excellent even for small N if the mismatch of the acoustic impedances at the boundary is not too large.

IV. OTHER EXAMPLES WITH THREEFOLD INVERSION SYMMETRY

A. The rectangular parallelepiped

The $2d_1 \times 2d_2 \times 2d_3$ rectangular parallelepiped is the shape considered by Demarest[3] and by Ohno[4] (both of whom used Legendre polynomials for their basis set $\{\Phi_i\}$) The numerical procedure to be followed here is identical to that for the ellipsoid. Only (14) is changed; it is now

$$f(p,q,r) = 8d_1^{p+1}d_2^{q+1}d_3^{r+1}/(p+1)(q+1)(r+1). \quad (19)$$

B. The cylinder

The right circular (or elliptical) cylinder also has threefold inversion symmetry. Again the numerical procedure is the same as before with

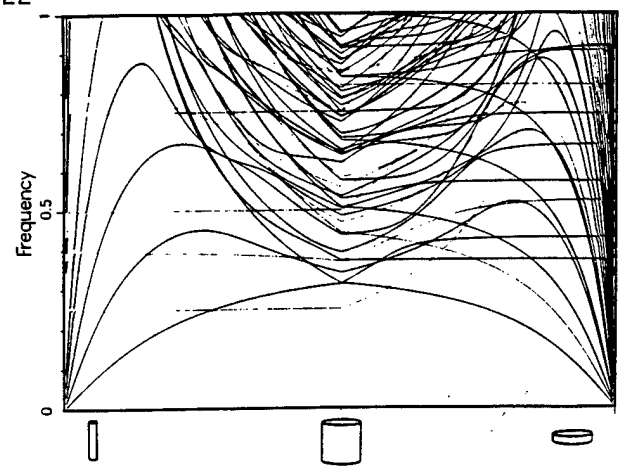


FIG. 4. Low-lying resonant frequencies of a family of cylinders. On the left half of the plot, the height of the cylinder is held constant at 2, and the aspect ratio (diameter/height) is varied linearly from 0 at the origin to 1 at the center of the plot. Then the diameter is held constant at 2 on the right half, and the height is decreased to 0 at the right end of the abscissa. There are a number of noteworthy features on this plot. First, on the left, there are modes whose frequencies are independent of the cylinder's diameter for any aspect ratio. These have frequencies that are integral multiples of $(\mu/\rho)^{1/2}/2h = 1/4$, where h is the height of the cylinder. They are the torsional modes[19] with h equal to multiples of half-wavelengths. Second, also on the left, there are two modes whose frequencies are independent of diameter as the diameter becomes small. These are the compressional "Young's modulus" modes with frequencies integral multiples of $(E/\rho)^{1/2}/2h = 0.395$, where E = Young's modulus $= \mu(3\lambda + 2\mu)/(\lambda + \mu) = 5/2$. Third, on the right, one finds two modes with frequencies independent of h even for large h. Their frequencies are close (but not within our computational uncertainty) to being given by $J_0(2\pi f) = 0$. Finally, on the right again, there are many modes whose frequencies become independent of h as h becomes small. They must be describable as vibrations of an elastic membrane with free edges, and might therefore be analytically tractable.

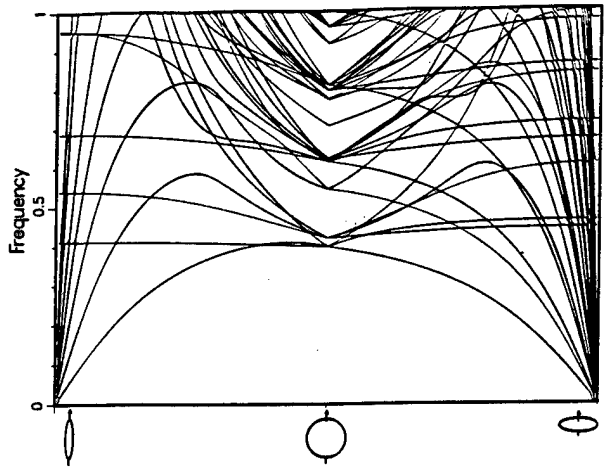


FIG. 5. Like Fig. 4, but for a family of spheroids. The sphere of unit radius is at the center, prolate spheroids with radius proportional to the abscissa on the left, culminating with the needle of length 2. To the right of center are oblate spheroids with unit radius ending with the zero-thickness pancake. In contrast to the cylinder, there are no frequencies here that are trivially calculable.

$$f(p,q,r) = 4\pi\, d_1{}^{p+1}\, d_2{}^{q+1}\, d_3{}^{r+1} \times (p-1)!!(q-1)!!/(r+1)(p+q+2)!!, \tag{20}$$

where $2d_3$ is the height and d_1, d_2 are the semiaxes of the elliptical cross section of the cylinder. We have computed the resonant frequencies for cylinders with various aspect ratios. The results are exhibited in Fig. 4, in which we have taken $N = 8$, unit density, and isotropic elasticity with unit Lamé parameters.

C. The spheroid

Very similar to the cylinder spectrum is that of the spheroid. Here, $f(p,q,r)$ is given by (14) for this shape, and its spectrum as a function of aspect ratio is shown in Fig. 5, with the prolate limit on the left, the sphere in the center, and the oblate limit on the right. Differences that stand out between this plot and Fig. 4 for the cylinder are: (1) the massive degeneracy of resonances in the spherical case and (2) the fact that many of the cylinder modes are rigorously independent of aspect ratio, not the case for the spheroid.

V. LESS SYMMETRIC OBJECTS

Intermediate between the potato and the ellipsoid, as far as symmetry is concerned, are the following objects that break the $z \to -z$ symmetry while retaining the x and y symmetries. Now the Γ matrix does not break up into eight diagonal blocks, as it did in for the ellipsoid, but only into four, since z parity is no longer conserved. From (16), we see that $k = 1$ mixes with $k = 2$, $k = 3$ mixes with $k = 4$, $k = 5$ with 6, and 7 with 8, so that the diagonal blocks can be labeled $k = 1$, 3, 5, and 7.

A. The sandwich

By the sandwich, we mean a rectangular parallelepiped in which there is a material discontinuity in the z direction, as illustrated in Fig. 6. For $z < b$ the elastic constants and the density are C_{ij}^{-}, ρ^{-}, and for $z > b$ they are C_{ij}^{+}, ρ^{+}. The changes in $f(p,q,r)$ are trivial and obvious. Some frequencies for this system are illustrated in Fig. 7, in which the abscissa

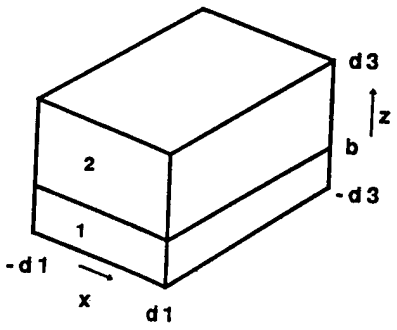


FIG. 6. A rectangular parallepipedal open-faced sandwich. Material 1 occupies the part of V with $z < b$, material 2 occupies the part with $z > b$. The origin is at the center of the block.

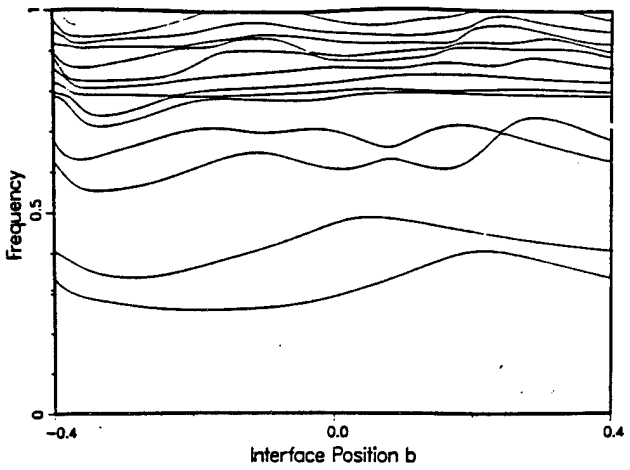


FIG. 7. The $k = 7$ normal-mode frequencies of the object illustrated in Fig. 6. The densities and Lamé parameters of the two materials are related as discussed in the text, so that the frequencies are identical at the right and left extremes, where the object is all stiff and heavy versus all soft and light. It is interesting that the third and fourth resonances cross once, and avoid crossing twice, so that the symmetries of these modes apparently change as b does. Here $(d_1,d_2,d_3) = (0.2,0.3,0.4)$, $(\rho,\lambda,\mu) = (2,0.4,0.4)$, $(10,2,2)$ in the bottom and top parts, respectively.

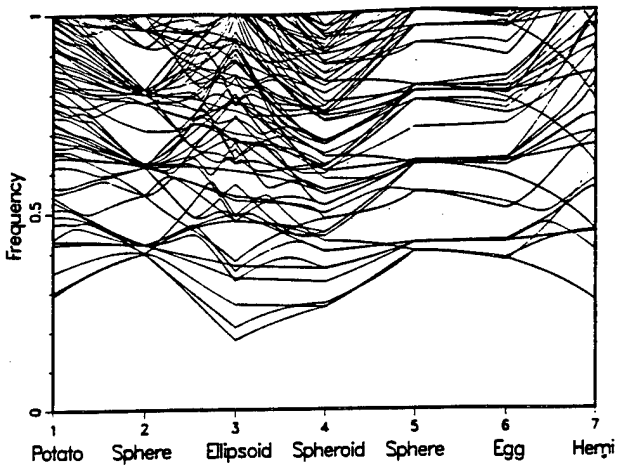


FIG. 8. Frequency spectra of a number of objects in the potato family. The seven stations correspond to shapes as labeled, with semiaxes as given in Table II. The sphere frequencies agree well with those in the literature for these material parameters (Poisson's ratio = 1/4).[20] The dimensional parameters $d_{1+},d_{1-},\ldots,d_{3-}$ are interpolated linearly between the seven stations here. Several interesting features invite comment. First, the potato has no degenerate lines, because of its low symmetry, and the sphere, conversely, has few lines that are nondegenerate. The ellipsoid has no degeneracies, and the spheroid, the egg, and the hemisphere (all being rotationally symmetric) do, but never more than doubly degenerate lines. Small deviations from the sphere in the egg direction do not change any of the frequencies to first order, because d_{3+} increases as much as d_{3-} decreases, compensating one another as far as affecting resonant frequencies is concerned. As in several other figures, apparent avoided crossings on this plot should be viewed with suspicion because the plotting program does not interchange line identities when physically the modes do, in fact, cross. Spectra are computed for 241 abscissa values here and elsewhere, which sets the scale on which avoided crossings may be spurious.

is b, the z position of the interface. We have chosen the densities and elastic constants to scale together; i.e., $\rho^+ = \alpha\rho^-$ and $C_{ij}^+ = \alpha C_{ij}^-$ with $\alpha = 5$, so that the frequencies, when the block is all one material, are identical to those when it is all the other material, although they vary considerably for intermediate compositions, and the identities of some of the modes are exchanged in the process.

B. The egg, the hemisphere, and the bell

The egg is a special case of the potato with $d_{1+} = d_{1-}$, $d_{2+} = d_{2-}$. It has twofold inversion symmetry; whereas the degeneracies of the spherical case are broken completely by the potato (as well as by the ellipsoid), they are only partly broken by the egg and the spheroid.

The hemisphere is a special case of the egg with $d_{3-} = 0$. The bell is a hemispheroid with a concentric hemispheroidal hole in it, formed by superimposing on the original hemispheroid a smaller one with negative density and elastic constants.

In Fig. 8 are shown resonant frequencies for a family of shapes: the potato, the ellipsoid, the spheroid, the sphere, the egg, and the hemisphere. The stations along the abscissa are labeled: The semiaxes characterizing the various shapes are listed in Table II. These parameters are linearly interpolated between the stations. Isotropic material with unit density and Lamé parameters is again assumed.

To illustrate the possibility of tuning the timbre of a bell by adjusting the bell's proportions, we show, in Fig. 9, the spectra of a family of bells. The abscissa is the logarithm of the aspect ratio $2d_1/d_{3+}$, which varies from 1/10 at the extreme left to unity in the middle (half a 2:1 prolate spheroidal shell) to 10 at the right-hand side. Thus, on the left, we have a chimelike object, and on the right we have something close to a cymbal. The semiaxes of the inner surface are 0.9 of those for the outside surface.

C. The cone and the pyramid

The cone fits easily into our recipe if we take its point to be at $z = 0$, and its base to be at $z = d_3$. Then, the cross sections of the cone parallel to the xy plane will be ellipses whose semiaxes are proportional to z. For this object, we find

$$f(p,q,r) = 2\pi d_1{}^{p+1} d_2{}^{q+1} d_3{}^{r+1} \times (p-1)!!(q-1)!!/(p+q+2)!!(p+q+r+3). \quad (21)$$

TABLE II. Semiaxes for the objects in Fig. 8.

Object	d_{1+}	d_{1-}	d_{2+}	d_{2-}	d_{3+}	d_{3-}
1. Potato	0.25	1.0	0.5	1.25	0.75	1.5
2. Sphere	1.0	1.0	1.0	1.0	1.0	1.0
3. Ellipsoid	0.5	0.5	1.0	1.0	1.5	1.5
4. Spheroid	1.0	1.0	1.0	1.0	1.5	1.5
5. Sphere	1.0	1.0	1.0	1.0	1.0	1.0
6. Egg	1.0	1.0	1.0	1.0	1.5	0.5
7. Hemisphere	1.0	1.0	1.0	1.0	1.0	0.0

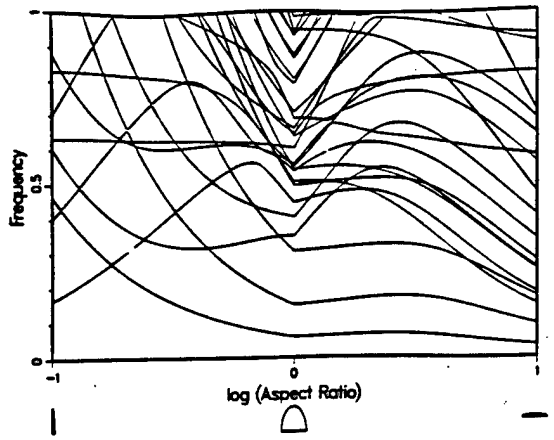


FIG. 9. Spectra of an idealized family of bells. These are hemispheroidal shells, with inner semiaxes always 0.9 of the outer ones. The material is our standard isotropic elastic medium with $\lambda = \mu = \rho = 1$. The abscissa is $\log_{10}$ (bell diameter/bell height). The height is kept constant at unity on the left half of the plot, and the diameter is kept constant on the right half. It should be noted that one cannot tell what a bell will sound like by examining a plot like this, because, for one thing, the extent to which a given mode will be excited depends on its amplitude at the clapper strike point. This information is available to us, but is more labyrinthine than the frequencies. Many other factors also enter, such as the Q's of the modes and their coupling to the air.

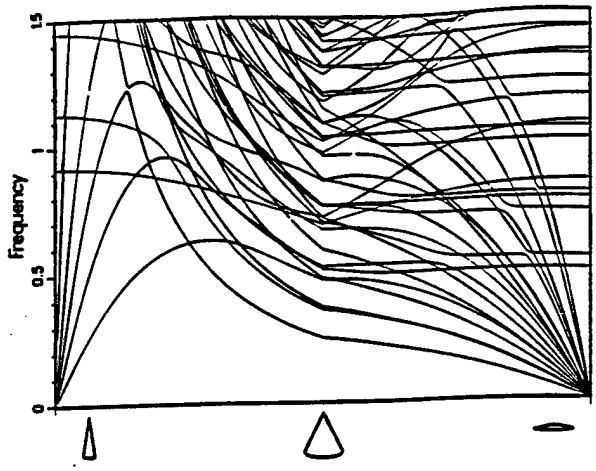


FIG. 10. These are the spectra for a family of right-circular cones. The height h is 2 at the left extreme, and the diameter is 0. The diameter increases linearly along the abscissa while h remains constant until at the center the aspect ratio is 1. Then the diameter is held constant at 2 while h decreases to 0 at the right extreme. There are some intriguing features on this graph, namely that at both the needle limit and the discus limit some of the frequencies are independent of the vanishing dimension, leading one to expect that a simple analytic description should be possible.

Likewise, the pyramid has its point at $z=0$ and its $2d_1 \times 2d_2$ base at $z=d_3$. For it,

$$f(p,q,r) = 4d_1{}^{p+1}d_2{}^{q+1}d_3{}^{r+1}/(p+q+r+3)(p+1)(q+1). \quad (22)$$

Figure 10 displays the resonant frequencies for a cone made of our standard isotropic material with $\lambda=\mu=\rho=1$.

D. The cylinder with a skewed end

A cylinder bounded by the ellipse $(x/d_1)^2+(y/d_2)^2=1$ and the planes $z=0$ and $z=d_3(1+\alpha x/d_3)$ has

$$f(p,q,r) = 2\pi(q-1)!!d_1{}^{p+1}d_2{}^{q+1}d_3{}^{r+1}g(p,q,r), \quad (23)$$

where

$$g(p,q,r) = \sum_{m=0}^{r+1} (\alpha d_1/d_3)^m r!(p+m-1)!! \times [m!(r+1-m)!(p+q+m+2)!!]^{-1} \quad (24)$$

and

$$\alpha = \tan\theta$$

is the slope of the top of the cylinder, which makes an angle θ with the horizontal. Note that this object breaks both the $x \to -x$ and $z \to -z$ symmetries, so the Γ matrix comprises only two diagonal blocks.

VI. ANISOTROPY

Since the word "anisotropic" appears in the title of this paper, it is fitting that we give some results for systems with anisotropic elastic constants. The simplest anisotropic material has transverse isotropy, or hexagonal symmetry. There are now five independent elastic constants. If the sixfold axis is in the z direction they are C_{33}, C_{12}, $C_{13}=C_{23}$, $C_{44}=C_{55}$, and C_{66}, with $C_{11}=C_{22}=C_{12}+2C_{66}$. To illustrate the effects of anisotropy, we will use a set of elastic constants, parametrized with a number g, close to our standard isotropic set, namely,

$$C_{11}=C_{22}=3, \quad C_{33}=3g, \quad C_{12}=C_{66}=1, \quad C_{13}=C_{23}=C_{44}=C_{55}=g, \quad (25)$$

when the symmetry axis is in the z direction.

A. The anisotropic cone

To illustrate the effects on the spectrum of variation of the magnitude of the anisotropy, we consider a right circular cone with unit aspect ratio. In Fig. 11 are shown the normal-mode frequencies of this object as functions of anisotropy. Here, $g=1$ at the center, where the material is isotropic, and g increases in both directions away from the center. On the left half of the plot the sixfold axis is the z axis, so rotational symmetry about the z axis obtains, and is evidenced by many twofold degeneracies. On the right half the sixfold axis is the x axis (so the elastic constants are obtained from those above by interchanging 1 with 3 and 4 with 6).

B. The anisotropic spheroid

To illustrate the effects of varying the direction of anisotropy with a fixed magnitude, we consider an oblate spheroid with aspect ratio 2. In Fig. 12 is shown the vibrational spectrum of this object as a function of angle between the spheroid axis and the sixfold axis, when $g=3.5$. To compute these curves one rotates the elastic tensor C_{ijkl} about the y axis through angles given by the value of the abscissa. In the

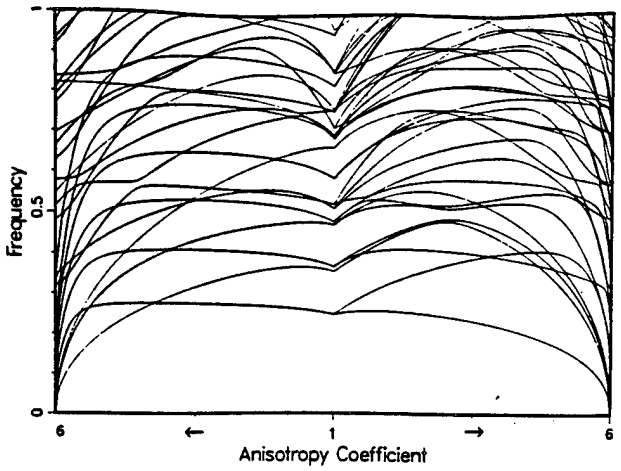


FIG. 11. These are the spectra for a right-circular cone with unit aspect ratio as functions of the anisotropy parameter g discussed in the text. At the center ($g = 1$) the material is elastically isotropic. As one goes to the left it has hexagonal symmetry with the sixfold axis in the z direction, so that rotational symmetry about the z axis holds, and the anisotropy parameter increases linearly to 6 at the left extreme. As one goes to the right from the center the sixfold axis is in the x direction, so the system lacks rotational symmetry. The degeneracies characteristic of rotational symmetry disappear when one crosses the center from left to right. Another interesting point is that some frequencies become imaginary at $g = 6$. The condition that the energy of the system be positive definite is that $\det|C_{ij}| > 0$. One can easily compute $\det|C_{ij}| = 4g^3(6 - g)$, so one would expect catastrophe not only for $g > 6$, but also, not unexpectedly, for $g < 0$.

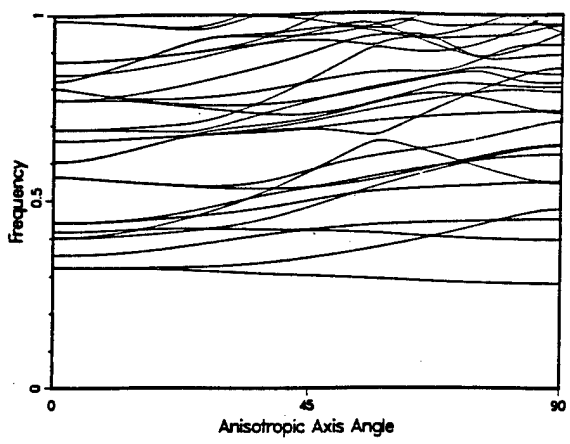


FIG. 12. These are the spectra for a 2:1 oblate spheroid composed of the same material as that in the preceding figure, i.e., an anisotropic hexagonal elastic tensor, with $g = 3.5$. The abscissa is the angle between the axis of the spheroid and the sixfold axis: They are parallel on the left, perpendicular on the right. Most of the states are twofold degenerate when the axes are parallel: Otherwise there are no degeneracies.

process, of course, one loses the advantages of threefold symmetry of the spheroid, and the computation takes as long as that for the anisotropic cone.

VII. DISCUSSION

Our computational scheme, which we call the *xyz* algorithm on account of Eq. (6), its basic distinction from the Demarest[2] algorithm, enables the calculation of normal modes for many classes of systems that have not been computed before by any method (even those as simple as the isotropic spheroid or cone, let alone the anisotropic potato), and could be calculated only with great difficulty by existing methods. The *xyz* algorithm, by its nature, is no more difficult to apply to general anisotropic systems with 21 (or even 81, the total number of elements of C_{ijkl}) independent elastic constants than to isotropic ones, and allows one easily to compute a wider variety of shapes than any other.

Although we have displayed only the resonant frequencies that these calculations have yielded, the same EISPACK subroutines also give the eigenvectors a_{μ} in (7). Thus the elastic displacements are known, and with enough persistence and determination, we could have shown vibrational shapes for each of the modes. But there are too many of them, so we will show none.

We believe we have demonstrated that the *xyz* algorithm is capable of quickly and accurately yielding the low-lying resonant frequencies for a wide variety of shapes with inhomogeneities and few restrictions on anisotropy. But perhaps its greatest virtue is its extreme simplicity when compared to other schemes that are less versatile, accurate, or speedy.

ACKNOWLEDGMENT

This work was supported by the USDOE.

APPENDIX

Here we give the listing of a Fortran program that will compute the normal-mode frequencies (and, with trivial modifications, the eigenfunctions, too) for all the objects we have discussed in this paper. This code does not exploit any of the symmetries of shape or elasticity tensor, and is therefore much slower than one that does for objects with symmetries, as is discussed in Sec. III.

```
      The program is named for the basis functions it uses.
      PROGRAM XYZ
            The arrays are dimensioned 252 here, sufficient for N=6.
      DIMENSION GAMMA(252,252),E(252,252),W(252),FV(252),FW(252),
     & C(3,3,3,3), LB(252),MB(252),NB(252),IC(252)
            The data following is the elastic tensor C_ijkl for our standard
            isotropic material. Any homogeneous anisotropy can be
            described by simply changing these data to include up to 81
            different elastic constants, for a general substance in the
            presence of magnetic fields.
      DATA C/3.,3*0.,1.,3*0.,1.,0.,1.,0.,1.,7*0.,1.,3*0.,1.,3*0.,1.,0.,1.,5*0.,1.,
     & 3*0.,3.,3*0.,1.,5*0.,1.,0.,1.,3*0.,1.,3*0.,1.,7*0.,1.,0.,1.,0.,1.,3*0.,1.,3*0.,3./
      DATA RHO/1./
      TWOPI=2.*ACOS(-1.)
      PRINT*,'PLEASE INPUT NN'
      READ*,NN
            The next 13 lines assign an index IG to each basis function (6).
      IG=0
      DO 1 I=1,3
      DO 2 L=1,NN+1
      DO 2 M=1,NN+1
      DO 2 N=1,NN+1
      IF(L+M+N.GT.NN+3) GO TO 2
      IG=IG+1
      IC(IG)=I
      LB(IG)=L-1
      MB(IG)=M-1
      NB(IG)=N-1
2     CONTINUE
1     CONTINUE
      NR=IG
```

```
*     In the next 11 statements the elements of the E and Γ
*     matrices are computed.
      DO 3 IG=1,NR
      DO 3 JG=IG,NR
      I=IC(IG)
      J=IG(JG)
      LS=LB(IG)+LB(JG)
      MS=MB(IG)+MB(JG)
      NS=NB(IG)+NB(JG)
      GAMMA(IG,JG)=
     & C(I,1,J,1)*FLOAT(LB(IG)*LB(JG))*F(LS-2,MS,NS)
     & +C(I,2,J,2)*FLOAT(MB(IG)*MB(JG))*F(LS,MS-2,NS)
     & +C(I,3,J,3)*FLOAT(NB(IG)*NB(JG))*F(LS,MS,NS-2)
     & +(C(I,1,J,2)*FLOAT(LB(IG)*MB(JG))+C(I,2,J,1)*
     & FLOAT(MB(IG)*LB(JG)))*F(LS-1,MS-1,NS)
     & +(C(I,1,J,3)*FLOAT(LB(IG)*NB(JG))+C(I,3,J,1)*
     & FLOAT(NB(IG)*LB(JG)))*F(LS-1,MS,NS-1)
     & +(C(I,2,J,3)*FLOAT(MB(IG)*NB(JG))+C(I,3,J,2)*
     & FLOAT(NB(IG)*MB(JG)))*F(LS,MS-1,NS-1)
      GAMMA(JG,IG)=GAMMA(IG,JG)
      IF(I.EQ.J) E(IG,JG)=F(LS,MS,NS)
3     E(JG,IG)=E(IG,JG)
*     The next line solves the eigenvalue problem (12) using the
*     EISPACK[13,14] subroutine RSG.
      CALL RSG(252,NR,GAMMA,E,W,0,Z,FV,FW,IERR)
*     Now obtain the frequencies from the eigenvalues. If C_ijkl is in
*     10^12 dynes/cm^2, ρ is in g/cm^3, and dimensions are in cm, then
*     frequencies W are in MHz.
      DO 4 I=1,NR
4     W(I)=SQRT(AMAX1(0.,W(I))/RHO)/TWOPI
*     The lowest 38 frequencies are printed out: the first 6 of
*     these are always zero if Γ is positive.
      PRINT*,"FREQUENCIES FOR CORNER PRISM, NN= ",NN
      PRINT 101,(W(I),I=1,36)
101   FORMAT(6G12.5)
      END
*     Next is a function subprogram for f(p,q,r) for the corner prism
*     (Eq. 15). It is straightforward here to substitute function
*     subprograms for any of the other objects we have
*     considered.
      FUNCTION F(IP,IQ,IR)
      DATA A,B,C/3*1./
      F=A**(IP+1)*B**(IQ+1)*C**(IR+1)*
     & FACT(IP)*FACT(IQ)*FACT(IR)/FACT(IP+IQ+IR+3)
      RETURN
      END
*     Factorial subprogram follows.
      FUNCTION FACT(N)
      FACT=1.
      IF(N.LT.2) RETURN
      DO 1 I=2,N
      FACT=FACT*FLOAT(I)
      RETURN
      END
```


[1] R. Holland, J. Acoust. Soc. Am. 43, 988 (1968).

[2] H. H. Demarest, Jr., Bachelor's thesis, Reed College, Portland, OR (1969).

[3] H. H. Demarest, Jr., J. Acoust. Soc. Am. 49, 768 (1971).

[4] I. Ohno, J. Phys. Earth 24, 355 (1976).

[5] Y. Sumino, I. Ohno, T. Goto, and M. Kumazawa, J. Phys. Earth 24, 263 (1976).

[6] I. Ohno, S. Yamamoto, O. L. Anderson, and J. Noda, J. Phys. Chem. Solids 47, 1103 (1986).

[7] A. Migliori, W. M. Visscher, S. E. Brown, Z. Fisk, S.-W. Cheong, B. Alten, E. T. Ahrens, K. A. Kubat-Martin, J. D. Maynard, Y. Huang, D. R. Kirk, K. A. Gillis, H. K. Kim, and M. H. W. Chan. Phys. Rev. B 41, 2098 (1990).

[8] A. Migliori, W. M. Visscher, S. Wong, S. E. Brown, I. Tanaka, H. Kojima, and P. B. Allen. Phys. Rev. Lett. 64, 2408 (1990).

[9] E. Mochizuki, J. Appl. Phys. 63, 5668 (1988).

[10] S. Isoda, H. Oda, I. Suzuki, and K. Seya, Tohoku Geophys. J. 32, 55 (1990).

[11] See the famous paper, "Can One Hear the Shape of a Drum?" by M. Kac, Am. Math. Month. 73, 1 (1966). For more recent work, see M. H. Protter, SIAM Rev. 29, 185 (1987).

[12] A. Stekel, J. L. Sarrao, T. M. Bell, Ming Lei, R. G. Leisure, W. M. Visscher, and A. Migliori (to be published).

[13] B. T. Smith, J. M. Boyle, J. J. Dongarra, B. S. Garbow, Y. Ikebe, V. C. Klema, and C. B. Moler, "Matrix Eigensystem Routines—EISPACK Guide," in *Lecture Notes in Computer Science No. 6*, edited by G. Goos and J. Hartmanis (Springer, New York, 1976).

[14] B. S. Garbow, J. M. Boyle, J. J. Dongarra, and C. B. Moler, "Matrix Eigensystem Routines—EISPACK Guide Extension," in *Lecture Notes in Computer Science No. 51*, edited by G. Goos and J. Hartmanis (Springer, New York, 1977).

[15] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (National Bureau of Standards Applied Mathematics Series No. 55, Washington, DC, 1965), p. 258.

[16] M. Marcus, *Basic Theorems in Matrix Theory* (National Bureau of Standards Applied Mathematics Series No. 57, Washington, DC, 1960), p. 12.

[17] F. Birch, J. Geophys. Res. 80, 756 (1974). This paper, concerned with vibrations of steel, glass, and rock spheres, provides a precedent for comparing measured and computed normal-mode frequencies of objects with simple geometries.

[18] This is an exercise that we leave for the interested reader. Hint: Start by evaluating the form factor $\langle \exp(k.r) \rangle$ for the tetrahedron.

[19] L. D. Landau, E. M. Lifshitz, A. M. Kosevich, and L. P. Pitaevskii, *Theory of Elasticity* (Pergamon, New York, 1986).

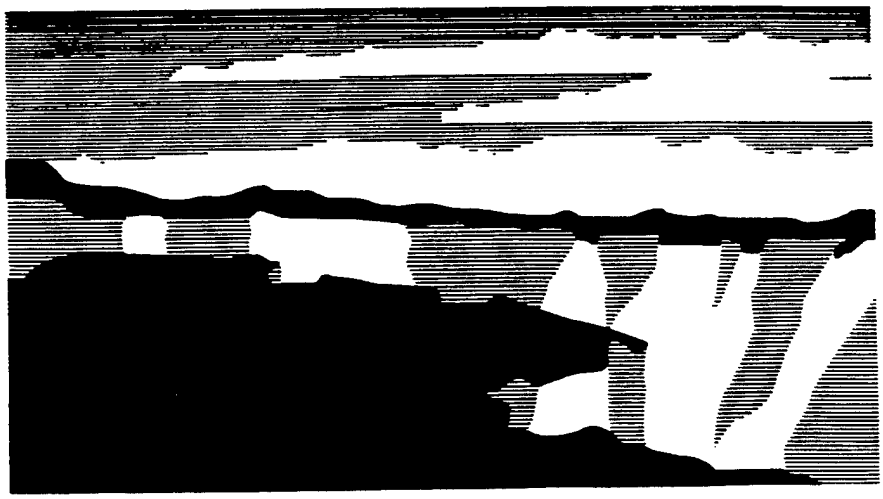
[20] H. Lamb, Proc. Math Soc. 13, 189 (1882).

LA-UR-93-3195

Title: Advanced Materials Testing Using Resonant Ultrasound Spectroscopy

Author(s):
A. Migliori
T.W. Darling
R.D. Dixon
G.W. Rhodes

Submitted to:



Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.

Los Alamos
Los Alamos National Laboratory

Advanced Materials Testing Using Resonant Ultrasound Spectroscopy

Albert Migliori, Timothy W. Darling, Raymond D. Dixon
Los Alamos National Laboratory

George W. Rhodes QUATRO Corporation

Advanced Materials Testing Using Resonant Ultrasound Spectroscopy

Albert Migliori, Timothy W. Darling, Raymond D. Dixon Los Alamos National Laboratory
George W. Rhodes QUATRO Corporation

The use of mechanical resonances to test properties of materials is perhaps older than the industrial revolution. Early documented cases of British railroad engineers tapping the wheels of a train and using the sound to detect cracks perhaps mark the first real use of resonances to test the integrity of high-performance alloys. Attempts were made in the following years to understand the resonances of solids mathematically, based on the shape and composition of an object. But Nobel Laureate Lord Rayleigh best summarized the state of affairs in his 1894 book "The Theory of Sound", where he states that "the problem has, for the most part, resisted attack". No real progress was made until electronic amplifiers were developed in the early part of this century. Then, accurate measurements of resonance frequencies could be made, and these were used crudely for some aspects of quality control. But still, no real quantitative use could be made of resonances for the determination of materials properties because only high-aspect-ratio objects such as thin plates and long thin rods of homogeneous isotropic materials were understood mathematically. The next advance was associated with the advent of modern computers. In the 1960's, Holland, and also Orson Anderson at Bell Labs, together with his student Harold Demarest developed new numerical methods that enabled the computation of the resonances of homogeneous anisotropic objects of any simple geometric shape, such as cubes, spheres, short cylinders, cones--in fact any shape that a materials scientist might use as a sample for testing. Of crucial importance was the ability to handle anisotropic materials such as single crystals, grain-oriented or textured alloys and ceramics, magnetic materials and the like. However, the electronics, and transducers were crude and intruded upon the measurement, and, surprisingly, the ability to compute the elastic properties from the resonances (the inverse problem) remained largely unsolved.

Finally, in 1987, Los Alamos began to study the new high temperature superconductors. These materials were the subject of a massive research effort because of the tremendous technological promise. Because of the powerful connection between the elastic response and the fundamental physics of superconductivity, ultrasound was a crucial tool. In fact it was ultrasound measurements that confirmed the Nobel winning theory of Bardeen, Cooper and Schreiffer that explained conventional superconductivity. But, the crystals that could be grown of these new superconductors had as many as nine independent elastic moduli, and were (and still are) available unflawed only in sub-millimeter sizes. Thus no ultrasound or other modulus measurement tool available could be applied. This was the initial motivation for LANL to attack the resonance problem.

The result of our efforts put resonances on the map. The LANL Resonant Ultrasound Spectroscopy (RUS) group developed new electronics, transducers and computational

procedures that make anisotropic modulus measurements of advanced materials routine while achieving the highest absolute accuracy for any modulus measurement technique in the world on smaller samples than can be used by any other ultrasound system, and do all of this at record speed over very broad temperature regimes. This award winning approach (1991 RD100, 1993 Federal Laboratory Award for Excellence in Technology Transfer by the LANL/QUATRO effort) is ripe for application to a broad variety of industry problems.

In addition, the mathematical models developed for such measurements also provide powerful tools to detect internal and surface materials problems in such newly important systems as stabilized ceramics, sintered alloys, ceramics--in fact in any material where advanced manufacturing techniques are still not perfected such that voids, phase separation, sintering defects, cracks and other flaws can exist. Because of the ease and speed of RUS, and the maturity of the RUS laboratory methods, it is just now becoming possible to develop non-expert systems to be used by the many small industrial efforts directed toward new materials development.

We provide here an overview of the advantages of using resonances, the unique electronics, transducers, and computational techniques, and the envelope of applicability of RUS for materials testing. Full details can be found in our review article in Physica B, 1,183(1993).

I. The advantages of Using Resonances

The elastic response of a solid is dependent on the intrinsic stiffness of the material, the shape, the density, magnetism, piezoeletric effects, the interfaces between grains or sintered particles, temperature, processing, and many other variables. Such variables are the subject of intense scrutiny in modern materials of complex microstructure that are in the first stages of development. Ultrasound has traditionally been the precision tool of choice for extraction of such information, and has been primarily used in the pulse-echo mode. In this mode, large flat transducers are very carefully bonded to the sample. Sound pulses are launched from the transducer, as narrow beams, and their transit time used to measure the speed of sound. During such a measurement the following conditions obtain.

1. The pulses are short so the electronics must be fast (and therefore expensive), and the transducers large to produce enough signal for positive detection of the brief echoes. Worse still, for most of such a measurement, no signal is present because of the brief pulses so that most of the time, the electronics are gathering in wideband background noise.

2. The arrival time of the short signal must be measured precisely. Thus if 0.1μs timing accuracy(in 0.1μs, sound travels 0.5mm in aluminum, a rather long way for precise measurements. That is, one only knows the return time to an accuracy corresponding to

0.5mm length error in the sample. Thus the sample had better be much longer than 0.5mm for any accuracy at all) is required, and the electronics must have a bandwidth of tens of megahertz, allowing even more noise.

3. The transducers must generate an accurate beam of sound and so they must be perfectly bonded to perfectly flat faces prepared on the sample. Thus for an alloy with fiber texture, such as a hard-drawn wire, five different measurements are required to determine the five moduli required for a complete description of the elastic response. The measurements are made in three perpendicular directions, requiring samples bigger than 20mm in all directions (not easy for wire) , and careful attachment and removal of transducers for each measurement.

It is no wonder then that much materials research forgoes the measurement of elastic stiffness in small samples.

In contrast, resonances circumvent ALL of the measurement difficulties associated with pulse-echo modulus measurements. Here's why.

1. To measure a resonance, one is attempting to detect the normal vibrational modes of an object. Because a vibrating object pretty much vibrates everywhere, it does not matter where the transducers attach, nor do they have to generate precise beams. Thus just weak point contact by a transducer is adequate to excite the object. No adhesives or flat faces are required.

2. To find a resonance, the excitation frequency is slowly swept. Thus signal is present at ALL times and we can take as long as we like to measure. Furthermore, because we know the instantaneous excitation frequency, we make the electronics only sensitive to that frequency, allowing almost no noise in. It is just like tuning a radio. When a resonance is reached, the sample responds strongly, greatly amplifying the signal. Overall, we can achieve a signal/noise ratio 10^6 better than with pulse-echo.

3. A 1mm cube extracted from drawn aluminum wire has enough resonances between 0.5MHz and 1.5MHz that we can easily determine all five moduli with unprecedented accuracy using much slower electronics.

The catch, of course, is that the resonances must be processed with very sophisticated algorithms to extract the moduli, whereas pulse-echo requires only a pocket calculator. Thus if we can work around the computational problem, pulse-echo becomes obsolete, as well as expensive, painful and slow.

II. Transducers and Electronics

To realize the promise of RUS, transducers had to be developed that did not produce resonances in the same range as the samples. Considering that transducers are solids

just like the samples, only careful design can prevent their resonances, essentially indistinguishable from the sample, from interfering. For larger samples (several mm) this is easy, but for the smallest ones (0.5mm), we have developed a lithium niobate/diamond composite transducer. This LANL-patented technology produces transducers without resonances to 4.3MHz. New LANL advances in solid-state bonding (patent pending) now enable us to produce such objects with no epoxies or other glues. Thus they are durable enough to survive on the foundry floor in non-expert hands, and can be used from 4K to 700K, a unique temperature range. This technology is also capable of producing highly reproducible transducers suitable for military applications such as phased sonar arrays. LANL can now produce prototypes suitable for industrial applications, and QUATRO is negotiating with LANL for commercial rights.

To drive the resonances and detect them, appropriate electronics must be used. It is simple to use very expensive commercial instruments and controllers to do this in a crude way, but with advances in digital frequency synthesizers and low noise operational amplifiers occurring in 1989 and 1991, it is possible to design a thermal-noise limited electronics package that will fit into a PC computer, making the technology more accessible and improving performance. The LANL RUS team has done just that, basing the system on plug-in modules and internally installed cards with an external low-noise receiver section. This design was licensed to QUATRO corp. who developed it, and produced user-friendly menu driven software to run it. It is now in production and is commercially available.

III. Computations

To circumvent the computational difficulties of using resonances to study materials, LANL developed new procedures for determining moduli from resonance measurements. These procedures are based on non-finite element Lagrangian minimization approaches and a Levenburg-Marquardt error minimization scheme, also patented by LANL, discussed at length in the Physica B article. A key point is that we can now run these codes in a user friendly way on an IBM PC 486-25 based machine in reasonable (40 minutes) time, instead of on the CRAY YMP used to develop them. Thus all of the technology is available for application to the small business environment.

The constraints on the applicability of RUS to a given material or object are governed by the computations. We envision three commercial modes of use for materials and processing diagnostics. They are:

1. Determination of major flaws, such as internal cracks exist in mass produced objects. In this case, a known good object, or an object before flaws appear is measured, and pattern-recognition techniques are applied to the resonances of production objects to see if they fall within an acceptability envelope. This approach is a sophisticated version of the one used by the railroad engineer, and requires no accurate computation of the

resonances of the object, but does require customization of the pattern recognition schemes for each object. This approach is being used on objects ranging from highway bridges to General Motors oxygen sensors.

2. Processing problems need to be detected in the material itself that occur on a macroscopic scale such as internal voids, inhomogeneities, cracks and the like. In this case, a full mathematical model of the objects is set up and the full force of the computations brought to bear. Flaws show up as a failure of a sample of material to fit the mathematical model. This is illustrated in the following computation on UCu_5, shown in table I. This material is of interest mostly for obscure reasons but is characterized by the undesirable growth of copper lamellae in the samples. The lamellae, produced by phase separation, make the material inhomogeneous. This, in turn, makes it impossible to produce an accurate correspondence between the model and the measurements, immediately indicating a flawed material. The flaw shows up in the large RMS (root-mean-square) error between best fit frequencies (fr) and measured ones (fex) in MHz. Note that the operator can choose how many moduli to fit, and must input the density (rho), and dimensions d1, d2, d3. The code also computes how much each resonance depends on each fitted parameter and produces other diagnostics including the estimated errors shown as the last three rows of entries. This last point is important and is one of our more important advances. The errors shown are for three different "directions" in the 5-dimensional parameter space that the fit uses. This parameter space consists of the parameters described as free moduli at the top of the table. A best fit occurs at a minimum in this space, but the minimum has different curvature for different parameters. Thus the error of 23.07 % for c_{12} reflects a very flat minimum for this parameter in one direction. The error is only 0.52% and 4.69% in other directions, but the worst case is the real test. The elastic moduli c_{ij} (dynes/cm^2), are computed based on the usual nomenclature, and displayed in orthorhombic notation only for computational reasons. Other choices are easily made such as, for example, shear modulus, Young's modulus, bulk modulus and Poisson's ratio, but these latter are less useful for anisotropic materials

3. One wishes to determine the moduli of a sample of material having no macroscopic defects to determine if processing, texturing, composition etc. are correct. In this case the code produces a very good fit, in contrast to the previous computation. The quality of the fit is the first test. If it is poor, macroscopic flaws exist and one stops there. If it is good, the material has no macroscopic flaws and is homogeneous. The fit shown is for a perfect single crystal of the ceramic component $SrTiO_3$. The RMS error is extremely small, indicating perfect material, and the three elastic moduli of this cubic material are determined with high accuracy.

The following two tables are the actual output of the code used to determine the moduli of rectangular parallelipiped samples.

1.The free moduli are those that are adjusted by the code for a best fit to the experimental frequencies. For an isotropic material, only c_{11} and c_{44} are necessary, while

for an orthorhombic material nine moduli are required. In this case, fiber texture is equivalent to an hexagonal symmetry for which the five free moduli are sufficient.

2. The order of the polynomial determines how accurately we approximate the vibrational motion on resonance. Typically a sixth order polynomial is adequate for a poor sample, while tenth order is required for a good fit. The computational time is proportional to the cube of this number, and reaches about 40 minutes tenth order on a 25 MHz 486 PC.

3. n is the mode number, fex the measure frequency, fr the best fit, %err is the percent error between fex and fr, wt is the weight used for the measured frequency in the fit. A weight of zero means that the measured frequency is ignored in determining the quality of a fit. k describes the symmetry of the vibrational motion, and can have one of eight values. i is the harmonic for a particular symmetry type. Thus a volume oscillation has k=5, and can have a fundamental (i=1) and harmonics. df/d(moduli) are numbers that determine the dependence of each frequency on each of the free moduli listed at the top of the output.

4. The numbers under c11, c22,...,c66 describe the moduli in orthorhombic basis in cgs units. Thus if the material is isotropic, a special case of orthorhombic, c11=c22=c33, c44=c55=c66, c12=c13=c23, and c11-c12=2c44. Therefore this line shows how the moduli are constrained by noting which are different, and also what are the best moduli, as determined by the fitting procedure. d1, d2, d3 are the dimensions in cm, and may differ from the input values if they are selected as free parameters by the user.

5. This fit took 8 iterations, resulted in a 1.5052% RMS error between measured and fitted frequencies, and converged because there was no change in the error on the eighth pass.

6. The length of the gradient vector is a measure of how bowl-shaped the minimum in elastic constant space was for this fit at convergence. A perfect bowl is flat (zero gradient vector) at the bottom.

7. The eigenvalues and eigenvectors describe the curvature of the "bowl". A highly curved bowl means that it is very clear where the bottom is, while a nearly flat bowl means that it is hard to determine the exact minimum accurately. Because we are fitting five moduli, the bowl is a 5-dimensional one with five different curvatures at the bottom, and five different directions for those curvatures.

8. At the bottom of the bowl, if we increase c_{33} by 1.57%, the χ^2 fit error increase by 2%, a measurable amount and a good measure of the accuracy with which c_{33} is determined. Similarly, c_{23} can shift by 23.07% before a 2% increase in the "goodness" of the fit occurs.

Table I

UCu5AM3 polycrystal
free moduli are c33, c23 ,c12 ,c44, c66
using 6 order polynomials mass= .0401 gm rho=10.478 gm/cc

n	fex	fr	%err	wt	k	i	df/d(moduli)				
1	.549700	.549789	.02	1.	4	1	.01	.00	.00	.45	.55
2	.710100	.709312	-.11	1.	4	2	.00	.00	.00	.89	.10
3	.737100	.720152	-2.30	1.	6	1	.02	-.02	.07	.26	.68
4	.744400	.767182	3.06	1.	1	1	.04	-.06	.10	.00	.91
5	.826600	.833677	.86	1.	8	2	.01	.00	.01	.93	.06
6	.868000	.862821	-.60	1.	3	2	.00	-.01	.01	.00	.99
7	.886300	.886839	.06	1.	6	2	.03	-.04	.10	.29	.62
8	.911800	.893731	-1.98	1.	2	2	.02	.00	.00	.93	.06
9	.928900	.939781	1.17	1.	5	1	.04	-.05	.07	.00	.94
10	.946100	.952608	.69	1.	2	3	.01	-.02	.07	.42	.52
11	.966400	.954632	-1.22	1.	7	1	.08	-.08	.10	.01	.89
12	.971000	1.025946	5.66	0.	8	3	.01	-.02	.07	.33	.61
13	1.048500	1.028327	-1.92	1.	5	2	.11	-.10	.10	.00	.88
14	1.110000	1.105052	-.45	1.	5	3	.21	-.17	.16	.00	.79
15	1.114400	1.127293	1.16	1.	1	2	.61	-.04	.01	.36	.06
16	1.166200	1.138763	-2.35	1.	2	4	.02	-.01	.01	.54	.44
17	1.168700	1.160718	-.68	1.	7	2	.55	-.01	.01	.37	.10
18	1.203300	1.197938	-.45	1.	4	3	.01	.00	.02	.58	.39
19	1.234600	1.214168	-1.65	1.	5	4	.57	-.02	.03	.06	.37
20	1.237500	1.245658	.66	1.	6	3	.02	-.04	.07	.31	.63

c11	c22	c33	c23	c13	c12	c44	c55	c66
1.3125	1.3125	1.5657	.2454	.2454	.1840	.4503	.4503	.5643

d1	d2	d3
.15510	.18090	.13640

loop# 8 rms error= 1.5052 %, changed by .0000 %

length of gradient vector= .000002 lambda= .00

eigenvalues	eigenvectors				
.05071	.35	.74	-.57	.02	.07
.33482	.81	.05	.55	-.11	-.13
.56130	.46	-.65	-.54	.16	.22
8.66666	-.01	-.03	-.02	-.85	.52
19.68808	-.01	.15	.30	.48	.81

chisquare increased 2% by the following % changes in independent parameters

1.57	23.07	17.33	-.23	-.13
1.28	.52	-9.58	-.19	.71
-.76	4.69	-6.08	-.08	1.10

Table II

SrTiO3 Monocrystal
free moduli are c11, c12, c44
using 10 order polynomials mass= .0140 gm rho= 5.081 gm/cc

n	fex	fr	%err	wt	k	i	df/d(moduli)		
1	.963173	.959916	-.34	1.	4	1	.01	.00	.99
2	1.121678	1.121881	.02	1.	1	1	1.13	-.29	.17
3	1.390726	1.389815	-.07	1.	6	1	1.08	-.27	.19
4	1.738738	1.740450	.10	1.	5	1	1.42	-.42	.00
5	1.762000	1.764861	.16	1.	8	1	.12	-.03	.91
6	1.767891	1.770092	.12	1.	1	2	.97	-.19	.22
7	1.812078	1.808547	-.19	1.	2	2	.34	-.09	.75
8	1.981056	1.985441	.22	1.	3	2	.82	-.25	.43
9	2.124899	2.122966	-.09	1.	7	1	1.13	-.30	.17
10	2.145693	2.137455	-.38	1.	5	2	1.33	-.36	.03
11	2.196230	2.196209	.00	1.	3	3	.49	-.05	.56
12	2.302864	2.294921	-.34	1.	4	2	.28	-.08	.81
13	2.453592	2.454139	.02	1.	5	3	1.31	-.31	.00
14	2.626425	2.627475	.04	1.	2	3	.25	-.06	.81
15	2.632761	2.629991	-.11	1.	7	2	1.23	-.38	.15
16	2.669778	2.673013	.12	1.	6	2	.50	-.15	.65
17	2.770135	2.772325	.08	1.	2	4	.25	-.05	.80
18	2.801711	2.807695	.21	1.	1	3	.47	-.12	.65
19	2.861851	2.861022	-.03	1.	8	2	.69	-.19	.50
20	2.894922	2.896918	.07	1.	4	3	.30	-.04	.74
21	3.133465	3.134292	.03	1.	1	4	.42	-.09	.66
22	3.157694	3.166496	.28	1.	3	4	.22	-.05	.83
23	3.325360	3.317645	-.23	1.	1	5	.44	-.11	.67

c11	c22	c33	c23	c13	c12	c44	c55	c66
3.1944	3.1944	3.1944	1.0471	1.0471	1.0471	1.2343	1.2343	1.2343

d1	d2	d3
.09050	.20200	.15050

loop# 4 rms error= .1822 %, changed by .0000 %

length of gradient vector= .000000 lambda= .00

eigenvalues	eigenvectors		
.00680	.65	.70	.31
1.13301	.76	-.60	-.24
3.47858	-.02	-.39	.92

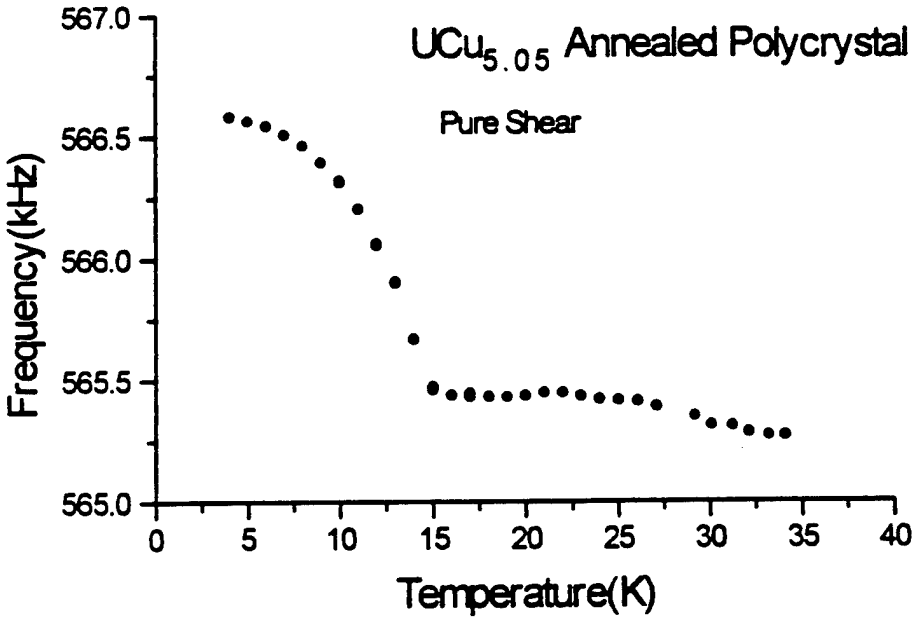
chisquare increased 2% by the following % changes in independent parameters

.47	1.66	-.03
.04	-.10	-.06
.01	-.02	.08

IV The Envelope of Applicability of RUS

In this section we simply present a sequence of RUS successes to illustrate the breadth and depth of the technique. Some of the materials used for illustration are deliberately obscure, but primarily because we wish to show limits. In Fig 1, for example, we illustrate the sensitivity of RUS to a magnetic transition in UCu_5.

Fig.1

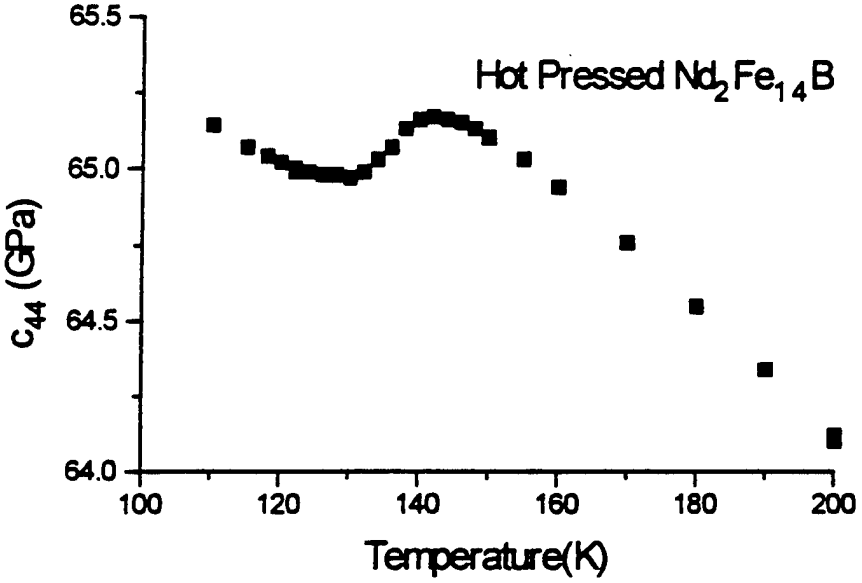


Note that a resonance frequency, proportional to 1/2 the shear modulus is plotted to illustrate the precision with which minute (0.2%) changes in mechanical properties associated with magnetism can be seen at cryogenic temperatures with RUS.

In Fig.2 we show the shear modulus of the GM developed $Nd_2Fe_{14}B$ permanent magnet material. This magnetic material is of extreme technological importance because it is the strongest permanent magnet known. As such it is used for motors and actuators where weight and efficiency are crucial, including military aircraft and GM automobiles. Because the material is not a permanent magnet until the grains are made small and lenticular, and the magnetic domains are locked with dopants, diagnosis of the magnetic properties

for production batches is important. Shown in Fig.2 is a spin-reorientation phenomena occurring at low temperatures that affects the shear modulus, and hence the room-temperature properties.

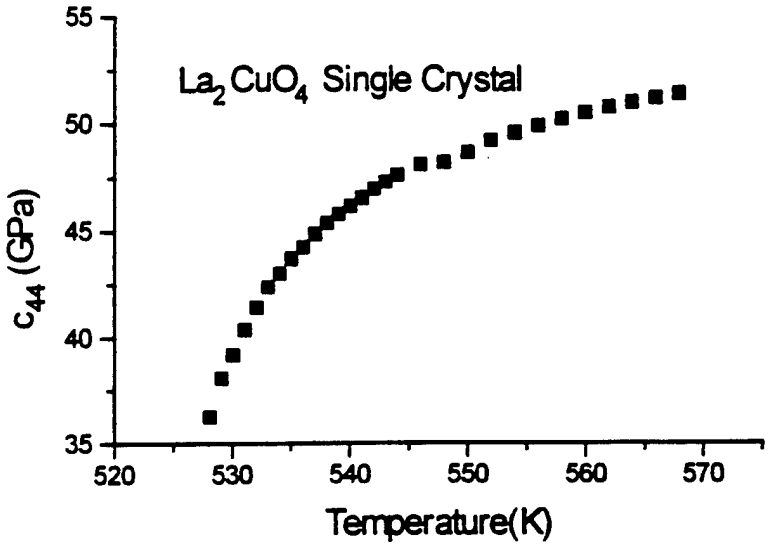
Fig.2



Structural changes are also easily detected. A crucial one occurs in the high-temperature superconductor parent material La_2CuO_4. In this material, a shear instability develops, shown in Fig.3 where c_{44} drops from 53 GPa to 36GPa, greatly changing the mechanical properties, a point of importance for wire fabrication. We use this particular illustration to emphasize the ability to make measurements at elevated temperatures. This sort of structural instability is also at least vaguely similar to those associated with martensitic changes in steel and α/β changes in Ti alloys.

We have also studied texturing in the stabilized Si_3N_4 ceramics used for bearing applications. Such materials vary widely in the dopants used for stabilization and for sintering aids, as well as in the processing of the raw material. Processing errors can render the material textured so that either it is impossible to grind a spherical ball or the ball becomes aspherical under load or at elevated temperatures. In Table III, we show a fit to the shear modulus and Poisson ratio of a perfect Si_3N_4 ball bearing. The extreme accuracy of the spherical shape produced by commercial ball grinding machines, and the lack of texturing in the blank used for this bearing produce a fit to its resonances accurate to an RMS error of 0.012%, making this the most accurate determination of moduli ever achieved by a routine modulus measurement system.

Fig.3



This 0.63500 cm diameter Si_3N_4 ceramic sphere has a density of 3.2325 gm/cm^3. The fit for μ=1.2374x10^{12} dyne/cm^2 and σ=0.2703 has a χ^2(%)=0.0124. This is sufficient to determine μ to about 0.01% and σ to about 0.05%. There are no corrections so these values are absolute.

Table III

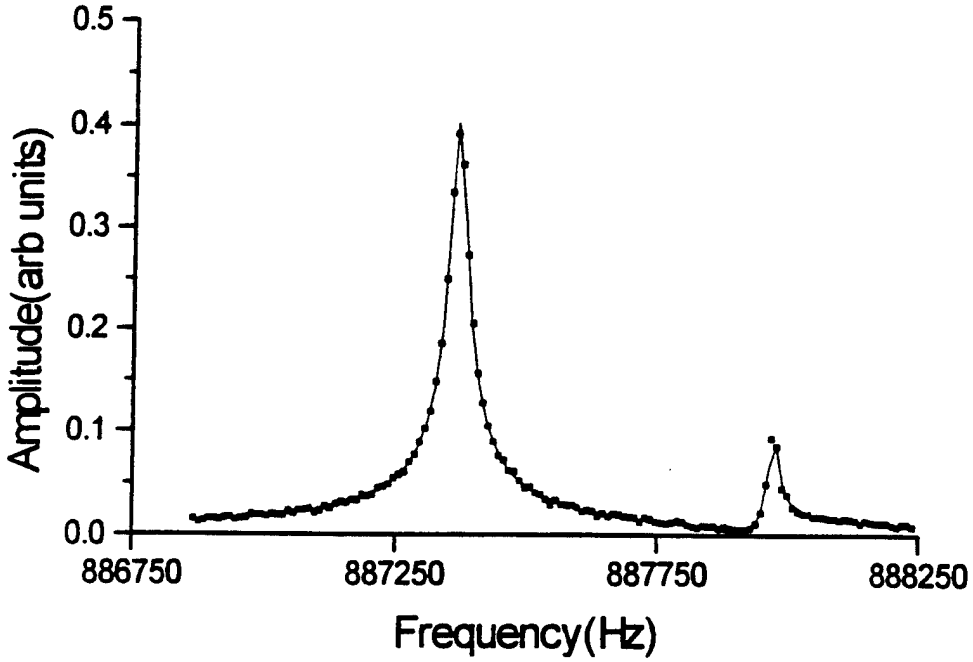
n	fex(MHz)	fm(MHz)	% error	(k,i)
1	.775706	.775707	-.000138	(6,1),(1,1),(4,1),(4,2),(7,1)
6	.819567	.819983	-.050778	(5,1),(3,1),(5,2),(8,1),(2,1)
11	1.075664	1.075399	.024614	(1,2),(7,2),(6,2)
14	1.988616	1.198505	.009239	(5,3),(2,2),(3,2),(8,2),(3,3)
				(8,3),(2,3)
21	1.217375	1.217850	-.039042	(1,3),(6,3),(7,3),(1,4),(6,4)
				(7,4),(4,3)
28	1.440760	1.440750	.000712	(5,4)
29	1.527080	1.526474	.039695	(5,5),(8,4),(3,4),(5,6),(2,4)
34	1.558358	1.558848	-.031448	(5,7),(5,8),(5,9),(3,5),(8,5)
				(3,6),(8,6),(2,6)
43	1.580067	1.579871	.012426	(6,5),(7,5),(7,6),(1,5),(4,4)
				(1,6),(6,6),(4,5),(4,6)

The (k, i) values in table III are used to describe the type of vibrational mode, just as for the rectangular parallelipiped. However, the spherical symmetry of a ball means that some modes are degenerate. That is, several mode types have the same frequency. The

number of identical frequencies, or the degeneracy of a mode is indicated in the first column by skipping numbers until the next distinct frequency occurs.

If the blank from which the ball is ground has been processed in an anisotropic way, then these frequencies separate into distinct resonances, illustrated in Fig.4 on a 7/32" ball ground from a hot-pressed Si_3N_4 blank. If the ball were perfect, the two peaks would merge. However, the

Fig. 4



hot-pressing aligns the grains. This grain alignment produced by such processing variables as cold working in metals or hot pressing of ceramics during sintering is called texturing and produces an anisotropic material not optimum for bearings. From the splitting of the frequencies in fig.4, we can extract detailed quantitative measurements of the texture. Note that the solid line, a fit to the data, determines the splitting to about 0.1Hz. Texturing in other ceramics and alloys is just as easily measured.

In addition to the above examples, there are many other types of information that can be extracted using RUS. These include the Curie temperatures of magnets, an important variable because at the Curie temperature, permanent magnetism disappears. For some rare earth magnets, this can occur as low as 600K, uncomfortably close to underhood temperature extremes in cars. Rus can also determine very precisely in small samples

the matrix fractions in metal-matrix composites (we have used RUS to determine to better than 1% the silica fraction in a sequence of small samples of Al/silica metal/matrix composite) because the matrix changes the elastic properties in a very predictable way. Finally, we note that the directional properties of graphite or fiberglass epoxy composites arise from their layered construction. Clearly a composite formed from graphite cloth and a binder is stiffer and stronger in the plane of the cloth. Such anisotropy is naturally determined by a RUS measurement of the elastic moduli because no additional measurements or computations are needed for anisotropic materials as compared to simple isotropic ones. But it is very difficult or impossible to measure anisotropic moduli using any other dynamic modulus measurement system. These examples, therefore, make it clear that RUS is a newly emerging and powerful tool for the material scientist.

Selected References:

1. "Complete Elastic Constants and Giant Softening of c_{66} in Superconducting $La_{1.86}Sr_{.14}CuO_4$", A.Migliori, William M. Visscher, S.Wong, S.E.Brown, I.Tanaka, H.Kojima, and P.B.Allen, Phys. Rev. Lett. 64, 2458 (1990)

2. "Normal State Transport and Elastic Properties of High T_c Materials and Related Compounds", P. B. Allen, Z. Fisk, and A. Migliori,in Physical Properties of High Temperature Superconductors I, ed. Donald M. Ginsberg, World Scientific Publishing Co., Teaneck, NJ. 1988 (invited book chapter)

3. "Importance of Structural Instability to High-Temperature Superconductivity", A. Bussmann-Holder, A. Migliori, Z. Fisk, J.L.Sarrao, R. G. Leisure, and S-W. Cheong, Phys. Rev. Lett. 67, 512 (1991).

4. "On the Normal Modes of Free Vibration of Inhomogeneous and Anisotropic Elastic Objects", W. M. Visscher, A. Migliori, T. M. Bell, and R. A. Reinert, J. Acoust. Soc. Am. 90 2154 (1991).

5. "Dopant-Induced Symmetry Breaking in the Resonant Ultrasound Spectrum of $La_{1.86}Sr_{.14}CuO_4$", A. Migliori, J. L. Sarrao, M. Lei, T. M. Bell, W. M. Visscher, I. Tanaka, and H. Kojima in Lattice Effects in High-T_c Superconductors

6. "Resonant Ultrasound Spectroscopy of Metal-Hydrogen System and High T_c Superconductors", R. G. Leisure, A. Migliori, R. B. Schwarz, in Perspectives in Physical Acoustics, World Scientific, eds. R. K. Sondfors, P. Sunthrothok, Y. Fu, 1992.(invited book chapter).

7. "Resonant Ultrasound Spectroscopic Techniques for Measurement of the Elastic Moduli of Solids", A. Migliori, J. L. Sarrao, W. M. Visscher, T. M. Bell, Ming Lei, Z. Fisk, and R. G. Leisure, Physica B 183, 1 (1993).

8. "Elastic Constants of a Monocrystal of $Y_1Ba_2Cu_3O_{7-\delta}$ Superconductor", Ming Lei, J. L. Sarrao, W. M. Visscher, T. M. Bell, J. D. Thompson, A. Migliori, U. W. Welp, B. Veal, to appear in Phys. Rev. B, March 1, 1993.

9. "Elastic Properties of U_2Zn_{17}", A. Migliori, J. L. Sarrao, D. G. Mandrus, Z. Fisk, and A. J. Balatsky, Bull Am. Phys. Soc. 38, 79 (1993).

10. "Elastic Properties of La_2CuO_4 to 580 K", J. L. Sarrao, D. G. Mandrus, Ming Lei, A. Migliori, P. C. Canfield, Z. Fisk, and P. D. Kodali, Bull. Am. Phys. Soc. 38, 389 (1993).

11. "Resonant Ultrasound Measurements of Anisotropic Elastic Properties of a Textured

Copper Polycrystal", Ming Lei, A. Migliori, J. L. Sarrao, S. R. Chen, and U. F. Kocks, Bull. Am. Phys. Soc. 38, 396 (1993).

12. "Resonant Ultrasound Measurements of Elastic Constants in Melt-Spun $Pr_2Fe_{14}B$ and $Nd_2Fe_{14}B$", C. D. Fuerst, J. F. Herbst, J. L. Sarrao, and A. Migliori.

13. "Room-Temperature Elastic Constants of Sc and $ScD_{0.18}$", R.G. Leisure, R. B. Schwarz, A. Migliori, and Ming Lei, submitted to Phys. Rev. B.

14. "Hydrogen Isotope Motion in Scandium Studied by Ultrasonic Measurements", R. G. Leisure, R. B. Schwarz, A. Migliori, D. R. Torgeson, and I. Svare, submitted to Phys. Rev. B.

15. "Deuterium Motion in Yttrium Studied by Ultrasonic Measurements", R. G. Leisure, R. B. Schwarz, A. Migliori, D. R. Torgeson, I. Svare, and I. S. Anderson, submitted to Phys. Rev. B.

16. "Analysis of Bond Quality by Resonant Ultrasound", W. M. Visscher, A. Migliori, and R. D. Dixon, to be published.

Wave Propagation in Periodic, Random, and Quasicrystalline Media, and Effects of Anderson Localization

J. D. Maynard
Department of Physics, the Pennsylvania State University, University Park, Pennsylvania 16802
(May 11, 1996)

The study of wave propagation in an array of scatterers dates back to Lord Rayleigh and others. Interesting effects for periodic arrays of scatters were discussed by Brillouin and are treated in texts on solid state physics; however, the effects are quite nontrivial when studied in detail. Remarkable effects for wave propagation in disordered arrays of scatterers have been understood relatively recently, with the discovery of Anderson localization. The discovery of quasicrystals has also led to interest in the consequences of waves propagating within this unusual symmetry. In this paper the effects of wave propagation in periodic, random, and quasiperiodic media will be treated in detail.

I. INTRODUCTION

The behavior of a wave propagating in a nonuniform medium plays a key role in many phenomena in nature, as for example in the electrical and thermal properties of solids, optical phenomena, seismology, oceanography, etc. Understanding the behavior of waves in nonuniform media is also important in practical applications, such as medical imaging, oil exploration, geophysical survey, radar and microwave propagation, and most recently, the possibility of developing new electronic devices based on man-made quantum mechanical waveguides and resonators. The fundamental nature of the problem is extended into a challenging new frontier of science when one considers the possiblities of nonlinear effects.

While wave propagation in nonuniform media has been studied for a long time, only certain limiting conditions are fully understood. Periodic media are understood, although some complexities are often overlooked, as will be illustrated in this paper. When the wavelength of the wave is significantly greater than the scale of the nonuniformity of the medium, then one may use an effective medium theory to describe the wave propagation. If the wave is much smaller than the scale of the nonuniformity, then one may adopt a diffusion approximation by using ensemble averaging or having a system which changes slowly in time. The challenging regime is the one where the wavelength is on the order of the scale of the nonuniformity of the medium, and there is virtually no inelastic scattering. This is the regime of interest in this paper.

Much of the recent literature on wave propagation in nonuniform media has dealt with the behavior of electron Schrodinger waves propagating in disordered metals, and some phenomena such as Anderson localization are most often referred to in the context of quantum electron transport. Of course, the phenomena are valid for wave propagation in any type of system with an analogous nonuniformity. We have been involved in a substantial research program using acoustic wave propagation in nonuniform quantum fluid and macroscopic mechanical systems, with recent emphasis on nonlinear effects. We have found that in order to understand the nonlinear systems, it is necessary to understand the linear systems in great detail; e.g. it is not sufficient to understand the natural frequencies and normal modes of a system, but one must understand the amplitudes of the resonantly driven system as well. In the course of this research we have developed a number of results, all of which are straightforward, but many of which are omitted in the usual treatment of wave propagation in the literature. The purpose of this paper is to present a detailed development of the equations for wave propagation in various nonuniform systems. Included is a tutorial on Anderson localization which is not just a description of the phenomenon, but is also a clear explanation, due to Luban, of why it occurs. The results and analysis of the nonlinear experiments will be presented in a later paper.

A. A Brief Description of the Physics Involved

When a wave in a nearly uniform medium encounters a change in the medium, the wave undergoes a reflection, and only a fraction of the wave is transmitted. If the wave encounters a sequence of scatters, then one might suppose that the transmitted wave becomes successively smaller. However, if the sequence of scatterers is periodic, then the wave can travel through the system (at certain frequencies) as though there were no scattering at all. A similar phenomena occurs in solid state physics, as for a quantum mechanical electron in a metal. In the metal the electron is very strongly scattered by the positive ions, and such scattering of the electron would result in low electrical conductivity. But metals have high electrical conductivity; because the ions in a metallic crystal are arranged periodically, the electron Schrodinger waves can propagate through the system as though there were no scatterers at all. This behavior, valid for any type of wave in a periodic potential field, is a consequence of Floquet's Theorem [1], also known as Bloch's theorem [2] in solid state physics.

According to Bloch's theorem, the eigenstates of a system described by a periodic potential are extended, having the same nominal amplitude at all positions in space. When a random element is added to the potential field, the eigenstate amplitude may not simply acquire a random variation but instead may become localized, decaying exponentially from a particular site with a characteristic length. [3,4] This phenomenon, referred to as Anderson localization, has been studied extensively in regard to the behavior of electrons in narrow wires and semiconductor channels. It is possible to observe localization effects in other wave systems, such as ones involving photons, phonons, or quantum fluid waves, where observations may be more straightforward. We have made such measurements [21,37] in an acoustic localization experiment where we can observe in detail eigenstate amplitude and phase, eigenvalue spectra and density of states, and dependence of localization length on strength of disorder, showing fluctuations as a function of position within a band, separation from nearby states, etc.

The analogy between electron wave and acoustic wave localization is mathematically precise. The wave functions ψ may be superpositions of energy states with time dependence $\exp(-iEt/\hbar)$ for electrons, or monotonal waves with time dependence $\exp(-i\omega t)$ for sound waves. For these states the wave equation may be written

$$\nabla^2\psi + \left[q^2 - V(r)\right]\psi = 0 \tag{1}$$

where for particles $q^2 = 2mE/\hbar^2$ (with m the particle mass), and for sound waves $q = \omega/c$ (with c the characteristic speed of sound). $V(r)$ is the potential (normalized with $2m/\hbar^2$) for particles and a combination of a stiffness operator and mass density for an acoustic medium. For a one- dimensional acoustic system V is easily determined from basic mechanics or acoustics.

Studying the acoustic analog systems has some definite advantages over studying the quantum systems directly. In the acoustic systems all parameters and conditions may be controlled, or at least measured, with high accuracy and precision. The acoustic systems may be carefully constructed so as to accurately coincide with the mathematical governing equations. The acoustic systems may be driven at very low amplitudes where they are quite linear, or at high amplitudes where the nonlinearity is quite evident. In an acoustic experiment, eigenvalues and eigenfunctions, including amplitude and phase, are readily measured. For nonlinear systems, extended waves as well as solitary waves may be studied.

In the following sections of this paper, various wave mechanical systems will be discussed, and derivations of results necessary for understanding the linear and nonlinear experiments will be presented.

II. THEORY FOR POINT MASSES COUPLED WITH MASSLESS SPRINGS

We consider a one-dimensional sequence of masses m_j $(j = 0, N+1)$ connected with ideal springs of stiffness s_j $(j = 0, N)$. The spring s_j lies between mass m_j and mass m_{j+1}. The displacement of the mass m_j is denoted y_j; y_0 and y_{N+1} serve as boundary conditions at each end of the system. Inside the boundaries (j=1,N) the motion of each mass is governed by Newton's law and Hooke's law:

$$m_j \frac{d^2 y_j}{dt^2} = -s_{j-1}(y_j - y_{j-1}) + s_j(y_{j+1} - y_j) \tag{2}$$

We assume that the motion can be Fourier analyzed in time, and adopt a time dependence $\exp(-i\omega t)$. Eq. (2) becomes

$$-m_j\omega^2 y_j = -s_{j-1}(y_j - y_{j-1}) + s_j(y_{j+1} - y_j) \tag{3}$$

At this point standard derivations make the sequence periodic and invoke Bloch's theorem, which provides the form for the normal modes of the coupled system in Eq. (3). However, Bloch's theorem is usually derived for an infinitely periodic system or a system which has periodic boundary conditions. In the derivation presented here, a technique will used which is valid for arbitrary boundary conditions, and which will furthermore provide useful insight into the properties of a disordered (nonperiodic) system.

Eq. (3) may be rewritten as

$$y_{j+1} = \left(1 + \frac{s_{j-1}}{s_j} - 4\frac{\omega^2}{\omega_j^2}\right) y_j - \frac{s_{j-1}}{s_j} y_{j-1} \tag{4}$$

where $\omega_j = 2\sqrt{s_j/m_j}$. The system of equations represented by Eq. (4) may be written as two-by-two matrix equations:

$$\begin{pmatrix} y_{j+1} \\ y_j \end{pmatrix} = \begin{pmatrix} 2\zeta_j & -\delta_j \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_j \\ y_{j-1} \end{pmatrix} \tag{5}$$

where $\zeta_j = (1+\delta_j)/2 - 2(\omega/\omega_j)^2$ and $\delta_j = s_{j-1}/s_j$. By repeated use of Eq. (5), one has

$$\begin{pmatrix} y_{N+1} \\ y_N \end{pmatrix} = \left[\prod_{j=1}^{N} \begin{pmatrix} 2\zeta_j & -\delta_j \\ 1 & 0 \end{pmatrix}\right] \begin{pmatrix} y_1 \\ y_0 \end{pmatrix} \tag{6}$$

The complex mechanical impedances at the ends are given by

$$Z_0 = im_0\omega + \frac{s_0}{i\omega}\frac{y_0 - y_1}{y_0} \tag{7}$$

and

$$Z_{N+1} = im_{N+1}\omega + \frac{s_N}{i\omega}\frac{y_{N+1} - y_N}{y_{N+1}} \tag{8}$$

If the boundary conditions (mechanical impedances) at the ends are known then Eqs. (6), (7), and (8) can be used to find the natural frequencies, and repeated use of Eq. (5) may then be used to find the normal modes y_j :

$$\begin{pmatrix} y_{j+1} \\ y_j \end{pmatrix} = \left[\prod_{j'=1}^{j} \begin{pmatrix} 2\zeta_{j'} & -\delta_{j'} \\ 1 & 0 \end{pmatrix} \right] \begin{pmatrix} y_1 \\ y_0 \end{pmatrix} \tag{9}$$

To proceed, we look for the diagonalization of the matrix in Eq. (5); that is, we need to find matrices D_j and inverses D_j^{-1} such that

$$\begin{pmatrix} 2\zeta_j & -\delta_j \\ 1 & 0 \end{pmatrix} = D_j^{-1} \begin{pmatrix} \lambda_j^+ & 0 \\ 0 & \lambda_j^- \end{pmatrix} D_j \tag{10}$$

The eigenvalues λ_j are found by solving the secular equation given by setting a determinant equal to zero:

$$\det \begin{pmatrix} 2\zeta_j - \lambda_j & -\delta_j \\ 1 & -\lambda_j \end{pmatrix} = (2\zeta_j - \lambda_j)(-\lambda_j) + \delta_j = 0 \tag{11}$$

The roots are

$$\lambda_j^{\pm} = \begin{pmatrix} \zeta_j \pm i\left(\delta_j - \zeta_j^2\right)^{1/2} & if\ \zeta_j^2 < \delta_j \\ \zeta_j \pm \left(\zeta_j^2 - \delta_j\right)^{1/2} & if\ \zeta_j^2 > \delta_j \end{pmatrix} \tag{12}$$

Note that $\lambda_j^+ \lambda_j^- = \delta_j$ and $\left(\lambda_j^+ + \lambda_j^-\right)/2 = \zeta_j$. The product in Eq. (9) can now be written as

$$\begin{pmatrix} y_{j+1} \\ y_j \end{pmatrix} = D_j^{-1} \begin{pmatrix} \lambda_j^+ & 0 \\ 0 & \lambda_j^- \end{pmatrix} D_j D_{j-1}^{-1} \begin{pmatrix} \lambda_{j-1}^+ & 0 \\ 0 & \lambda_{j-1}^- \end{pmatrix}$$

$$D_{j-1} \cdots\cdots\cdots D_1^{-1} \begin{pmatrix} \lambda_1^+ & 0 \\ 0 & \lambda_1^- \end{pmatrix} D_1 \begin{pmatrix} y_1 \\ y_0 \end{pmatrix} \tag{13}$$

We now specialize to the case of a periodic system. We let $m_j = m$, $s_j = s$, $\omega_j = \omega_0 = 2\sqrt{s/m}$, $\delta_j = 1$, $D_j = D$, and

$$\zeta_j = \zeta = 1 - 2\left(\omega/\omega_0\right)^2 \tag{14}$$

Because of the cancellation of DD^{-1} in the product in Eq. (13), we have a simple expression for the displacements y_j :

$$\begin{pmatrix} y_{j+1} \\ y_j \end{pmatrix} = D^{-1} \begin{pmatrix} (\lambda^+)^j & 0 \\ 0 & (\lambda^-)^j \end{pmatrix} D \begin{pmatrix} y_1 \\ y_0 \end{pmatrix} \tag{15}$$

The expression for $\lambda^{\pm}$ for the periodic system is

$$\lambda^{\pm} = \begin{pmatrix} \zeta \pm i\left(1 - \zeta^2\right)^{1/2} & if\ 0 < \omega < \omega_0 \\ \zeta \pm \left(\zeta^2 - 1\right)^{1/2} & if\ \omega > \omega_0 \end{pmatrix} \tag{16}$$

Note that if $\omega > \omega_o$, then $|\lambda^+| < 1$ and $|\lambda^-| > 1$. Because of the powers of j in Eq. (15), the larger eigenvalue grows exponentially while the smaller eigenvalue decays; these solutions correspond to evanescent waves, which are usually not regarded as solutions of interest.

If $0 < \omega < \omega_0$, then $\lambda^+ = (\lambda^-)^*$, $|\lambda^{\pm}| = 1$, and we can write

$$\lambda^{\pm} = e^{\pm ika} \tag{17}$$

The choice of notation for the phase, written as ka, will be evident later. The relation $\left(\lambda^+ + \lambda^-\right)/2 = \zeta$ becomes

$$\left(e^{+ika} + e^{-ika}\right)/2 = \zeta = 1 - 2\left(\omega/\omega_0\right)^2 \tag{18}$$

or

$$\omega^2 = \omega_0^2\left[1 - \cos(ka)\right]/2 = \omega_0^2 \sin^2(ka/2) \tag{19}$$

which finally yields the dispersion relation

$$\omega_k = \omega_0 \sin(ka/2) \tag{20}$$

The subscript k has been added to emphasize the dispersion. Note that $0 < k < \pi/a$. This dispersion relation is illustrated in Fig. 1.

We now consider some limits relevant to a continuous system. We assume that the springs connecting the masses have been stretched to a periodic spacing a with the application of a static tension T_0, so that $T_0 = sa$. We now take the limit $m \to 0$ and $a \to 0$, while keeping the mass per unit length, $\mu = m/a$, constant. As $a \to 0$, then $\sin(ka/2) \to ka/2$, and the dispersion relation becomes

$$\begin{aligned} \omega_k &= \omega_o \sin(ka/2) \\ &\to (a\omega_o/2)\, k \\ &= \left[a\sqrt{(s/m)}\right] k \\ &= \left[(sa)/(m/a)\right]^{1/2} k \\ &= \left[T_0/\mu\right]^{1/2} k \\ &= ck \end{aligned} \tag{21}$$

where $c = \sqrt{T_0/\mu}$ is the usual speed of sound in a string of mass per unit length μ and tension T_0. Thus the dispersion relation $\omega = ck$ for the continuum limit is obtained.

If one considers the dispersion relation for the discrete system (Eq. (20)) in the limit of small k, one also obtains $\omega = ck$. However, if this is extrapolated to the maximum $k = \pi/a$, one finds a frequency $\omega = c\pi/a$, whereas the actual frequency should be ω_o, which is a factor of $2/\pi$ smaller then the extrapolated value of $c\pi/a$. The dispersion given by Eq. (20) is a fundamental difference between the discrete coupled oscillators appropriate for phonons and sound waves in a continuous system.

To continue the problem of the periodic coupled oscillators, we must now find the matrices D and D^{-1}. We need to find the eigenvectors $\left(\psi_1^+, \psi_0^+\right)$ and $\left(\psi_1^-, \psi_0^-\right)$ such that

$$\begin{pmatrix} 2\zeta - \lambda^{\pm} & -1 \\ 1 & -\lambda^{\pm} \end{pmatrix} \begin{pmatrix} \psi_1^{\pm} \\ \psi_0^{\pm} \end{pmatrix} = 0 \tag{22}$$

One finds $\psi_0^+/\psi_1^+ = (2\zeta - \lambda^+) = \lambda^-$ and $\psi_0^-/\psi_1^- = 1/\lambda^-$. These eigenvectors, suitably normalized, are used as the columns of the matrix D^{-1}; one finds:

$$D^{-1} = \begin{pmatrix} 1/\chi & \lambda^-/\chi \\ \lambda^-/\chi & 1/\chi \end{pmatrix} \tag{23}$$

and

$$D = \begin{pmatrix} 1/\chi & -\lambda^-/\chi \\ -\lambda^-/\chi & 1/\chi \end{pmatrix} \tag{24}$$

where $\chi^2 = \left[1 - (\lambda^-)^2\right] = [(2i)\exp(-ika)\sin(ka)]$. From Eq. (15) we now have for the displacement y_j at any site j :

$$\begin{pmatrix} y_{j+1} \\ y_j \end{pmatrix} = D^{-1} \begin{pmatrix} \left(e^{ika}\right)^j & 0 \\ 0 & \left(e^{-ika}\right)^j \end{pmatrix} D \begin{pmatrix} y_1 \\ y_0 \end{pmatrix} \tag{25}$$

Using the expressions for D, D^{-1}, and χ, we obtain

$$y_j = e^{ik(ja)}U_k + e^{-ik(ja)}U_{-k} \tag{26}$$

where

$$U_k = \frac{y_1 - y_0 e^{-ika}}{(2i)\sin(ka)} \tag{27}$$

The values of y_0, y_1, U_k, and restrictions on k (eigenvalues) are determined by the boundary conditions and initial conditions. For example, for a system clamped at the ends, $y_0 = 0$, $y_{N+1} = 0$,

$$y_j = \sin[k(ja)]\left(\frac{y_1}{\sin(ka)}\right) \tag{28}$$

and k is quantized with $k = n\pi/L$, $n = 1, 2, \ldots N$, with $L = (N+1)a$. For an infinite system or a system with periodic boundary conditions the normal modes correspond to $U_k = 0$ or $U_{-k} = 0$. If one uses a coordinate system with the origin at the mass m_0, then the position of the j^{th} mass is $x = ja$, and we can write for infinite or periodic boundary conditions

$$y_k(x) = e^{ikx}U_k \tag{29}$$

which are Bloch waves for the discrete system. The subscript k has been added to the displacement field $y(x)$ as a label for the normal mode. The final solution will be a linear combination of the normal modes:

$$y_j = \sum_k \left(e^{ik(ja)}U_k + e^{-ik(ja)}U_{-k}\right)e^{-i\omega_k t} \tag{30}$$

If we assume periodic boundary conditions, so that $\exp[ika(N+1)] = 1$, we have

$$\sum_{j=0}^{N}\left(e^{ikaj}\right)\left(e^{ik'aj}\right)^* = \frac{1 - e^{ika(N+1)}e^{-ik'a(N+1)}}{1 - e^{ika}e^{-ik'a}}$$

$$= (N+1)\,\delta_{k,k'} \tag{31}$$

which shows the orthogonality of the normal modes. For clamped end conditions we have

$$\sum_{j=0}^{N}\sin(kaj)\sin(k'aj) = \frac{1}{2}(N+1)\,\delta_{k,k'} \tag{32}$$

Using the orthogonality conditions we can derive the expression for the total energy of the system (with periodic boundary conditions):

$$E = \frac{1}{2}M\sum_k\left(|U_k|^2 + |U_{-k}|^2\right)\omega_k^2 \tag{33}$$

where $M = (N+1)m$. If only one amplitude $|U_k|$ is appreciable, then the amplitude is given by $\sqrt{2E/M\omega_k^2}$, corresponding to the classical turning points of a simple harmonic oscillator.

The normalization of the eigenfunctions and the energy expression are important in the analysis of nonlinear effects and in the comparison of classical and quantum mechanical results.

III. HARMONICALLY COUPLED QUANTUM PARTICLES

For a system of harmonically coupled quantum mechanical particles, we use a Lagrangian formulation instead of the Newton's law formulation in Eq. (2). [7–9] The Lagrangian is

$$L = \frac{1}{2}\sum_j m_j\left(\frac{dy_j}{dt}\right)^2 - \frac{1}{2}\sum s_j\left(y_{j+1} - y_j\right)^2 \tag{34}$$

The canonical momenta are $p_j = m_j\left(dy_j/dt\right)$, and the Hamiltonian is

$$H = \sum_j p_j\left(\frac{dy_j}{dt}\right) - L \tag{35}$$

The system is quantized with the commutation relations

$$[y_j, p_{j'}] = i\hbar\delta_{j,j'} \tag{36}$$

We now assume a periodic system with $m_j = m$ and $s_j = s$. Guided by our results with the classical system, we let

$$y_j = \sum_k \left(e^{ikja} Y_k + e^{-ikja} Y_{-k}\right) \tag{37}$$

The Hamiltonian becomes

$$H = \frac{1}{2m} \sum_k P_k P_{-k} + \frac{1}{2} m \sum_k \omega_k^2 Y_k Y_{-k} \tag{38}$$

where $P_k = m\,(dY_{-k}/dt)$ and ω_k is defined as before in Eq. (20). One could use this Hamiltonian and construct a Schrodinger wave equation for a field $\psi(Y)$, where Y is a point in the 2N-dimensional Y_k space, and $P_k \psi(Y) = -i\hbar\,(\partial\psi/\partial Y_k)$. However, a better approach is to use another transformation and construct the properties of the eigenfunctions and eigenvalues using the commutation relations in Eq. (36). We define

$$a_k^+ = \left(\frac{m}{2\hbar\omega_k}\right)^{1/2} \left(\omega_k Y_{-k} - \frac{i}{m} P_k\right) \tag{39}$$

and

$$a_k = \left(\frac{m}{2\hbar\omega_k}\right)^{1/2} \left(\omega_k Y_k + \frac{i}{m} P_{-k}\right) \tag{40}$$

The Hamiltonian becomes

$$H = \sum_k \hbar\omega_k \left(N_k + \frac{1}{2}\right) \tag{41}$$

where $N_k = a_k^+ a_k$. Using Eq. (36), one finds the commutation relation for a_k^+ and a_k

$$\left[a_k^+, a_{k'}\right] = \delta_{k,k'} \tag{42}$$

One also finds the relations

$$N_k a_k = a_k\,(N_k - 1) \tag{43}$$

and

$$N_k a_k^+ = a_k^+\,(N_k + 1) \tag{44}$$

We now write ψ_{kn} for an eigenfunction of the operator N_k with eigenvalue n_k. One can show that the eigenvalues must be positive. Using the relations in Eqs. (43) and (44), one finds that $a_k \psi_{kn}$ is an eigenfunction with eigenvalue $(n_k - 1)$, and $a_k^+ \psi_{kn}$ is an eigenfunction with eigenvalue $(n_k + 1)$. We then consider the sequence of eigenfunctions given by $(a_k)^p \psi_{kn}$, with p a positive integer. The eigenvalues form the sequence $(n_k - 1), (n_k - 2), \ldots (n_k - p) \ldots$ If this sequence were allowed to continue indefinitely, then for $p > n_k$, the eigenvalues would become negative, which cannot happen. If n_k is such that $(n_k - p) = 0$, then the sequence will terminate without going negative. Thus we have $n_k = p$; i.e. n_k is quantized to be a positive integer.

A general state of the system will be superposition of eigenstates:

$$\psi = \sum_k \sum_n C_{kn} \psi_{kn} \tag{45}$$

where $\mid C_{kn} \mid^2$ gives the probability that the system is in the state ψ_{kn}. The expectation value of the total energy is (using $E\psi = H\psi$):

$$< \psi \mid E \mid \psi >= \sum_k \sum_n \mid C_{kn} \mid^2 \hbar\omega_k \left(n_k + \frac{1}{2}\right) \tag{46}$$

If the expectation value of the position operator Y_k is examined, it is found that it couples a state ψ_{nk} only to the state $\psi_{(n+1)k}$ or $\psi_{(n-1)k}$. As a consequence the time dependence will behave like

$$e^{i[(n+1)-n]\omega_k t} + e^{i[(n-1)-n]\omega_k t} = 2\cos(\omega_k t) \tag{47}$$

If $C_{nk} \sim C_{(n\pm1)k}$, and is appreciable for only one k and for $n_k >> 1/2$, then [7]

$$< \psi \mid Y_k \mid \psi >= \sqrt{\frac{2E'}{m\omega_k^2}} \cos(\omega_k t) \tag{48}$$

where E' is the square of the expectation value of the square root of the energy operator. Eq. (48) corresponds to the normal mode oscillating between the classical turning points at the frequency ω_k. It is interesting to note that in order to observe a "classical sound wave", the quantum system must not only have a large n_k so that $n_k\hbar\omega_k$ corresponds to the macroscopic classical energy, the quantum state must contain a superposition of two eigenstates, ψ_{nk} and $\psi_{(n\pm1)k}$.

The term phonon can now be defined in the following sense: we can say that a phonon is emitted or absorbed when a system of harmonically coupled quantum particles makes a transition from a state ψ_{nk} to a state $\psi_{(n-1)k}$ or $\psi_{(n+1)k}$.

A. Larger unit cells and higher dimensions

Instead of having a periodic system such that $m_j = m$, etc., one may have different masses or spring constants, but still have a periodic system. That is, several different masses and springs may be joined to form a unit cell, and this unit cell may be repeated periodically. For M different masses or springs in three dimensions, the matrix in Eq. (5) will enlarge from 2x2 to 6Mx6M, and instead of a single dispersion relation as in Eq. (20), there will be multiple dispersion relations, and multiple "branches" in the eigenvalue spectrum. If the system is extended into two or three dimensions, the matrix representing the dynamics of the system is further enlarged, the displacement fields become vector fields, and the eigenfunctions may be characterized as transverse or longitudinal. A thorough discussion of the dynamics of a three dimensional system of coupled oscillators is available in Ref. [8].

B. Phonon Density of States

In statistical physics, one calculates the properties of a system, such as specific heat, by taking a product of the quantity of interest with a distribution function, and integrating over all possible states. The distribution functions are functions of the energy of the state, and for this reason it is usually easier to transform the integration variable from the state label, such as the Bloch wavevector k in our one-dimensional example, to the energy, which is related to ω in our example. In our one-dimensional example, the transformation is accomplished using

$$dk = \left(\frac{d\omega}{dk}\right)^{-1} d\omega = d\omega/v_g \tag{49}$$

where $v_g = d\omega/dk$ is the group velocity, which will be discussed in detail later. For the general case in higher dimensions, the Jacobian of the transformation is the density of states, which also varies as $1/v_g$. Values of the state parameter (e.g. k) where the dispersion relation ω_k is flat are called critical points, and at these points the group velocity is zero and the density of states is singular. Such singularities are called van Hove singularities. In our one dimensional example

$$\frac{d\omega}{dk} = \frac{2\omega_0}{a}\cos(ka/2) = \frac{2}{a}\left(\omega_0^2 - \omega^2\right)^{1/2} \tag{50}$$

so that $d\omega/dk = 0$, and the density of states is singular, when $\omega = \omega_0$. In two and three dimensions the singularities may be classified according to the nature of $\mid \vec{\nabla}_k \omega \mid = 0$, whether it is a maximum, minimum, or one of two types of saddlepoint. The van Hove singularities are important in determining the properties of a material relating to phonons, such as the specific heat or thermal conductivity.

IV. POINT MASSES COUPLED WITH MASSIVE SPRINGS: THE KRONIG-PENNY SYSTEM

We now consider a system of point masses coupled with springs which have mass. If the springs have mass, then the springs themselves can propagate waves. If an ideal spring is stretched with a tension T_0 and has mass per unit length μ, then the speed of a wave in the spring is $c = \sqrt{T_0/\mu}$. An equivalent picture of such a system is a string, stretched with tension T_0 and having mass density μ, which is loaded at intervals with point masses. In this model the waves may be transverse waves, with speed $c = \sqrt{T_0/\mu}$. Mathematically this system corresponds to a wave equation with a Kronig-Penny potential field.

The coordinate and indexing system is as follows: N masses, m_j, are located at positions x_j, with $1 \leq j \leq N$. The system is terminated with a complex mechanical impedance Z at the position x_{N+1}. The driving point impedance Z_0 and the reflection coefficient for the system are referenced to the point $x_0 = 0$. The spacing between successive masses is defined with $a_j = x_j - x_{j-1}$. Between the masses m_j and m_{j+1} the eigenfunction must be of the form

$$\psi(x,t) = Re\left[\left(u_j e^{iq(x-x_j)} + v_j e^{-iq(x-x_j)}\right) e^{-i\omega t}\right] \tag{51}$$

where Re indicates taking the real part, and $q = \omega/c$. The eigenfunction is defined piecewise, with j selected so that $x_j < x < x_{j+1}$.

The continuity of the string at the positions x_j requires

$$u_j + v_j = u_{j-1}e^{iqa_j} + v_{j-1}e^{-iqa_j} \tag{52}$$

Newton's Law at a mass m_j is written as

$$-T_0\frac{\partial\psi_-}{\partial x} + T_0\frac{\partial\psi_+}{\partial x} = m_j\frac{\partial^2\psi}{\partial t^2} \tag{53}$$

where ψ_- and ψ_+ are the wave fields just to the left and right of the mass m_j, respectively. This system can also be described with a time-independent Schrodinger equation

$$\nabla^2\Psi + \left[q^2 - V(x)\right]\psi = 0 \tag{54}$$

where $V(x)$ is a Kronig-Penny potential field given by

$$V(x) = \sum_{j=1}^{N} V_j \delta(x - x_j) \tag{55}$$

with $V_j = -(m_j/\mu)\,q^2$. Inserting Eq. (51) into Eq. (53) yields

$$-T_0 iq\left(u_{j-1}e^{iqa_j} - v_{j-1}e^{-iqa_j}\right) + T_0 iq\left(u_j - v_j\right)$$

$$= -m_j\omega^2\left(u_j + v_j\right) \tag{56}$$

Eqs. (52) and (56) can be put in a matrix form

$$\begin{pmatrix} u_{j-1} \\ v_{j-1} \end{pmatrix} = \begin{pmatrix} \alpha_j & \beta_j \\ \beta_j^* & \alpha_j^* \end{pmatrix} \begin{pmatrix} u_j \\ v_j \end{pmatrix} \tag{57}$$

where

$$\alpha_j = (1 - i\eta_j)\,e^{-iqa_j} \tag{58}$$

$$\beta_j = -i\eta_j e^{-iqa_j} \tag{59}$$

with $\eta_j = m_j q/2\mu$. The determinant of the matrix in Eq. (57) is unity:

$$\mid \alpha_j \mid^2 - \mid \beta_j \mid^2 = 1 \tag{60}$$

Eq. (60) is necessary for conservation of power flow through the system; that is, it is necessary in order to have

$$| u_j |^2 - | v_j |^2 = | u_{j-1} |^2 - | v_{j-1} |^2 \tag{61}$$

Eqs. (57) and (60) are valid for any system of energy conserving scatterers, not just the point masses used in our model system. For general scatterers, the α's and β's can be given in terms of complex reflection and transmission coefficients: $\alpha_j = 1/T_j$ and $\beta_j = (R_j/T_j)^*$. The results for layered media are presented in the Appendix. If the scatterers are points, such that the wave field is continuous across the scatterers, then $\eta_j = i\,(R_j/T_j)^* \exp(iqa_j)$ and we have the extra equation:

$$\alpha_j e^{iqa_j} + \beta_j^* e^{-iqa_j} = 1 \tag{62}$$

The inverse of the matrix in Eq. (57) is readily found

$$\begin{pmatrix} u_{j+1} \\ v_{j+1} \end{pmatrix} = \begin{pmatrix} \alpha_j^* & -\beta_j \\ -\beta_j^* & \alpha_j \end{pmatrix} \begin{pmatrix} u_j \\ v_j \end{pmatrix} \tag{63}$$

An actual boundary value problem involves specifying the boundary condition at x_{N+1}. One sets $u_N = 1$, and

$$v_N = e^{i2qa_{N+1}} \frac{\mu c - Z}{\mu c + Z} \tag{64}$$

where Z is the complex mechanical impedance at the point x_{N+1}. If the string continues to ∞ without any further masses, then $Z = \mu c$, and $v_N = 0$. If the string is clamped at x_{N+1}, then $Z = \infty$, and $v_N = -e^{i2qa_{N+1}}$.

Once v_N is determined by the end condition, the wave field at $x_0 = 0$ may be determined from

$$\begin{pmatrix} u_0 \\ v_0 \end{pmatrix} = \prod_{j=1}^{N} \begin{pmatrix} \alpha_j & \beta_j \\ \beta_j^* & \alpha_j^* \end{pmatrix} \begin{pmatrix} u_N \\ v_N \end{pmatrix} \tag{65}$$

The complex reflection coefficient for the whole system, referenced to $x_0 = 0$, is $R = v_0/u_0$. The complex mechanical impedance at $x_0 = 0$ is

$$Z_0 = \mu c \frac{1-R}{1+R} = \mu c \frac{u_0 - v_0}{u_0 + v_0} \tag{66}$$

The power P delivered to the system by a driving force $F_0 \cos(\omega t)$ at $x_0 = 0$ is given by

$$\mu c P / F_0^2 = \frac{\mu c Re Z_0}{| Z_0 |^2} = \frac{| u_0 |^2 - | v_0 |^2}{| u_0 - v_0 |^2} \tag{67}$$

Note that for strong scatterers, $v_0 \simeq -u_0$, and $\mu c P/F_0^2$ is approximately equal to one fourth the modulus of the complex transmission coefficient, squared: $| T |^2 = 1 - | R |^2$.

The power P delivered to the system by a velocity (displacement) drive $V_0 \cos(\omega t)$ at $x_0 = 0$ is given by

$$P/\mu c V_0^2 = \frac{1}{\mu c} Re Z_0 = \frac{| u_0 |^2 - | v_0 |^2}{| u_0 + v_0 |^2} \tag{68}$$

If the real system has damping, then in all expressions one must replace qa with a complex quantity $\tilde{q}a = qa[1 + i/(2Q)] = qa + i\delta$, where Q is the quality factor of the system. In particular α_j and β_j are replaced with $\alpha_j e^{\delta}$ and $\beta_j e^{\delta}$, and α_j^* and β_j^* are replaced with $\alpha_j^* e^{-\delta}$ and $\beta_j^* e^{-\delta}$. Eq. (61) becomes

$$| u_j |^2 - | v_j |^2 = e^{-2\delta} | u_{j-1} |^2 - e^{+2\delta} | v_{j-1} |^2 \tag{69}$$

showing the decay of right and left power fluxes. The real part of $Z_0/\mu c$ becomes

$$\frac{1}{\mu c} Re Z_0 = \frac{| u_0 |^2 - | v_0 |^2 - Im(u_0 v_0^*)/Q}{| u_0 + v_0 |^2} \tag{70}$$

For high Q systems, the last term in the numerator of Eq. (70) is negligible.

Some insight into the general problem can be obtained by diagonalizing the matrix in Eq. (57). One must find a matrix D_j such that

$$D_j \begin{pmatrix} \alpha_j & \beta_j \\ \beta_j^* & \alpha_j^* \end{pmatrix} D_j^{-1} = \begin{pmatrix} \lambda_j^+ & 0 \\ 0 & \lambda_j^- \end{pmatrix} \tag{71}$$

Equivalently, one has

$$\begin{pmatrix} \alpha_j & \beta_j \\ \beta_j^* & \alpha_j^* \end{pmatrix} = D_j^{-1} \begin{pmatrix} \lambda_j^+ & 0 \\ 0 & \lambda_j^- \end{pmatrix} D_j \tag{72}$$

The λ_j^+ and λ_j^- are found from

$$\det \begin{pmatrix} (\alpha_j - \lambda_j) & \beta_j \\ \beta_j^* & (\alpha_j^* - \lambda_j) \end{pmatrix} = 0 \tag{73}$$

or, using Eq. (60),

$$\lambda_j^2 + 2\,(Re\alpha_j)\,\lambda_j + 1 = 0 \tag{74}$$

The roots are

$$\lambda_j^{\pm} = (Re\alpha_j) \pm \left[(Re\alpha_j)^2 - 1\right]^{1/2} \tag{75}$$

It should be noted that $(\lambda_j^+ + \lambda_j^-)/2 = Re\alpha_j$, and $\lambda_j^+ \lambda_j^- = 1$. If $| \lambda_j^+ | > 1$, then $| \lambda_j^- | < 1$. The product in Eq. (65) can now be written as

$$\begin{pmatrix} u_o \\ v_o \end{pmatrix} = D_1^{-1} \begin{pmatrix} \lambda_1^+ & 0 \\ 0 & \lambda_1^- \end{pmatrix} D_1 D_2^{-1} \begin{pmatrix} \lambda_2^+ & 0 \\ 0 & \lambda_2^- \end{pmatrix}$$

$$\cdots\cdots\cdots D_N^{-1} \begin{pmatrix} u_N \\ v_N \end{pmatrix} \tag{76}$$

If $| \lambda_j^+ | > 1$, and if the matrices D_j and D_{j+1}^{-1} tend to "cancel", then in the product in Eq. (76) the larger diagonal element grows exponentially, and the smaller one tends to decrease exponentially. This is a hint as to how evanescent and exponentially localized eigenfunctions occur.

It is instructive to now consider the case where the masses on the string are periodic, so that $m_j = m, \eta_j = \eta, a_j = a, \alpha_j = \alpha, \beta_j = \beta, \lambda_j^{\pm} = \lambda^{\pm}$, and $D_j = D$. In order to avoid evanescent solutions, we must have $Re\alpha < 1$ and $\mid \lambda^{\pm} \mid = 1$. We write $\lambda^{+} = \exp(-ika)$ and $\lambda^{-} = \exp(+ika)$. The parameter k is found from the relation $(\lambda^{+} + \lambda^{-})/2 = Re\alpha$:

$$\frac{1}{2}\left(e^{-ika} + e^{+ika}\right) = Re\alpha \tag{77}$$

For our model with point scatterers, Eq. (77) yields the dispersion relation:

$$\cos(ka) = \cos(qa) - \eta \sin(qa) \tag{78}$$

which can be rewritten as

$$\cos(ka) = \sqrt{1+\eta^2}\cos\left(qa + \tan^{-1}\eta\right) \tag{79}$$

It should be recalled that η depends on q. However if $\eta > 3$, then q is approximately equal to an integer multiple of π/a, so its effect in Eq. (79) is weak. The behavior of Eq. (79) can be studied graphically, as in Fig. 2. In Fig. 2, the right hand side of Eq. (79) is plotted simply as a cosine with amplitude $\sqrt{1+\eta^2}$ and shifted so that $qa = 0$ corresponds to $ka = 0$. Solutions for $\cos(ka)$ must lie within ± 1, so that bands and gaps for qa are formed. Taking the inverse cosine of the ordinate in Fig. 2 to obtain k, and plotting $\omega (= cq)$ versus k, we obtain Fig. 3, which is the conventional picture of band structure as in solid state physics. Some points of interest and the eigenfunctions to which they correspond are indicated by the letters a through e in Figs. 2 and 4. Letters a and b correspond to low frequency modes, with the string between the masses bending very little. For higher frequency modes the string has an increasingly larger bend; the actual bending wavelength in the string is $2\pi/q = 2\pi c/\omega$, not $2\pi/k$. For the modes c through e, the behavior of eigenfunction for adjacent bays should be noted. For $\eta >> 1$, the eigenfunctions are essentially standing waves between the scatterers.

The widths of the bands can be determined from Eq. (79). The band edges are given by $\cos(ka) = \pm 1$. One of the band edges is given by the unperturbed point $qa = n\pi$, where n is the band number, with $n = 0$ being the lowest band. For the other band edge we write $qa = n\pi + \Delta$. Eq. (79) becomes

$$(-1)^{n+1} = \sqrt{1+\eta^2}\cos\left(n\pi + \Delta + \tan^{-1}\eta\right)$$

$$= (-1)^{n}\sqrt{1+\eta^2}\cos\left(\Delta + \tan^{-1}\eta\right) \tag{80}$$

Solving for Δ yields

$$\Delta = \cos^{-1}\left[\frac{-1}{\sqrt{1+\eta^2}}\right] - \tan^{-1}\eta$$

$$= \left[\frac{\pi}{2} + \tan^{-1}\frac{1}{\eta}\right] - \left[\frac{\pi}{2} - \tan^{-1}\frac{1}{\eta}\right]$$

$$= 2\tan^{-1}\frac{1}{\eta} \tag{81}$$

The fractional band width is given by

$$\Delta/qa = \Delta/n\pi = \frac{2}{n\pi}\tan^{-1}\frac{1}{\eta} \tag{82}$$

If the scatterers were infinitely strong ($\eta = \infty$), then Eq. (82) yields "sharp levels" ($\Delta = 0$) at $qa = n\pi$; in this case each section of wave medium between the scatterers acts like a "local oscillator" consisting of medium of length a clamped at each end. If the scatterers are large but not infinite (large η), then the coupling between the local oscillators is weak, and the band widths are narrow.

Once the diagonal elements $\lambda^{\pm}$ are found, the matrix D must be found. One solves the following equations, using $\lambda^{+} = \exp(-ika)$:

$$\begin{pmatrix} (\alpha - e^{-ika}) & \beta \\ \beta^{*} & (\alpha^{*} - e^{-ika}) \end{pmatrix}\begin{pmatrix} D_{++} \\ D_{+-} \end{pmatrix} = 0 \tag{83}$$

and finds

$$D_{+-}/D_{++} = \left(e^{-ika} - \alpha\right)/\beta \equiv \gamma_k \tag{84}$$

Using $\lambda^{-} = \exp(+ika)$, one similarily finds $D_{--}/D_{-+} = 1/\gamma_k^{*} = \gamma_{-k}$. Using Eq. (77) we have

$$\gamma_{\pm k} = \frac{1}{i\beta}\left(Im\alpha \pm \sqrt{1 - (Re\alpha)^2}\right) \tag{85}$$

In the case of a continuous wave field, where Eq. (62) applies, we have

$$\gamma_k = \frac{\tan(qa/2) - \tan(ka/2)}{\tan(qa/2) + \tan(ka/2)}e^{iqa} \tag{86}$$

For the coefficients of the matrix D, we normalize with $\sqrt{1 - \mid \gamma_k \mid^2}$, and set

$$D_{++} = 1/\sqrt{1 - \mid \gamma_k \mid^2} \qquad D_{-+} = \gamma_k^{*}/\sqrt{1 - \mid \gamma_k \mid^2}$$

$$D_{+-} = \gamma_k/\sqrt{1 - \mid \gamma_k \mid^2} \qquad D_{--} = 1/\sqrt{1 - \mid \gamma_k \mid^2} \tag{87}$$

Then $\mid D_{++} \mid^2 - \mid D_{+-} \mid^2 = 1, \mid D_{-+} \mid^2 - \mid D_{--} \mid^2 = -1$, and $D_{++}D_{-+}^{*} - D_{+-}D_{--}^{*} = 0$. We now can write

$$D^{-1} = \begin{pmatrix} D_{++} & D_{-+} \\ D_{+-} & D_{--} \end{pmatrix} \tag{88}$$

and

$$D = \begin{pmatrix} D_{++}^{*} & -D_{+-}^{*} \\ -D_{-+}^{*} & D_{--}^{*} \end{pmatrix} \tag{89}$$

We now transform the eigenfunction with the following definitions:

$$\begin{pmatrix} u_j^+ \\ v_j^- \end{pmatrix} = D \begin{pmatrix} u_j \\ v_j \end{pmatrix} \tag{90}$$

The inverse is

$$\begin{pmatrix} u_j \\ v_j \end{pmatrix} = D^{-1} \begin{pmatrix} u_j^+ \\ v_j^- \end{pmatrix} \tag{91}$$

For the new coefficients u_j^+ and v_j^- we have

$$\begin{pmatrix} u_{j-1}^+ \\ v_{j-1}^- \end{pmatrix} = D \begin{pmatrix} u_{j-1} \\ v_{j-1} \end{pmatrix}$$

$$= D \begin{pmatrix} \alpha & \beta \\ \beta^* & \alpha^* \end{pmatrix} \begin{pmatrix} u_j \\ v_j \end{pmatrix}$$

$$= D \begin{pmatrix} \alpha & \beta \\ \beta^* & \alpha^* \end{pmatrix} D^{-1} \begin{pmatrix} u_j^+ \\ v_j^- \end{pmatrix}$$

$$= \begin{pmatrix} e^{-ika} u_j^+ \\ e^{+ika} v_j^- \end{pmatrix} \tag{92}$$

Thus we have

$$u_j^+ = e^{+ijka} u_0^+ = e^{+ikx_j} u_0^+ \tag{93}$$

$$v_j^- = e^{-ijka} v_0^- = e^{-ikx_j} v_0^- \tag{94}$$

If the relations in Eq. (91), (93), and (94) are substituted into the expression for the eigenfunction in Eq. (51), one obtains:

$$\psi_k(x,t) = Re\left[\left(A_k e^{+ikx} U_k(x) + B_{-k} e^{-ikx} U_{-k}(x)\right) e^{-i\omega t}\right] \tag{95}$$

where

$$A_k = \frac{C_k}{\sqrt{1 - |\gamma_k|^2}} u_0^+ \tag{96}$$

$$B_{-k} = \frac{C_{-k}}{\sqrt{1 - |\gamma_k|^2}} u_0^- \tag{97}$$

with

$$C_k = \frac{1 + \gamma_k}{|1 + \gamma_k|} \times$$

$$\left[2\left(1 + |\gamma_k|^2\right) + 4\left(\frac{\sin(qa)}{qa}\right) Re\left(\gamma_k e^{-iqa}\right)\right]^{1/2} \tag{98}$$

and

$$U_k(x) = \frac{1}{C_k}\left[e^{i(q-k)(x-x_j)} + \gamma_k e^{-i(q+k)(x-x_j)}\right] \tag{99}$$

In deriving Eq. (95), we use the relations $1/\gamma_k^* = \gamma_{-k}, (C_k/\gamma_k)^* = C_{-k}$, and $U_k^*(x) = U_{-k}(x)$. Note that $U_k(x)$ is defined piecewise. If x is between x_j and x_{j+1}, then $x + a$ is between x_{j+1} and x_{j+2}. Then

$$U_k(x+a) =$$

$$\frac{1}{C_k}\left[e^{i(q-k)[(x+a)-x_{j+1}]} + \gamma_k e^{-i(q+k)[(x+a)-x_{j+1}]}\right]$$

$$= \frac{1}{C_k}\left[e^{i(q-k)(x-x_j)} + \gamma_k e^{-i(q+k)(x-x_j)}\right]$$

$$= U_k(x) \tag{100}$$

since $x_{j+1} = x_j + a$. Thus $U_k(x)$ is periodic with the period of the potential field. For large η, $U_k(x) \simeq \sin(qx)$ between any pair of masses. Functions of the form $\exp(ikx)\, U_k(x)$ are referred to as Bloch waves.

The coefficient C_k is chosen so that

$$\int_0^a |U_k(x)|^2 \, dx = \frac{a}{2} \tag{101}$$

and the phase of $U_k(0)$ is zero.

The wave function in Eq. (95) is a superposition of two Bloch waves. Note that in deriving this form it was not necessary to have periodic symmetry in an infinite system nor to invoke periodic boundary conditions. The wave function in Eq. (95) is valid for arbitrary boundary conditions.

It should be noted that Bloch waves are not real waves in the sense of displacements of a string; in a Bloch wave it is not the string which is waving, but rather pieces of the function $U_k(x)$ that are waving. Real waves have certain relations with real force, velocity, momentum, energy, etc., and Bloch waves do not have straightfoward relations with such quantities. For example, a Bloch wave traveling in one direction may have energy traveling in the opposite direction.

While Bloch waves are not actual waves in the string, it is possible to form wave packets which do behave like ordinary wave packets traveling with a group velocity $d\omega/dk$, which may be small or negative. It will be shown in a later section that energy in the system travels at the group velocity. With these considerations in mind, one may define a Bloch wave reflection coefficient (with phase referenced at $x = a/2$) as B_k/A_k, using Eqs. (96) and (97).

Bloch waves are often misrepresented as a real periodic function modulated by a phase factor $\exp(ikx)$ or by trigonometric functions $\sin(kx)$ or $\cos(kx)$. For example, it is tempting to write a Bloch wave for a system clamped at $x = 0$ as

$$\psi_k(x,t) = Re\left[A_k \sin(kx)\, U_k(x)\, e^{-i\omega t}\right] \tag{102}$$

However this is not correct, because the periodic functions $U_k(x)$ have a phase which varies with position. As seen in Eq. (99), a part of the phase is $-k(x-x_j)$, which piecewise cancels the phase of $\exp(ikx)$. If we define a phase $\phi(x)$ with

$$e^{i\phi(x)} = e^{ik(x-x_j)} U_k(x) / \mid U_k(x) \mid \tag{103}$$

then

$$\psi_k(x,t) =$$

$$Re\left[\left(A_k e^{i(kx_j+\phi(x))} + B_{-k} e^{-i(kx_j+\phi(x))}\right) \mid U_k(x) \mid e^{-i\omega t}\right] \tag{104}$$

A plot of a typical Bloch wavefunction is shown in Fig. 5a, with a plot of $\mid U_k(x) \mid$ shown in Fig. 5b, and $\sin(kx_j + \phi(x))$ shown in Fig. 5c; Fig. 5c is not what is naively assumed.

A. The Undamped System, Clamped at Each End

If we have a system which is clamped at $x_0 = 0$, then $v_0/u_0 = -1$, and from Eq. (90) we have $v_0^-/u_0^+ = (1+\gamma_k)/(1+\gamma_k)^*$. From Eqs. (96) and (97) we have

$$B_{-k}/A_k = -\frac{(1+\gamma_k)/C_k}{(1+\gamma_k)^*/C_k^*} = -\frac{U_k(0)}{U_k^*(0)} = -1 \tag{105}$$

Now

$$\psi_k(x,t) = Re\left[2A_k \sin(kx_j + \phi(x)) \mid U_k(x) \mid e^{-i\omega t}\right] \tag{106}$$

which satisfies the boundary condition because $\phi(0) = 0$. If the wave field is continuous, then Eq. (86) can be used to show that $\phi(a) = ka$, and $U_k(a) = U_k(0)$. A boundary condition $\psi(x = L, t) = 0$, with $L = (N+1)a$, yields

$$kx_N + \phi(a) = k(N+1)a = kL = \nu\pi \tag{107}$$

so k is quantized as $k = \nu\pi/L$, with $\nu = 0 \cdots\cdots (N+1)$. The eigenfrequencies ω_ν are found by solving Eq. (79) for q and using $\omega = cq$.

The eigenfunction in Eq. (107) is normalized as follows:

$$\int_0^L \sin^2(kx_j + \phi(x)) \mid U_k(x) \mid^2 dx =$$

$$\frac{1}{4}\int_0^L \left(e^{i(kx_j+\phi(x))} \mid U_k(x) \mid - e^{-i(kx_j+\phi(x))} \mid U_k(x) \mid\right)^2 dx$$

$$= \frac{1}{2}\int_0^L \mid U_k(x) \mid^2 dx -$$

$$\frac{1}{2} Re \int_0^L e^{i2(kx_j+\phi(x))} \mid U_k(x) \mid^2 dx$$

$$= \frac{N+1}{2}\int_0^a \mid U_k(x) \mid^2 dx -$$

$$\frac{1}{2} Re \sum_{j=0}^{N} e^{i2kx_j} \int_{x_j}^{x} e^{i2\phi(x)} \mid U_k(x) \mid^2 dx$$

$$= \frac{(N+1)a}{4} -$$

$$\frac{1}{2} Re\left(\int_0^a e^{i2\phi(x)} \mid U_k(x) \mid^2 dx\right) \sum_{j=0}^{N} \left(e^{i2ka}\right)^j$$

$$= \begin{pmatrix} L/4 & when & 0 < ka < & \pi \\ L/2 & when & ka = 0, & \pi \end{pmatrix} \tag{108}$$

The sum in the second term vanishes because of the boundary conditions, except when $ka = 0$ or π. When $ka = 0$ or π, the sum is $(N+1)$ and $\phi(x) = \pi/2$, so that the second term is an additional $L/4$.

B. The Driving Point Impedance for a Periodic System Clamped at the End

The treatment above provides the natural frequencies and normal modes for the undamped, undriven system. We now wish to consider the damped, driven system. In particular, we shall find the driving point impedance for a wave medium with periodic scatterers clamped at $x = L = (N+1)a$. As already discussed, the clamped end condition can be expressed with $u_N = 1$, and $v_N = -e^{i2qa}$. The coefficients at the driven end are

$$\begin{pmatrix} u_0 \\ v_0 \end{pmatrix} = D^{-1} \begin{pmatrix} e^{-iNka} & 0 \\ 0 & e^{iNka} \end{pmatrix} D \begin{pmatrix} 1 \\ -e^{i2qa} \end{pmatrix}$$

$$= \frac{e^{iqa}}{(\mid \gamma_k \mid^2 - 1)^2} \begin{pmatrix} e^{-ikL} u_k^*(a) - \gamma_k^* e^{ikL} u_k(a) \\ \gamma_k e^{-ikL} u_k^*(a) - e^{ikL} u_k(a) \end{pmatrix} \tag{109}$$

where $u_k(a) = e^{-ika}\left(e^{iqa} + \gamma_k e^{-iqa}\right)$. If the wave field is continuous, then Eq. (62) can be used to show $u_k(a) = u_k(0) = (1+\gamma_k)$. In this case we have

$$\begin{pmatrix} u_0 \\ v_0 \end{pmatrix} =$$

$$\frac{-e^{iqa}}{\left(|\gamma_k|^2-1\right)^2}\begin{pmatrix}|\gamma_k|^2 e^{ikL}-e^{-ikL}+\gamma_k^* 2i\sin(kL)\\ -|\gamma_k|^2 e^{-ikL}+e^{ikL}+\gamma_k 2i\sin(kL)\end{pmatrix} \tag{110}$$

Omitting the negligible factor $(1+i/(2Q))$, the driving point impedance is:

$$\frac{Z_0}{\mu c}=\frac{u_0-v_0}{u_0+v_o}=-\frac{|\gamma_k|^2-1}{|\gamma_k+1|^2}\,i\cot(kL)-i\frac{2Im(\gamma_k)}{|\gamma_k+1|^2} \tag{111}$$

In the absence of damping, the impedance would be singular when kL = nπ. Also, the first factor is singular when $qa = n\pi$; note that this is the condition for the normal mode which has a unique normalization.

To study the singularities, we must use the explicit form for γ_k for continuous wave fields (point scatterers). Letting $\theta = qa - n\pi$, where n is the band number, with $n = 0$ being the lowest band (n is approximately the number of half wavelengths which fit between the scatterers), we have

$$\gamma_k=-(-1)^n|\gamma_k|e^{-iqa}=-|\gamma_k|\cos\theta-i|\gamma_k|\sin\theta \tag{112}$$

where

$$|\gamma_k|=\cos\theta+\frac{1}{\eta}\sin\theta-$$

$$\frac{(-1)^n}{\eta}\left[\sin\theta\left(2\eta\cos\theta-\left(\eta^2-1\right)\sin\theta\right)\right]^{1/2} \tag{113}$$

is positive definite. With some algebra one finds

$$\frac{|\gamma_k|^2-1}{|\gamma_k+1|^2}=\left[\frac{2F\cos\theta+\left(F^2-\sin\theta\right)\sqrt{\sin\theta}}{F^2+\sin\theta}\right]\frac{1}{\sqrt{\sin\theta}} \tag{114}$$

where

$$F=\frac{1}{\eta}\sqrt{\sin\theta}-\frac{(-1)^n}{\eta}\left[2\eta\cos\theta-\left(\eta^2-1\right)\sin\theta\right]^{1/2} \tag{115}$$

The function F and the factor in brackets in Eq. (114) are not singular.

When there is damping in the system, the impedance is no longer singular. To incorporate this behavior into the equations, we replace the frequency parameter with the complex parameter $\tilde{q}a = qa(1+i/(2Q)) = qa + i\delta$. Now

$$\sin(\tilde{q}a-n\pi)=\sin\theta\cosh\delta+i\cos\theta\sinh\delta \tag{116}$$

and

$$\frac{1}{\sqrt{\sin\theta}}=\left(S^+-iS^-\right)/S \tag{117}$$

where

$$S=\sqrt{\sin^2\theta+\sinh^2\delta} \tag{118}$$

and

$$S^{\pm}=\sqrt{\left(S\pm\sin\theta\cosh\delta\right)/2} \tag{119}$$

Since the frequency parameter is complex, then the Bloch parameter k is also complex. Writing $\tilde{k}a = k'a + i\kappa$, we have, using Eq. (78):

$$\cos\tilde{k}a=\cos k'a\cosh\kappa-i\sin k'a\sinh\kappa$$

$$=\cos\tilde{q}a-\eta\sin\tilde{q}a$$

$$=\cos ka\cosh\delta-i(-1)^n\left(\sin\theta+\eta\cos\theta\right)\sinh\delta \tag{120}$$

Equating real and imaginary parts yields

$$\kappa=(-1)^n\sinh^{-1}\Bigg[$$

$$\sqrt{\left(\sin^2 ka-\left(\eta^2+1\right)\sinh^2\delta\right)^2/4+\left(\sin^2 ka+\eta^2\right)\sinh^2\delta}$$

$$-\left(\sin^2 ka-\left(\eta^2+1\right)\sinh^2\delta\right)/2\Bigg]^{1/2} \tag{121}$$

and

$$k'a=\cos^{-1}\left(\cos ka\cosh\delta/\cosh\kappa\right) \tag{122}$$

With $k'L = (N+1)k'a$ and $\kappa' = (N+1)\kappa$, we have

$$i\cot\tilde{k}L=\frac{\cosh\kappa'\sinh\kappa'+i\cos k'L\sin k'L}{\sin^2 k'L+\sinh^2\kappa'} \tag{123}$$

Finally we obtain for the real part of the driving point impedance:

$$\frac{1}{\mu c}Re(Z_0)=-\left(\frac{2F\cos\theta+\left(F^2-\sin\theta\right)\sqrt{\sin\theta}}{F^2+\sin\theta}\right)\times$$

$$\left(\frac{\cosh\kappa'\sinh\kappa'S^+ +\cos k'L\sin k'LS^-}{S\left(\sin^2 k'L+\sinh^2\kappa'\right)}\right) \tag{124}$$

A similar expression is obtained for the imaginary part, with S^- replaced with S^+, and S^+ replaced with $-S^-$, and with the second (non-singular) term in Eq. (111) included. The second term in Eq. (111) is given by:

$$-i\frac{2Im(\gamma_k)}{|\gamma_k+1|^2}=i\frac{2\left(\cos\theta+F\sqrt{\sin\theta}\right)}{F^2+\sin\theta} \tag{125}$$

A plot of the real part of the driving point impedance as a function of $qa \geq \pi$ is shown in Fig. 6. The fact that the impedance is much larger at one edge of the band than the other is of crucial importance for the nonlinear system.

C. The Driving Point Impedance for a Periodic System Terminated with Z = μc

In this section we calculate the driving point impedance for a periodic system terminated with an impedance $Z = \mu c$, i.e. the wave medium continues to infinity after N scatterers. Using the D matrices of Eqs. (88) and (89), the coefficients at the beginning are given by

$$\begin{pmatrix} u_0 \\ v_0 \end{pmatrix} = D^{-1} \begin{pmatrix} e^{-iNka} & 0 \\ 0 & e^{iNka} \end{pmatrix} D \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$= \frac{1}{1 - |\gamma_k|^2} \begin{pmatrix} e^{-iNka} - |\gamma_k|^2 e^{iNka} \\ -\gamma_k \left(e^{iNka} - e^{-iNka}\right) \end{pmatrix}$$

$$= \frac{1}{|\gamma_k|^2 - 1} \times$$

$$\begin{pmatrix} \left(|\gamma_k|^2 - 1\right)\cos(Nka) + i\left(|\gamma_k|^2 + 1\right)\sin(Nka) \\ i2\gamma_k \sin(Nka) \end{pmatrix} \tag{126}$$

where γ_k is given by Eq. (85) (or 4.62 and 4.63). We have

$$|u_0|^2 - |v_0|^2 = \cos^2(Nka) + \sin^2(Nka) = 1 \tag{127}$$

if k is real. Damping can be added by making k complex, as in the previous section. We use $\gamma_k = -(-1)^n |\gamma_k| \exp(iqa) = -|\gamma_k| \exp(i\theta)$ where $\theta = qa - n\pi$ and n is the band number. Then

$$\frac{1}{\mu c} ReZ_0 = \frac{|u_0|^2 - |v_0|^2}{|u_0 + v_0|^2}$$

$$= \left(|\gamma_k|^2 - 1\right)\Big[\left(|\gamma_k|^2 - 1\right)\cos^2(Nka) +$$

$$4\left(|\gamma_k|^2 - 1\right)|\gamma_k| \sin\theta \cos(Nka)\sin(Nka) +$$

$$\left[\left(|\gamma_k|^2 + 1\right) - 4\left(|\gamma_k|^2 + 1\right)|\gamma_{k|} \cos\theta + 4|\gamma_k|^2\right] \times$$

$$\sin^2(Nka)\Big]^{-1}$$

$$= \Bigg[\cos^2(Nka) + \left(\frac{4|\gamma_k|\sin\theta}{|\gamma_k|^2 - 1}\right)\cos(Nka)\sin(Nka) +$$

$$\left[\left(\frac{|\gamma_k| - 1}{|\gamma_k| + 1}\right)^2 + \left(\frac{4\left(|\gamma_k|^2 + 1\right)|\gamma_k|(1 - \cos\theta)}{\left(|\gamma_k|^2 - 1\right)^2}\right)\right] \times$$

$$\sin^2(Nka)\Bigg]^{-1} \tag{128}$$

D. Expressions for Magnitudes of Various Quantities

For purposes of estimation, plotting graphs, etc., it is useful to have some expression for the nominal magnitude for various quantities, such as the driving point impedance, power delivered, etc. Of course, the transmission and reflection coefficients have maximum amplitudes of one. The spectrum of power delivered by a force, normalized as in Eq. (67), is roughly equal to the transmission spectrum, and also has a maximum amplitude of one. The spectrum of power delivered by a velocity drive, given by Eq. (68), is a more complex problem. If the system has no damping and a purely reactive end impedance Z, then the the spectrum is singular. If a periodic system has purely reactive end conditions but has damping in the wave medium, then it is possible to find the maximum power delivered, and this can be used as an upper bound for disordered systems.

It is found that the maximum power for a periodic system occurs when $qa = \pi$, and the wavefunction for this frequency is unperturbed by the scatterers because the nodes of the wavefunction are at the sites of the scatterers. In this case the driving point impedance is the same as that for a simple one dimensional wave medium with damping. As found in common acoustic texts, the impedance will have a denominator which is a trigonometric function of $\tilde{q}L = qL - i\alpha L$, where $L = (N+1)a$ is the length of the system. Assuming α is small, it can be shown that the resonance peaks in $ReZ_0/\mu c$ have a maximum value of $1/\alpha L$ and a quality factor $Q = \omega/2\alpha c$. If the Q is a known parameter, then $ReZ_0/\mu c$ scales as

$$\frac{1}{\alpha L} = \frac{2Qc}{\omega}\frac{1}{L} = \frac{2Q}{q}\frac{1}{Na} = \frac{2Q}{N\pi} \tag{129}$$

It can also be shown that the maximum amplitude of the wavefunction, divided by the amplitude at the drive $|u_0 + v_0|$, is also $1/\alpha L = 2Q/N\pi$. For $qa = \pi$, the maximum amplitude occurs for $x - x_j = a/2$, and is given by $|u_0 - v_0|$. Thus $\max|Z| = \max(|u_0 - v_0|/|u_0 + v_0|) = 2Q/N\pi$.

When a system has no damping but is terminated with an impedance with resistance, then $ReZ/\mu c$ is not singular, but its maximum is difficult to calculate because there are no standing wave solutions with nodes at the sites of the scatterers. A case of interest is when the system is terminated with an impedance μc, i.e. after N scatterers the wave medium continues to infinity. The analytic equation for this case can be studied for the two limits of strong scatterers and weak scatterers, and an expression for the maximum of $ReZ/\mu c$ can be found.

We consider only the first band and let $\theta = qa - \pi$ and $\delta = \pi - ka$. We first assume strong scatterers, so η is large, qa is limited to a narrow band, and $\theta << 1$. We let $\sin\theta \simeq \theta$ and $1 - \cos\theta \simeq \theta^2/2$. From Eq. (113)

$$|\gamma_k| = \cos\theta + \frac{1}{\eta}\sin\theta +$$

$$\frac{1}{\eta}\left[\sin\theta\left(2\eta\cos\theta-\left(\eta^2-1\right)\sin\theta\right)\right]^{1/2}$$

$$\simeq 1+\sqrt{2\theta/\eta} \tag{130}$$

We let $\mid \gamma_k \mid -1 \simeq \sqrt{2\theta/\eta}$. To second order in $\sqrt{2\theta\eta}$ Eq. (128) becomes

$$\frac{1}{\mu c}ReZ_0=\left[\cos^2(N\delta)-\sqrt{2\theta\eta}\cos(N\delta)\sin(N\delta)+\right.$$

$$\left.\left(\frac{1+\eta^2}{4\eta^2}\right)(2\theta\eta)\sin^2(N\delta)\right]^{-1}$$

$$=\left[\left(\cos(N\delta)-\frac{1}{2}\sqrt{2\theta\eta}\sin(N\delta)\right)^2+\right.$$

$$\left.\frac{1}{4\eta^2}(2\theta\eta)\sin^2(N\delta)\right]^{-1} \tag{131}$$

In terms of θ and δ, the dispersion relation is $\cos\delta = \cos\theta - \eta\sin\theta$. It is found that the maximum in the spectrum occurs at the beginning of the band, so that δ is small. Thus

$$2\theta\eta \simeq 2\eta\sin\theta = 2(\cos\theta-\cos\delta) \simeq 2(1-\cos\delta) \simeq \sin^2\delta \tag{132}$$

Eq. (131) for $ReZ_0/\mu c$ is approximately maximum when the first term in parentheses vanishes:

$$\left(\cos(N\delta)-\frac{1}{2}\sin\delta\sin(N\delta)\right)=0 \tag{133}$$

To a good approximation, this transcendental equation has the root $\delta = \pi/(2N+1)$. Evaluating Eq. (131) for this value yields

$$\max\left(\frac{1}{\mu c}ReZ_0\right)=\frac{16}{\pi^2}\eta^2\left(N+\frac{1}{2}\right)^2 \tag{134}$$

This is a good approximation for $\eta N > 2$.

We now consider the limit of weak scattering, or small η. For the first band the dispersion relation becomes $\cos\delta = \cos\theta - \eta\sin\theta \simeq \cos\theta$; thus $\sin\theta \simeq \sin\delta$. Since $\mid \gamma_k \mid$ has terms proprtional to $1/\eta$, then we take $\mid \gamma_k \mid$ to be large and expand Eq. (128) to first order in $1/\mid \gamma_k \mid$:

$$\frac{1}{\mu c}ReZ_0=\left[1-\frac{4}{\mid \gamma_k \mid}\sin(N\delta)\left[\sin\delta\cos(N\delta)+\right.\right.$$

$$\left.\left.\left(1-\frac{1}{2}\sin^2\delta\right)\sin(N\delta)\right]\right]^{-1} \tag{135}$$

This is maximum when the inside term in brackets is maximum. This occurs approximately when $\delta = \pi/2(N+1)$, and we have

$$\max\left(\frac{1}{\mu c}ReZ_0\right)=\left[1-\frac{4}{\mid \gamma_k \mid}\right]^{-1} \tag{136}$$

When $\eta = 0.1$ and $N = 2$, then $\mid \gamma_k \mid \simeq 10$ and $ReZ_0/\mu c \simeq 1.6$. However, for any small value of ηN $ReZ_0/\mu c$ is between 1 and 2, and it is found that a satisfactory approximation is obtained by simply modifying the earlier expression (Eq. (134)) for the limit of large η. Thus we use

$$\max\left(\frac{1}{\mu c}ReZ_0\right)=\frac{16}{\pi^2}\left[1+\eta^2\left(N+\frac{1}{2}\right)^2\right] \tag{137}$$

This also has the value $\sim$1.6 when $\eta = 0.1$ and $N = 2$, and works well for all η and N.

V. SYSTEMS WITH DISORDER: ANDERSON LOCALIZATION

If one takes the modulus of an eigenfunction for a periodic system (e.g. Eq. (29)), evaluated at some nominal point in each unit cell, then one obtains a constant, independent of position in the system. That is, the eigenfunction has the same nominal amplitude throughout the entire system. Such eigenfunctions are said to be extended. Extended eigenfunctions are very important in solid state theory; for example, they explain why metallic crystals have high electrical conductivity. In a metal, an electron is very strongly scattered by the positive ions, and on this basis a metal should have a very low electrical conductivity. But because the ions in a crystal are arranged periodically, the quantum mechanical eigenfunctions for the electron are extended, allowing the electron to propagate freely and resulting in high electrical conductivity.

That the eigenfunctions for a periodic system are extended is guaranteed by Bloch's theorem. We now consider what happens if we take a periodic system and introduce some disorder throughout the system. One would suppose that the eigenfunctions, instead of having uniform amplitude everywhere in the system, would simply obtain some variations about the original constant value, but would otherwise remain extended. Surprisingly, this is not what happens; instead, the eigenfunctions become exponentially locallized, having a maximum value at some site in the system and, at least in a probabilistic sense, decaying exponentially from that site. This phenomena, referred to as Anderson localization, was used by P. W. Anderson and Sir Neville Mott to explain the electrical conductivity of disordered metals, and this accomplishment was cited in their Nobel prize in 1977. [10]

While much of the research with Anderson localization has dealt with electron Schrodinger waves in disordered

metals, the phenomenon is relevant for any form of wave propagation in a disordered system. Anderson localization is important for the understanding of phonon quantum wave propagation and its effects on specific heat and thermal conductivity, and for the understanding of classical wave propagation in disordered systems, as in acoustic geophysical survey, oil exploration, medical imaging, the vibration of complex structures, sound propagation in the ocean, radar propagation in vegetation, etc. Anderson localization remains an active area of research for both quantum and classical systems in condensed matter physics.

Many of the advances in the theory of Anderson localization have occurred through the application of renormalization group methods and other modern techniques of statistical physics. As with critical phenomena in statistical physics, Anderson localization is strongly dependent on the dimensionality of the system. In one dimension it is found that any amount of disorder will result in some degree of localization; very weak disorder will result in localization, but the characteristic length for the exponential decay may be large. In three dimensions it is believed that a critical amount of disorder must be exceeded before all of the eigenstates become localized. Two dimensions is the "critical dimension", where theory becomes difficult; it is believed that in two dimensions all eigenstates are localized, as in one dimension, but the localization is much weaker. In one dimension there are rigorous theorems which describe Anderson localization, but in two and three dimensions there are no comprehensive rigorous theorems. In the paragraphs below Anderson localization in one dimension, based on the coupled oscillator model described in the first sections of this paper, will be treated; some comments on two and three dimensions will follow.

While it is easy to describe Anderson localization, it is very difficult to understand how it occurs; one would naively expect that disorder, especially weak disorder, would simply put variations in the amplitude of the otherwise extended eigenfunctions. Of the many papers treating Anderson localization, most simply assume the phenomena as an established fact and few offer explanations. The rigorous theory for one dimension was found by Furstenburg in 1963; [11] this is a long and difficult paper, with references to earlier theorems, and so it offers little physical insight into the phenomenon. Fortunately there is one excellent paper, by Marshall Luban and James Luscombe, [12] which presents a clear and understandable model of Anderson localization. After some preliminary observations, this model will be described.

A careful statement of the Anderson localization effect reveals some of the difficulty. Anderson localization does not refer to the behavior of a single realization of a disordered system, but rather refers to the results of averaging over an ensemble of realizations, with the disorder of each realization randomly determined with some distribution function. Furthermore, the rigorous result occurs in the limit as the system becomes infinite in size. The actual eigenfunctions of a single realization do not have the form $\exp(-2 \mid x - x_l \mid /\Lambda)$ (here x_l would be the site of localization and Λ would be the localization length); in an actual realization, an eigenfunction at some site far from x_l may fluctuate above the small exponential tail, requiring an unbounded prefactor to the exponential. When averages are taken, the fluctuations can be ignored, and the eigenfunctions, while not actually having the form $\exp(-2 \mid x - x_l \mid /\Lambda)$, behave as though they did, in some statistical sense.

In the treatment of the coupled oscillators presented in the first sections, it was found that the eigenfunctions could be expressed as products of matrices, as in Eq. (9). For an ensemble of disordered systems, we consider a product of random matrices for a single realization, which we write symbolically as

$$P_j = \prod_{j'=1}^{j} M_{j'} \tag{138}$$

We are interested in the ensemble average $\bar{P}_j$ of the product of random matrices. While this is a difficult mathematical problem, we consider an analogy with a much more familiar problem, the ensemble average of a sum of random variables $X_{j'}$ with distribution function $F(X)$:

$$S_j = \sum_{j'=1}^{j} X_{j'} \tag{139}$$

We expect from intuition, and know from rigorous theorems, that the average of the sum is just the average of the random variable, $\bar{X} = \int XF(X)\,dX$, times j :

$$\bar{S}_j = j\bar{X} \tag{140}$$

in the limit as j goes to infinity. Comparing Eq. (138) with Eq. (139) we imagine taking the log of the product of random matrices and obtain a sum of random variables:

$$\ln(P_j) = \sum_{j'=1}^{j} \ln(M_{j'}) \tag{141}$$

We might then expect the average of $\ln(P_j)$ to be j times some mean $\bar{M}$:

$$\overline{\ln(P_j)} = j\bar{M} \tag{142}$$

Exponentiating both sides of Eq. (142) suggests:

$$\bar{P}_j = e^{j\bar{M}} \tag{143}$$

which indicates that the ensemble average of a product of random matrices may grow or decay exponentially as j goes to plus or minus infinity. This effect might also be seen in Eq. (13) or (76), showing the Singular Value Decompositions of a product of random matrices. If the product in Eq. (13) or (76) is ensemble averaged, one

might expect that the combinations $D\cdots D^{-1}$ would cancel; the diagonal matrices cannot cancel, leaving a matrix of the form:

$$P_j = \begin{pmatrix} \Pi_{j'=1}^{j}\lambda_{j'}^{+} & 0 \\ 0 & \Pi_{j'=1}^{j}\lambda_{j'}^{-} \end{pmatrix} \tag{144}$$

For the periodic system we would have $\mid \lambda_{j'}^{\pm} \mid = 1$, resulting in extended eigenfunctions. However, for the disordered system we have $\mid \lambda_{j'}^{-} \mid > 1$ and $\mid \lambda_{j'}^{+} \mid < 1$, so that in Eq. (144) the singular values (and the eigenfunctions) grow or decay exponentially.

The above discussions of the average of a product of random matrices are not rigorous, but they do suggest what Furstenburg proved rigorously. The statistical nature of the Anderson localization is actually stated somewhat differently using Furstenburg's theorem: for waves propagating in a disordered system, the eigenfunctions decay exponentially to zero at plus or minus infinity with probability one. This probabilistic statement permits actual wavefunctions for a given realization to fluctuate above an exponential envelope.

The problem with the rigorous statement of Anderson localization is that it is difficult to see how to apply it to practical situations. That is, real experiments may have only one finite-size realization available; one cannot examine the system out to infinity, and even if one could it would not be evident if something were happening with "probability one".

A treatment of Anderson localization which illustrates the localization and its probalistic nature in a single finite-size realization is the one by Luban and Luscombe. [12] In this treatment one considers the effects of the rigorous theorems on a computer evaluation of a finite product of random matrices.

For the Luban and Luscombe approach, one first notes that the problem, as represented by Eq. (3) (involving a coupling over two sites) or by the wave equation (Eq. (54)), is second order and hence must have two linearly independent solutions, which we write as R_j and L_j. Using the periodic system of masses coupled with massless springs as an example,these two solutions are $R_j = \exp(ikja)\,U_k$ and $L_j = \exp(-ikja)\,U_{-k}$, which are Bloch waves propagating to the right and left respectively. For the disordered system the solutions are those which, if one could continue them to infinity, would grow or decay exponentially with probability one. The Wronskian of the discrete problem is

$$W = L_j R' - R_j L'$$

$$= \left(L_j R_{j-1} - L_{j-1} R_j\right)/a_{j-1} \tag{145}$$

where the prime indicates the derivative with respect to site, defined as

$$y' = \frac{1}{2}\left[\left(y_j - y_{j-1}\right)/a_{j-1} + \left(y_{j+1} - y_j\right)/a_j\right] \tag{146}$$

and the distance between sites is $a_j = T/s_j$, as in the discussion of the continuum limit for masses coupled with massless springs. One notes that the Wronskian is a constant in the problem, independent of j. (A similar result is obtained for the continuous system.) Thus if one of the linearly independent solutions exponentially decays, then the other must exponentially grow to keep W constant. The actual solution which one has depends on the end conditions y_0 and y_1. We now suppose that we know exactly the end conditions which would give the solution which would exponentially decay with probability one as j goes to infinity. We put these end conditions into a computer program which iteratively multiplies the matrices with increasing j. We would find that after a number of iterations the eigenfunction would begin growing rather than decaying. The reason is that the computer has a finite digital word size, and that even if we knew the exact end conditions for the exponentially decaying solution, the digital end conditions which we give the computer will necessarily contain a small admixture of the solution which grows exponentially. After a number of iterations, usually less than a hundred in actual applications, the exponentially growing solution will dominate the exponentially decaying solution, and the exponentially decaying solution will become lost in the numerical roundoff. This is a well known phenomenon in numerical computation, which occurs for example when using iteration schemes to calculate special functions.

We now note that we can use the computer to find the exponentially increasing solution without actually observing it exponentially grow to infinity. We simply start with arbitrary end conditions and iterate a number of times. We then change the end conditions and iterate from the beginning again; if the latter part of the eigenfunction is the same in both cases (within some small numerical limit) then we can conclude that the latter part of the eigenfunction is the one which would exponentially grow if we continued to infinity. If the latter part of the eigenfunction is different, then we simply increase the number of iterations until we have enough to allow the growing solution to dominate.

The Anderson localized eigenfunctions which we are seeking are solutions which exponentially decay in both directions; i.e. with increasing j to plus infinity, and decreasing j to minus infinity. We have found that the computer has the ability to find exponentially increasing solutions, and we use this ability to find the localized eigenfunction in the following way. We first pick some value for the eigenvalue parameter (ω in Eq. (3)). We then take arbitrary end conditions at a low value of j (e.g. $j = 0$) and iterate with increasing j until we reach a site j_c where we find the exponentially increasing solution dominating. We note that the latter values of y_j $(j \lesssim j_c)$ which we have found are those which would exponentially decay with probability one if we had infinite numerical precision and iterated in the opposite direction, decreasing j to minus infinity. We next jump to a large value of j $(> j_c)$, again use arbitrary end conditions, and use

the inverse of the matrices to iterate with decreasing j until we again reach the site j_c. If we have not found the dominating solution at this point, we simply start again, jumping to a larger value of j. We thus find values y_j $(j \tilde{>} j_c)$ which would decay with probability one if we had infinite numerical precision and iterated in the direction of j increasing to plus infinity. We now note that the values of the displacement at the overlapping site j_c from the forward and reverse iterations will not necessarily agree; thus our two solutions do not form an eigenfunction. We must now slightly change the value of the eigenvalue parameter ω and try again. Eventually we find a value ω for which the increasing and decaying solutions match within some acceptable precision; we have thus found an eigenvalue and an eigenfunction. The entire processes can be repeated to find a set of eigenvalues and eigenfunctions.

It should be noted that in this scheme it is not necessary to use infinite values of j. Solutions which would exponentially decay as j goes to plus or minus infinity are obtained by using reverse iteration and finding solutions which dominate after a finite number of sites. It should also be noted that it is not necessary that the solutions have a monotonic exponential decay or growth. The solutions may fluctuate by large amounts; all that is required is that one solution dominate over the other. Finally, it should be noted how the probalistic nature of Anderson localization is accounted for with the numerical roundoff of the computer. From our understanding of numerical roundoff we are confident that the computer will find the exponentially increasing solution, that is, it will find it with "probability one".

The most common description of Anderson localization, namely that "the eigenfunctions are exponentially localized" is not particularily useful for single realizations of finite size systems. The eigenfunctions only behave as though they were exponentially localized if many systems are ensemble averaged. A better way of describing the effects of disorder is with a quantity called the participation ratio, $P(N)$, (or $P(L)$) which is defined with

$$P(N) = \frac{\left(\sum_0^N \mid y_j \mid^2\right)^2}{\left(\sum_0^N \mid y_j \mid^4\right)} \tag{147}$$

for eigenfunctions of discrete systems, or

$$P(L) = \frac{\left(\int_{-L/2}^{L/2} \mid \psi(x) \mid^2 dx\right)^2}{\left(\int_{-L/2}^{L/2} \mid \psi(x) \mid^4 dx\right)} \tag{148}$$

for continuous systems. The participation ratio is a measure of the localization of a function: a function which is spread out has a larger value of $P(L)$ than a function which is more localized. If $\mid \psi(x) \mid = \psi_0$ then $P(L) = L$, i.e. the entire system participates. If $\mid \psi(x) \mid = \psi_0 \exp(-2 \mid x \mid / \Lambda)$, then $P(L) \to \Lambda$, the localization length, as $L \to \infty$. A pratical method of determining whether a particular eigenfunction is localized or not is to monitor $P(L)$ as L is increased beyond an anticipated localization length Λ. If $P(L)$ keeps increasing as L is increased, then the state is not localized; if $P(L)$ levels off at a finite value as L is increased, then the state is localized. In either case $P(L)$ may fluctuate about L or Λ; the actual eigenfunction only behaves as though it were of the form $\exp(-2 \mid x \mid / \Lambda)$ if the fluctuations are averaged out.

A useful statement of Anderson localization, at least for one dimensional systems, is the following:

Given any quantity A and any parameter B relevant to a wave system, if the system is disordered, then A is a rapidly varying function of B. The dynamic range of A is so large that A as a function of B must be plotted with a logrithmic scale.

Instances of applying this statement of Anderson localization are: the transmission coefficient for a wave traversing a finite array of scatterers having fixed disorder is a rapidly varying function of the frequency of the wave, or the total transmission coefficient integrated over frequency is a rapidly varying function of the amount of disorder, etc. Some examples will be presented in the section on numerical calculations.

One of the comments made about one dimensional disordered systems is that "all states are localized". There are some pathological exceptions, such as the case when point scatterers have random strengths but equal spacing; the unique states which haves nodes at the scattering sites are clearly extended. However, such states comprise a "set of measure zero". Of greater significance is the case of a disordered potential field which is "spatially band limited". That is, one could take a disordered potential field, $V(x)$, perform a spatial Fourier transform, roll-off the transform to zero above some cutoff k_c, and then inverse transform. The resulting potential field would still be disordered, but states with wavevectors greater than k_c would be extended, not Anderson localized. [13]

A. Anderson Localization in Two and Three Dimensions: The Scaling Argument

The properties and effects of Anderson localization in one dimension are firmly established with rigorous theorems. As already mentioned, in two and three dimensions there are no rigorous theorems. There is a scaling argument [14] which treats one, two, and three dimensions, and this argument provides an adequate explanation for the behavior of electrons in disordered solids. However, the scaling argument is based on an assumption which may be valid for electrons in solids at finite temperatures, but which may be invalid for ideal (lossless) waves in disordered systems in the regime of strong localization. Some of the results of the scaling argument are known to be incorrect in one dimension, [15] and one of the most important scaling argument results in two dimensions is currently being questioned. Indeed, more than twenty

years ago there were computer simulations for 2D systems which disagreed with the predictions of the scaling argument, [16] but the disagreement was attributed to an effect of the finite size of the computer simulations, whereas the argument was supposed to be appropriate for infinite systems. Very recently there have been new large scale computer simulations which reveal an entirely new aspect of the problem. [17]

The scaling argument of localization was developed in order to explain experimental measurements of the electrical conductivity of disordered metals. The argument uses renormalization group techniques to examine how the conductivity of a finite size sample of the metal varies as the sample size is systematically scaled up toward infinite size. A key assumption of the argument is that as the size of the sample becomes infinite, the conductance of the sample begins to obey Ohm's law. For describing experimental conductivity measurements, this is a valid assumption, because the experiments are performed at low but finite temperatures so that there is always some probability of inelastic scattering of the electrons. Inelastic scattering destroys the wave nature of the electrons and results in the validity of Ohm's law when the sample size exceeds the mean free path for the inelastic scattering. However, the fundamental problem in which we are interested concerns the behavior of waves in a static disordered system, where there is absolutely no inelastic scattering or attenuation of the waves, even at boundaries. The assumption of Ohm's law for large size samples is equivalent to the approximation that waves in a disordered system diffuse, at least for the purpose of calculating ensemble averaged quantities. This is based on the supposition that ensemble averaging or averaging over initial conditions can convert the wave equation to the Boltzman transport equation.

When a wave propagates through a disordered system of lossless (elastic) scatterers, the result is a complicated interference pattern. It is usually assumed that ensemble averaging over many interference patterns will result in a smooth pattern which could be predicted with a diffusion equation. However, in the "derivation" of the Boltzman transport equation from the wave equation, it is necessary to average while the wave is propagating, rather than after all of the interference has been established. In the case of pure elastic scattering the derivation is not rigorous. Boltzman himself was aware that his transport equation was based on an unproven ansatz. In higher dimensions when there is sufficient inelastic scattering, the ansatz is a good approximation, but there are still counterexamples such as coherent backscatter and Anderson localization itself.

Coherent backscatter was treated in detail by researchers in other fields prior to 1966. [18] The effect was rediscovered by researchers interested in quantum (wave mechanical) transport for electrons in disordered mescopic wires. The effect was greatly publicised and called "weak localization". [19] Coherent backscatter is a consequence of the interference of waves, as is Anderson localization, but this does make coherent backscatter and localization in any way equivalent. In any case, the theoreticians treating mesoscopic electron wave systems brought great progress to the field of quantum transport. It should be noted that in the mesoscopic electron systems there is always significant inelastic scattering so that ensemble self-averaging and the Boltzman ansatz are valid. In some classical systems, this is not the case.

In any case, one dimension is pathalogical. If there is disorder, then most eigenstates are localized and there is no diffusion. Strictly speaking, one should not refer to an "elastic scattering length" or a "diffusion length" in a one dimensional disordered system; in such a system the Anderson localization length is the only relevant length.

One of the aspects of wave propagation in disordered media which was supposed to have been solved by the scaling argument involves the effects of dimensionality on the nature of the eigenfunctions for different amounts of disorder. A major prediction of the scaling argument was that all eigenstates in a two dimensional disordered system were localized, as in one dimension; i.e. there can be no "Anderson transition" in two dimensions. [14] It was this prediction for two dimensions which was contradicted by early numerical simulations. Now recent computer simulations [17] have indicated a new nature for eigenfunctions in higher dimensions which indicates a new type of transition for two dimensions and also suggests different critical behavior for the Anderson transition in three dimensions.

The new nature of the eigenfunctions is revealed not by considering the effects of eigenfunctions after ensemble averaging, but instead by examining actual eigenfunctions in individual realizations of the disordered system. The recent computer simulations on large (1000x1000) systems indicate that actual eigenfunctions are fractal structures, which may or may not percolate throughout an infinite system. Thus as the amount of disorder is varied in an infinite two dimensional system, there may not be an Anderson transition of "ensemble averaged eigenfunctions", but there may be a percolation transition of real eigenfunctions. This is consistent with the observations of the other computer simulations which detected a transition in two dimensions. [16] In three dimensions the critical nature of the Anderson transition may be altered by the possibility of the percolation transition. A rigorous understanding of Anderson transition in two and three dimensions is currently not available. For a review of the situation, see Ref. [20].

VI. NUMERICAL CALCULATIONS FOR A ONE-DIMENSIONAL SYSTEM

While the formulas for the periodic system may be complicated, they can at least be written in a few lines, even for arbitrarily large systems. For the disordered system, it is possible to write a closed form solution for

the driving point impedance, etc., using the transmission line formula. That is, given the impedance Z_j at a site x_j, one could back up through the wave medium of length a_j and obtain an impedance given by

$$Z_{j-1} = \mu c \frac{i\mu c + Z_j \cot(qa_j)}{iZ_j + \mu c \cot(qa_j)} \tag{149}$$

The impedance of a scatterer at this site could be added, and the formula used again. The process could be used repetitively, connecting the end impedance Z at x_{N+1} to the driving point impedance Z_0 at x_0. However, for large systems, the formula would be quite unwieldly. For practical purposes, it is simpler to numerically compute using the multiplication of two-by-two complex matrices indicated in Eq. (65). The driving point impedance given by Eq. (66) can be used to determine power or transmission resonance frequencies, and if the intermediate results of the matrix multiplications are saved, one also obtains the eigenfunctions and the participation ratio (Eq. (147), with y_j being the wavefunction evaluated at a particular point in each section).

Examples of calculations for periodic and random systems are shown in Figs. 7, 8, and 9. Fig. 7a shows the real part of the driving point impedance for a periodic system of 32 scatterers, each with reflection coefficient of .99, with no damping, but with the system terminated with an impedance μc (i.e. the wave medium continuing to infinity after N scatterers). The first three peaks have been clipped so as to fit in the figure; the expression for the maximum peak value is given in Eq. (137). The first (maximum) peak is 50 times larger than the fourth peak (not clipped in Fig. 7a). The peaks are relatively narrow (requiring fine numerical scans in qa) because the only dissipation is due to the resistance μc at the end of the system. The position of the first, sharpest peak is given by Eq. (133).

Fig. 7b shows the real part of the driving point impedance for a system similar to that if Fig. 7a, but with the end clamped, and a Q of 2000. This figure, calculated by repeated matrix multiplication, is identical to Fig. 6, calculated with the analytic formula in Eq. (124).

Figs. 7a and b are proportional to the power delivered by a constant velocity drive, $V_0 \cos\omega t$. Fig. 7c is for a system identical to that of Fig. 7b, but with a constant force drive, $F_0 \cos\omega t$.

Fig. 8a shows the transmission spectrum (transmission coefficient as a function of frequency) for the band corresponding to qa $\sim \pi$, for a periodic system of 64 scatterers, each with reflection coefficient of .99, with no damping, but terminated with an impedance μc. Dispersion (the reduced spacing between the peaks near the band edges) is evident in the figure. Figs. 8b and c are eigenfunctions for qa near the band edge and near the band center, respectively. Figs. 8d and e are pulses, to be discussed later.

Fig. 9 shows plots for a system similar to that of Fig. 8, but with 2% disorder in the spacing of the scatterers. Fig. 9a is the transmission spectrum. The total transmission has significantly decreased. Since there is no damping in this system, waves for the most part are reflected at the entrance; disorder does not cause incoming waves to become attenuated or localized, but instead they simply never enter the system. It should also be noted that the transmission peaks are relatively narrow. This is related to a property of an infinite disordered system: Usually we have that a finite size system has a spectrum with discrete eigenvalues, while an infinite system has a continuous spectrum. However if there is disorder, then even for the infinite system, the spectrum is discrete, or "pure point".

Fig. 9b and c show examples of Anderson localized eigenstates corresponding to two peaks in the transmission spectrum. If a system has damping and is driven at one end, then virtually all of the states will be localized near the driven end, as in Fig. 9b. The reason is that delivering power, $F \cdot V$, to a state requires sufficient response amplitude at the drive, but states localized at distant sites are not likely to satisfy this condition.

Experimental measurements for periodic and disordered systems may be found in Ref. [21].

VII. THE BEHAVIOR OF PULSES IN PERIODIC AND DISORDERED SYSTEMS

Except for a few instances, the behavior of pulses in systems of scatterers is considerably more difficult to comprehend than continuous wave behavior. For a linear system, calculating the time evolution of a pulse is readily formulated: it is simply the Fourier transform of the continuous wave result, possibly with some window function $W(\omega)$. For example, if a delta function pulse is incident upon a system, then the time evolution of the displacement at the exit of a system is the Fourier transform of the transmission spectrum $T(\omega)$. However, spectra such as those shown in Figs. 8 and 9 do not present obvious clues as to how their transforms will behave, especially if a number of bands are included. Indeed, there are some myths about the behavior of pulses which are incorrect in one dimension and which have counterexamples in higher dimensions. As an example, it is often stated that pulses "diffuse" through systems of disordered scatterers. The approximations of the diffusion equation are discussed in the section on Anderson localization.

One aspect of pulses in systems of scatterers which can be treated rigorously involves the group velocity. This is particularly useful in the case of a periodic system when the width of the pulse exceeds the lattice constant. Before treating this case, we first consider the notion of group velocity in some detail.

A common non-rigorous treatment of group velocity is as follows. We suppose that a pulse may be described as a Fourier transform of a transmission spectrum $T(\omega)$:

$$\psi(x,t) = \int W(\omega) T(\omega) e^{i(kx-\omega t)} d\omega \tag{150}$$

If there is dispersion, then k is a function of ω. If the argument of the exponential is large for some value of x and t, then the integrand oscillates rapidly and causes cancellation in the integration over ω, so that the resulting $\psi(x,t)$ is small. It may be assumed that the peak of the pulse occurs for values of x and t such that the argument of the exponential is minimum. This is given by

$$\frac{d}{d\omega}(kx - \omega t) = \frac{dk}{d\omega}x - t = 0 \tag{151}$$

This relationship between x and t indicates that the peak of the pulse moves with a velocity given by $d\omega/dk = v_g$, the group velocity.

Another non-rigorous treatment involves the superposition of two simple harmonic traveling with wavevectors and frequencies (ω_1, k_1 and ω_2, k_2) which differ slightly. Using a trigonmetric identity, we have

$$\psi(x,t) = \cos(k_1 x - \omega_1 t) + \cos(k_2 x - \omega_2 t)$$

$$= 2\cos\frac{1}{2}(\Delta k x - \Delta\omega t)\cos(kx - \omega t) \tag{152}$$

where k, ω are the average wavevector and frequency, and $\Delta k, \Delta\omega$ are their differences. It can be seen that the envelope of the superposition travels with the velocity $\Delta\omega/\Delta k \simeq d\omega/dk = v_g$.

A rigorous treatment of group velocity is based on Foster's theorem, which involves wave energy density and power flux. Before examining Foster's theorem, it will be useful to review the relationship between wave energy density and power flux.

We begin with Newton's law for continuous media

$$\rho\left(\frac{\partial\vec{v}}{\partial t} + \vec{v}\cdot\vec{\nabla}\vec{v}\right) = -\nabla\vec{P} \tag{153}$$

and conservation of mass

$$\frac{\partial\rho}{\partial t} + \vec{\nabla}\cdot(\rho\vec{v}) = 0 \tag{154}$$

where ρ is the mass density, P is the pressure, and $\vec{v}$ is the particle velocity. For small displacements from equilibrium we can neglect the second term in Eq. (153), and write $dP = (\partial P/\partial\rho)\, d\rho = c^2 d\rho$. Now

$$\frac{\partial P}{\partial t} = c^2\frac{\partial\rho}{\partial t} = -\rho c^2\vec{\nabla}\cdot\vec{v} \tag{155}$$

For an arbitrary volume V enclosed by a surface S, the total energy inside is the volume integral of the energy density:

$$E = \int_V \frac{1}{2}\left(\rho v^2 + \frac{1}{\rho c^2}P^2\right) dV \tag{156}$$

The time rate of change of the energy is, using Eqs. (153) and (155)

$$\frac{dE}{dt} = \int_V \left(\vec{v}\cdot\rho\frac{\partial\vec{v}}{\partial t} + \frac{1}{\rho c^2}P\frac{\partial P}{\partial t}\right) dV$$

$$= -\int_V \left(\vec{v}\cdot\vec{\nabla}P + P\vec{\nabla}\cdot\vec{v}\right) dV$$

$$= -\int_V \vec{\nabla}\cdot(P\vec{v})\, dV$$

$$= -\int_S (P\vec{v})\cdot\hat{n}\, dS \tag{157}$$

where the last step follows from the divergence theorem. Thus we have

$$\frac{d}{dt}\left[\int_V \frac{1}{2}\left(\rho v^2(\vec{r},t) + \frac{1}{\rho c^2}P^2(\vec{r},t)\right) dV\right]$$

$$= -\int_S P(\vec{r},t)\,\vec{v}(\vec{r},t)\cdot\hat{n}\, dS \tag{158}$$

which is the usual relation between energy density (the volume integrand) and power flux (the surface integrand). Note that for a lossless medium and a uniform acoustic field (i.e. a simple harmonic plane wave) both sides of Eq. (158) are zero. If the acoustic field is not uniform, as for a pulse, then the equation is not trivial.

We now consider the analog of Eq. (158) in the temporal frequency domain. The fields are related by Fourier transforms, as follows

$$P(\vec{r},t) = \int P(\vec{r},\omega)\, e^{i\omega t} d\omega \tag{159}$$

$$\vec{v}(\vec{r},t) = \int \vec{v}(\vec{r},\omega)\, e^{i\omega t} d\omega \tag{160}$$

From Eqs. (153) and (155) we have

$$\rho\frac{\partial\vec{v}}{\partial t} = \int \rho i\omega\vec{v}(\vec{r},\omega)\, e^{i\omega t} d\omega$$

$$= -\int \vec{\nabla}P(\vec{r},\omega)\, e^{i\omega t} d\omega \tag{161}$$

$$\frac{\partial P}{\partial t} = \int i\omega P(\vec{r},\omega)\, e^{i\omega t} d\omega$$

$$= -\rho c^2\int \vec{\nabla}\cdot\vec{v}(\vec{r},\omega)\, e^{i\omega t} d\omega \tag{162}$$

Using inverse transforms we have

$$\rho i\omega\vec{v}\left(\vec{r},\omega\right)=-\vec{\nabla}P\left(\vec{r},\omega\right) \tag{163}$$

$$i\omega P\left(\vec{r},\omega\right)=-\rho c^{2}\vec{\nabla}\cdot\vec{v}\left(\vec{r},\omega\right) \tag{164}$$

We can take derivatives of Eqs. (163) and (164) with respect to ω to obtain

$$\vec{\nabla}\frac{\partial P}{\partial\omega}\left(\vec{r},\omega\right)=-i\rho\vec{v}\left(\vec{r},\omega\right)-i\rho\omega\frac{\partial\vec{v}}{\partial\omega}\left(\vec{r},\omega\right) \tag{165}$$

$$\vec{\nabla}\cdot\frac{\partial\vec{v}}{\partial\omega}\left(\vec{r},\omega\right)=-\frac{i}{\rho c^{2}}P\left(\vec{r},\omega\right)-\frac{i\omega}{\rho c^{2}}\frac{\partial P}{\partial\omega}\left(\vec{r},\omega\right) \tag{166}$$

We now examine the following quantity

$$\vec{\nabla}\cdot\left(P\frac{\partial\vec{v}^{*}}{\partial\omega}+\frac{\partial P^{*}}{\partial\omega}\vec{v}\right)$$

$$=\vec{\nabla}P\frac{\partial\vec{v}^{*}}{\partial\omega}+P\vec{\nabla}\cdot\frac{\partial\vec{v}^{*}}{\partial\omega}+\left(\vec{\nabla}\frac{\partial P^{*}}{\partial\omega}\right)\cdot\vec{v}+\frac{\partial P^{*}}{\partial\omega}\vec{\nabla}\cdot\vec{v}$$

$$=\left[-\rho i\omega\vec{v}\right]\cdot\frac{\partial\vec{v}^{*}}{\partial\omega}+P\left[\frac{i}{\rho c^{2}}P^{*}+\frac{i\omega}{\rho c^{2}}\frac{\partial P^{*}}{\partial\omega}\right]+$$

$$\left[i\rho\vec{v}^{*}+i\rho\omega\frac{\partial\vec{v}^{*}}{\partial\omega}\right]\cdot\vec{v}+\frac{\partial P^{*}}{\partial\omega}\left[-\frac{i\omega}{\rho c^{2}}P\right]$$

$$=i\left[\rho\vec{v}\cdot\vec{v}^{*}+\frac{1}{\rho c^{2}}PP^{*}\right]$$

$$=i\left[\rho\mid\vec{v}\left(\vec{r},\omega\right)\mid^{2}+\frac{1}{\rho c^{2}}\mid P\left(\vec{r},\omega\right)\mid^{2}\right] \tag{167}$$

Integrating over the volume V and using the divergence theorems yields the analog of Eq. (158):

$$\int_{V}\frac{1}{2}\left[\rho\mid\vec{v}\left(\vec{r},\omega\right)\mid^{2}+\frac{1}{\rho c^{2}}\mid P\left(\vec{r},\omega\right)\mid^{2}\right]dV$$

$$=\frac{1}{2i}\int_{S}\left[P\left(\vec{r},\omega\right)\frac{\partial\vec{v}^{*}}{\partial\omega}\left(\vec{r},\omega\right)+\frac{\partial P^{*}}{\partial\omega}\left(\vec{r},\omega\right)\vec{v}\left(\vec{r},\omega\right)\right]\cdot\hat{n}dS \tag{168}$$

Eq. (168) is Foster's theorem [22] for scalar wave fields. For plane wavefronts depending only on x, we have

$$\int_{x_{1}}^{x_{2}}\frac{1}{2}\left[\rho\mid\vec{v}\left(x,\omega\right)\mid^{2}+\frac{1}{\rho c^{2}}\mid P\left(x,\omega\right)\mid^{2}\right]dx$$

$$=\frac{1}{2i}\left[P\left(x,\omega\right)\frac{\partial v^{*}}{\partial\omega}\left(x,\omega\right)+\frac{\partial P^{*}}{\partial\omega}\left(x,\omega\right)v\left(x,\omega\right)\right]_{x_{1}}^{x_{2}} \tag{169}$$

For plane waves traveling in the $+x$ direction, with possible dispersion $k=k\left(\omega\right)$, we write

$$P\left(x,\omega\right)=P_{0}\left(\omega\right)e^{-ik(\omega)x} \tag{170}$$

$$v\left(x,\omega\right)=v_{0}\left(\omega\right)e^{-ik(\omega)x} \tag{171}$$

In this case we have

$$\int_{x_{1}}^{x_{2}}\frac{1}{2}\left[\rho\mid\vec{v}\left(x,\omega\right)\mid^{2}+\frac{1}{\rho c^{2}}\mid P\left(x,\omega\right)\mid^{2}\right]dx$$

$$=\frac{1}{2}\left[\rho\mid v_{0}\left(\omega\right)\mid^{2}+\frac{1}{\rho c^{2}}\mid P_{0}\left(\omega\right)\mid^{2}\right]\left(x_{2}-x_{1}\right) \tag{172}$$

and

$$\frac{1}{2i}\left[P\left(x,\omega\right)\frac{\partial v^{*}}{\partial\omega}\left(x,\omega\right)+\frac{\partial P^{*}}{\partial\omega}\left(x,\omega\right)v\left(x,\omega\right)\right]_{x_{1}}^{x_{2}}$$

$$=\frac{1}{2i}\left[\left(P_{0}\left(\omega\right)v_{0}^{*}\left(\omega\right)+P_{0}^{*}\left(\omega\right)v_{0}\left(\omega\right)\right)i\frac{dk}{d\omega}\left(x_{2}-x_{1}\right)\right]$$

$$=Re\left(P_{0}\left(\omega\right)v_{0}^{*}\left(\omega\right)\right)\frac{dk}{d\omega}\left(x_{2}-x_{1}\right) \tag{173}$$

Thus from Foster's theorem we have

$$v_{g}=\frac{d\omega}{dk}=\frac{Re\left(P_{0}\left(\omega\right)v_{0}^{*}\left(\omega\right)\right)}{\frac{1}{2}\left[\rho\mid v_{0}\left(\omega\right)\mid^{2}+\frac{1}{\rho c^{2}}\mid P_{0}\left(\omega\right)\mid^{2}\right]} \tag{174}$$

This is the rigorous relation for the group velocity. Being proportional to the ratio of power flux to energy density, the group velocity gives the rate at which energy is transported by the wave. Note that the derivation also works for Bloch waves, with $\left(x_{2}-x_{1}\right)=a$.

We can use the dispersion relation in Eq. (78) to calculate the group velocity for Bloch waves, keeping in mind that η is a function of $\omega=cq$. One obtains

$$v_{g}=\frac{\sin\left(ka\right)}{\eta\cos\left(qa\right)+\left(1+\frac{1}{a}\frac{d\eta}{dq}\right)\sin\left(qa\right)} \tag{175}$$

At the middle of the band $ka=\pi/2$, $\sin\left(qa\right)=\left(-1\right)^{n}/\sqrt{1+\eta^{2}}$, and $\cos\left(qa\right)=\left(-1\right)^{n}\eta/\sqrt{1+\eta^{2}}$, where n is the band index. In this case

$$v_{g}=\left(-1\right)^{n}\frac{c}{\sqrt{1+\eta^{2}}}\left[1+\frac{1}{\left(1+\eta^{2}\right)}\frac{1}{a}\frac{d\eta}{dq}\right]^{-1} \tag{176}$$

If η is large, the the group velocity may be approximated with $v_{g}=\left(-1\right)^{n}c/\sqrt{1+\eta^{2}}$.

It is informative to perform numerical simulations and observe pulses propagating through systems of scatterers as a function of time. There are two basic regimes: one when the pulse width is greater than the spacing between

scatterers, and the other when the width of the pulse is less than the spacing of the scatterers. In cases where disorder is not too severe and the remnants of bands may be discerned, then the first case requires superposition only within one band, and the second requires the superposition of states over a number of bands.

In the first case for a periodic system, a rigorous picture may be formulated. Since the width of the pulse is greater than the lattice constant, then one may obtain a valid idea of the behavior of the pulse by evaluating the pulse field at the same point in each unit cell. In this case the periodic part of the Bloch wavefunctions simply becomes a constant, and the wavefunctions take on the simple dispersive plane wave form of Eq. (170) and (171). Now the behavior of the pulse may be determined be examining how the pulse is formed from a window $W(\omega)$ on a transmission spectrum, such as that shown in Fig. 8a. In general, the pulse may be formed with states with different group velocities; as time progresses, the different states will propagate different distances, the synthesis of the pulse field $\psi(x,t)$ will be different, and the pulse will change its shape as it propagates. However, if the window is such that only states near the center of the band have non-negligible amplitudes, then the group velocity may be relatively constant, and the pulse may propagate without changing its shape. This situation is shown in Figs. 8d and e, where the pulse if formed with states in the middle third of the spectrum in Fig. 8a, superimposed with a Hamming window. For $R = .99$, $\eta = 7$, and a group velocity about one-seventh of c is observed.

The behavior of a pulse similar to that just discussed in a system with disorder is shown in Figs. 9d and e. In such a case it is useful to think of the wave medium with scatterers as a system of coupled local oscillators (the "tight binding" model). At the entrance to the system, the pulse transfers some energy to the first local oscillator. Recall that for two coupled simple harmonic oscillotors, energy is transfered between the oscillators at a rate proportinal to the difference in the normal mode frequencies, $\Delta\omega$. For a system with a number of local oscillators, energy is transfered at the group velocity, also proportional to $\Delta\omega \propto d\omega/dk$. For the disordered system of scatterers, only a part of the energy is transfered forward, and the amplitude of the "pulse front", moving at the group velocity, rapidly diminishes. To some degree energy gets preferentially stored at localization sites for the states which comprise the pulse, represented by the peaks in the transmission spectrum (Fig. 9a). In disordered systems, most ot the energy gets reflected at the entrance, as with continuous waves. It is a remarkable property of pulses formed with Bloch wave states that each local oscillator transfers essentially all of its energy to subsequent local oscillators, as in Figs. 8d and e.

The second basic situation, where the width of the pulse is less than the spacing between scatterers, is more difficult to calculate, since states from a number of bands must be included in the superposition forming the incident pulse. Since the pulse width is less than the distance between the scatterers, then one might think it would be possible to understand the behavior of the pulse by keeping track of its reflections and transmissions as it encounters the scatterers. However, it is still difficult to keep proper track of phase and interference. It is interesting to note that there are numerous attempts to do this in the literature, and by an accident, the results have fostered the notion that pulses diffuse through a random sequence of scatterers in one dimension. The problem is that in trying to simplify the numerical calculations, a special kind of disordered medium was used. This is a layered medium, where each layer has a random acoustic impedance, $\mu_j c_j$. However, in order to simplify keeping track of the reflections of the pulse, each layer is given a thickness a_j such that the transit time of the pulse accross each layer, a_j/c_j, is a constant. This type of medium is called a Goupillaud medium [23]. Some results for such a system are shown in Fig. 10. Fig. 10a shows the transmission spectrum for a periodic version which consists of alternating layers of two different media. The wave displacement at he exit of the system as a function of time is shown in Fig. 10b. Note that in the case when the pulse width is smaller than the spacing between scatterers (the interfaces between the layers), even the periodic case has non-trivial behavior. Fig. 10d shows the exit displacement for a disordered Goupillaud system. By comparing Fig. 10b and d, it appears that the pulse has spread, as would be predicted with a diffusion equation. However, this is a complete artifact of the Goupillaud medium; a hint that something non-random is happening here may be seen in the transmission spectrum for the Goupillaud medium, as in Fig. 10c. The bands are disordered, but the disorder appears symmetrically! Unfortunately the false "spreading gaussian" of the Goupillaud medium has supported the notion that pulses diffuse in one-dimensional disordered media.

By numerically calculating a large number of periodic and disordered systems, a general feature concerning the behavior of pulses in one-dimensional disordered systems emerges. In one-dimension the only effect which matters is Anderson localization, and the consequences which this has depends on the nature (frequency dependence) of the scatterers. If the strength of the scatterers increases with frequency, then the higher frequency states will be more localized, and the Anderson localization acts like a low pass filter. If the strength of the scatteres decreases with frequency, then the low frequency states will be more localized, and the Anderson localization acts like a high pass filter. Characteristic examples from a large number of simulations, all showing this effect (with the exception of the Goupillaud medium), are shown in Figs. 11 and 12. In Fig. 11 the scattering strength decreases with frequency. For a periodic system, the result is that for the higher frequencies the bands are wider and the gaps are narrower, as expected. Fig. 11c shows the high pass filtering due to Anderson localization in the disordered system. By comparing the exit displacements of the periodic and disordered systems, Fig. 11 b and c, one sees

that disorder actually makes the pulse appear narrower.

Results with a system where the strength of the scatterer increases with frequency are shown in Fig. 12. The transmission spectrum and pulse behavior for a periodic system are shown in Fig. 12a and b. Results for a disordered system, illustrating low pass filtering by the Anderson localization, are shown in Fig. 12c and d.

VIII. THE QUASICRYSTALLINE POTENTIAL FIELD

One of the original goals of the research with the acoustic simulations was to study two-dimensional Anderson localization. This research goal was diverted with the discovery of quasicrystals. Until a few years ago it was believed that solid matter could exist in two basic forms: crystalline and amorphous (glassy). In an amorphous material the atoms are positioned at random positions, and a macroscopic sample would be homogeneous and isotropic. Crystals are, of course, quite different; crystals are formed by taking a unit cell and periodically repeating it to fill three dimensional space. In order to fill all space with long range periodic order, only unit cells of particular shapes are allowed; only fourteen shapes fit together to fill space, and these form the bases for the fourteen Bravis lattices of crystallography. Since there are only fourteen possible unit cells, only certain rotational symmetries are allowed; in particular, five-fold rotational symmetry is forbidden. However a few years ago Shechtman [24] at the National Bureau of Standards made an aluminum alloy whose x-ray diffraction pattern indicated five-fold rotational symmetry. While it is possible to have atoms in clusters with five-fold (icosahedral) order, the diffraction patterns had sharp spots, indicating long range order. At first the existance of such materials seemed impossible, but theorists pointed out that while it is impossible to have five-fold rotational symmetry and long range periodic order, it is possible to have five-fold rotational symmetry and another type of long range order which is quasiperiodic. Thus atoms may minimize their energy by arranging themselves quasiperiodically, forming quasicrystals.

In one dimension the notion of quasiperiodicity is relatively easy to understand. For a line with periodic lattice sites at a spacing a, the Fourier transform (or diffraction pattern) has a sharp line at π/a. If a line has points positioned randomly, then the Fourier transform would show a broad spectrum. Now suppose one takes a line with periodic lattice spacing b, and superimposes it on the line with periodic spacing a. The result may appear similar to the line with random spacing. However, if the two lattice constants a and b are commensurate, i.e. their ratio a/b is equal to the ratio of two integers, then the superposition will be periodic. The period may be much larger than a or b, but it will be exactly periodic, and the powerful theorems pertaining to periodic systems (Bloch's theorem, group theory, etc.) will apply, and the properties of the system which follow from symmetry may be calculated exactly. On the other hand, if the two lattice constants a and b are not commensurate, i.e. their ratio a/b is equal to an irrational number, then the superposition is not periodic, and there is in general no easy way to calculate the properties. Nevertheless, the superposition does possess long range order; it is constructed from simple rules, and its Fourier transform contains sharp peaks at π/a and π/b. The superposition of periodic lines with incommensurate lattice constants is an example of one-dimensional quasiperiodicity.

While quasiperiodicity in one-dimension is fairly straightfoward, in two and three dimensions the notion is considerably more interesting. For higher dimensions, another method of generating quasiperiodic patterns is used. One begins with a periodic lattice in a higher dimension, e.g. a square lattice in two dimensions. The higher dimensional lattice is traversed with a lower dimensional surface, e.g. a straight line passing through the square lattice, making an angle θ with one of the lattice directions. Next a window width is defined, and lattice points falling within that window are projected onto the lower dimensional surface. If the direction cosines describing the orientation of the lower dimensional surface are irrational, then the sites projected onto the surface will form a quasiperiodic pattern. In the one-dimensional example a particularily interesting case occurs when the slope of the line is such that $\tan\theta = \tau = \left(\sqrt{5}+1\right)/2$, the Golden Mean (also known as the Devine Ratio), which is "the most irrational number." In this case the pattern of sites can be related to a Fibbonacci sequence, and rigorous theorems describing the properties of this special quasiperiodic symmetry can be derived. An algorithm which generates this sequence is

$$a_j = \tau - (\tau - 1)\left[Int\left(\frac{j+1}{\tau}\right) - Int\left(\frac{j}{\tau}\right)\right] \qquad (177)$$

where $Int\,(x)$ indicates the integer part, and $j = 0, 1, 2, 3...$ etc. The quantity a_j will take on the values τ and 1 in a quasiperiodic sequence.

General theories for one-dimensional systems have been reviewed by Simon, [25] and renormalization group and dynamic mapping techniques have been introduced by Kohmoto, Kadonoff, and Tang [26] and by Ostlund, et al. [27] Since the discovery of the aluminum alloy quasicrystals, [24] there has been considerably more work with renormalization group techniques and numerical calculations. [28,29] Some special properties of 1-D quasiperiodic systems are: (1) the eigenvalue spectrum is a Cantor set, (2) there may exist a mobility edge and a metal-insulator transition, (3) the eigenfunctions may be extended, localized, or critical. One-dimensional quasiperiodic systems have received some experimental attention, through measurements of infrared reflectivity in quasiperidic superlattices formed with molecular beam epitaxy. [30]

If one begins with a six-dimensional periodic lattice (e.g. a hypercubic lattice), and intersects it with a three-dimensional surface, then one may obtain a three-dimensional quasicrystalline pattern; such patterns may describe the recently discovered aluminum alloy quasicrystals. If one begins with a five-dimensional periodic lattice and intersects it with a two-dimensional (plane) surface, then one may obtain a two-dimensional quasiperiodic pattern, or Penrose tiling; [31,32] one such pattern is illustrated in Fig. 13. If one tries to tile a plane with only one tile shape (unit cell), then as in three dimensions only certain shapes and rotational symmetries are allowed. However if one is allowed to use tiles of two or more shapes then five-fold and other symmetries are possible. Articles and cover stories on Penrose tilings have appeared in Scientific American, [32] American Scientist, [33] etc.

Although Penrose tile patterns are not periodic, they do have some symmetry properties. [30,32] Special patterns may have "inflation symmetry", so that a decoration of the tiles with certain lines will produce a replication of the original Penrose pattern, but with a reduced scale. There is also Conway's theorem which states that given any local pattern (having some nominal diameter), an identical local pattern will be found within a distance of two diameters. The typical Penrose tiling shown in Fig. 13 is formed with two different rhombuses, a fat one and a skinny one; the ratio of the areas is the Golden Mean, $\left(\sqrt{5}+1\right)/2$.

Quasiperiodic patterns are a new symmetry of Nature, and it will be important to understand how this new symmetry effects the properties of materials. A fundamental question is: given a wave equation (for Schrodinger waves or sound waves) with a potential field having quasicrystalline symmetry, what are the eigenvalues and eigenfunctions. Unlike the theorems for the quasiperiodic patterns in one dimension, theorems for two and three dimensions, if they exist, have not yet been discovered. For a pattern which is periodic Bloch's theorem may be used to obtain exact solutions. For a pattern which is random statistical methods may be used to make predictions about properties (e.g. Anderson localization). However a quasiperiodic pattern is not periodic, so that Bloch's cannot be used; and it is not random, so that statistical methods cannot be used.

In the absence of a "quasi-Bloch theorem", progress in understanding the consequences of 2-D quasicrystalline symmetry has relied mostly on numerical calculations. [34–36] However these calculations have found no features which reflect the quasiperiodic symmetry.

Despite the notable efforts in searching for unique consequences of 2-D quasiperiodic symmetry, existing studies do not treat the problem which we wish to address, namely, the properties of a wave equation with a quasiperiodic potential. All of the current theoretical research has dealt with a hopping Hamiltonian involving a matrix embodying the quasiperiodic topology, but having all non-zero matrix elements identical. For an actual wave equation the problem can be reduced to a similar matrix, but the non-zero matrix elements would be complex functions of the eigenvalue, whose determination would involve solving a complicated transcendental equation (an example is presented below). The resulting eigenvalue spectrum would be quite different from the ones found with the existing theoretical models, because now one must contend with the possibility of phase coherence effects in a system of scatters with a quasiperiodic pattern. Another way of viewing the situation is to note that another length, the wavelength, has entered the problem. The relation between the wavelength and the inflation and pattern repetition properties of the quasiperiodic pattern results in new features in the density of the states.

A convenient description of an acoustic 2-D quasiperiodic system as an analog of the Schrodinger equation may be obtained with a "tight-binding" approach, where local oscillators, which when isolated have a single sharp eigenvalue or resonant frequency, are coupled together to produce a spectrum of eigenvalues. While this seems identical to the models used in the numerical studies, there is a crucial difference: in the numerical studies the coupling has been analogous to "massless springs", whereas in the acoustic system the coupling is through a wave medium (a spring with a finite mass) involving a wavelength, phase coherence, etc.

As an example, consider a system where each local oscillator is a mass m on a massless spring with resonant frequency ω_0, and the coupling to nearest neighbors (NN) is through waves in other springs with sound speed c, individual masses m_{NN}, and lengths a_{NN}. Following the notation of Ref. [36], where $\psi(x)$ is the amplitude at a site x and $\psi(x')$ is that of a nearest neighbor, the equations of motion are

$$E\psi(x) = -\sum_{NN}\left[\alpha\psi(x') - \beta\psi(x)\right] \tag{178}$$

where in the acoustic case the eigenvalue E is equal to the square of the acoustic resonance frequency, $E=\omega^2$, and

$$\alpha = \omega_{NN}^2\left(\frac{\omega a_{NN}}{c}\csc\frac{\omega a_{NN}}{c}\right) \tag{179}$$

$$\beta = \frac{\omega_o^2}{z} + \omega_{NN}^2\left(\frac{\omega a_{NN}}{c}\cot\frac{\omega a_{NN}}{c}\right) \tag{180}$$

where $\omega_{NN} = (c/l_{NN})\sqrt{m_{NN}/m}$ and z is the number of nearest neighbors. These equations illustrate the comments made in the earlier paragraph: if the coupling springs were massless, the speed of sound, proportional to $1/\sqrt{m_{NN}}$, would be infinite (ω_{NN} would remain finite), the factors in parentheses in Eqs. (179) and (180) would be unity, and the matrix elements α and β would become constants, no longer functions of the eigenvalue.

This is the model used in existing numerical calculations. However, in order to have a wave equation, c must be finite, and determining the eigenfrequencies ω with α and β functions of ω involves a complicated transcendental equation.

An experiment which provided experimental results for the consequences for quasicrystalline symmetry in a two-dimension wave medium is presented in Ref. [37].

APPENDIX A: LAYERED MEDIA

Here we consider a system composed of layers, each with a characteristic impedance $Z_j = \mu_j c_j$. The coordinate and indexing system is as in section 4, which described sections of length $a_j = x_j - x_{j-1}$ between point scatterers. We now insert in each section a_j an intermediate layer of thickness b_j extending from $x_j - b_j$ to x_j. We also allow each layer to have its own wave speed c_j, so that each layer has its own wavevector $q_j = \omega/c_j$. Now the wave fields between x_{j-1} and $x_j - b_j$, $x_j - b_j$ and x_j, and x_j and $x_{j+1} - b_{j+1}$ are given by

$$u_{j-1}e^{iq_j(x-x_{j-1})} + v_{j-1}e^{-iq_j(x-x_{j-1})} \tag{A1}$$

$$u'_j e^{iq'_j(x-x_j)} + v'_j e^{-iq'_j(x-x_j)} \tag{A2}$$

$$u_j e^{iq_{j+1}(x-x_j)} + v_j e^{-iq_{j+1}(x-x_j)} \tag{A3}$$

respectively. The characteristic impedances for the layers are Z_j, Z'_j, and Z_{j+1}, respectively. Continuity of the wavefield and force at the interface at $x_j - b_j$ gives

$$u_{j-1}e^{iq_j(a_j-b_j)} + v_{j-1}e^{-iq_j(a_j-b_j)}$$

$$= u'_j e^{-iq'_j b_j} + v'_j e^{iq'_j b_j} \tag{A4}$$

$$Z_j\left[u_{j-1}e^{iq_j(a_j-b_j)} - v_{j-1}e^{-iq_j(a_j-b_j)}\right]$$

$$= Z'_j\left[u'_j e^{-iq'_j b_j} - v'_j e^{iq'_j b_j}\right] \tag{A5}$$

Continuity of the wavefield and force at the interface at x_j gives

$$u'_j + v'_j = u_j + v_j \tag{A6}$$

$$Z'_j\left[u'_j - v'_j\right] = Z_{j+1}\left[u_j - v_j\right] \tag{A7}$$

Eliminating u'_j and v'_j from Eqs. X through Z results in a matrix equation of the form of Eq. (57), with

$$\alpha_j = \frac{1}{2}\left[\left[1 + \left(\frac{Z_{j+1}}{Z_j}\right)\right]\cos\left(q'_j b_j\right) -\right.$$

$$\left. i\left[\left(\frac{Z'_j}{Z_j}\right) + \left(\frac{Z_{j+1}}{Z_j}\right)\left(\frac{Z_j}{Z'_j}\right)\right]\sin\left(q'_j b_j\right)\right]e^{-iq_j(a_j-b_j)} \tag{A8}$$

$$\beta_j = \frac{1}{2}\left[\left[1 - \left(\frac{Z_{j+1}}{Z_j}\right)\right]\cos\left(q'_j b_j\right) -\right.$$

$$\left. i\left[\left(\frac{Z'_j}{Z_j}\right) - \left(\frac{Z_{j+1}}{Z_j}\right)\left(\frac{Z_j}{Z'_j}\right)\right]\sin\left(q'_j b_j\right)\right]e^{-iq_j(a_j-b_j)} \tag{A9}$$

Note that the determinant of the matrix is

$$|\,\alpha_j\,|^2 - |\,\beta_j\,|^2 = \left(\frac{Z_{j+1}}{Z_j}\right) \tag{A10}$$

The equations simplify if we have a uniform medium with non-uniform slabs inserted; i.e. $Z_j = Z_{j+1} = Z \neq Z'_j$, and $q_j = q_{j+1} = q \neq q'_j$. Furthermore, for thin slabs we can take the limit that b_j goes to zero, but $b_j Z'_j$ remains finite. Then we get α_j and β_j as in Eqs. (58) and (59), but with $\eta_j = q'_j b_j Z'_j/2Z$.

[1] G. Floquet, Ann. Sci. Ecole Norm. Sup. **12**, 47 (1883)

[2] F. Bloch, Z. Physik **52**, 555 (1928)

[3] P. W. Anderson, Phys. Rev. **109**, 1492-1505 (1958), "Absence of diffusion in certain random lattices"

[4] N. F. Mott, Adv. Phys. **16**, 49-144 (1967), "Electrons in Disordered Structures"

[5] Shanjin He and J. D. Maynard, Phys. Rev. Lett. **57**, 3171 (1986), "Direct Measurements of Inelastic Effects in Anderson Localization"

[6] Shanjin He and J. D. Maynard, Phys. Rev. Lett. **62**, 1888 (1989), "Eigenvalue spectrum, density of states, and eigenfunctions of a two-dimensional quasicrystal"

[7] E. Merzbacher, *Quantum Mechanics* (John Wiley, New York, 1961) pp. 51-64.

[8] A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1966) pp. 432-448.

[9] C. Kittel, *Quantum Theory of Solids* (John Wiley, New York, 1963) pp 13-23.

[10] P. W. Anderson, Rev. Mod. Phys. **50**, 1978, "Nobel Lectures in Physics for 1977"

[11] H. Furstenberg, Trans. Amer. Math. Soc. **108**, 337 (1963), "Noncommuting random products"

[12] M. Luban and J. H. Luscombe, Phys. Rev. **B35**, 1987, "Localized eigenstates of one dimensional tight-binding systems: A new algorithm"

[13] R. Fisch, Phys. Rev. **B34**, 1662 (1986), "Behavior of wavefunctions in smooth aperiodic potentials: Anderson localization in the continuum"

[14] P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 1985, "Disordered Electronic Systems"

[15] V. I. Perel and D. G. Polyakov, Soviet Physics JETP, **59**, 1984, "Probability distribution for the transmission of an electron through a chain of randomly placed centers"

[16] W. M. Visscher and J. E. Gubernatis, in *Dynamical Properties of Solids,* **4,** ed. G.K. Horton and A. A. Maradudin, North Holland, Amsterdam, 1980, "Computer experiments and disordered solids"

[17] M. Schreiber, in *Localization 1990,* ed. K. A. Benedict and J. T. Chalker, Institute of Physics, New York, 1991, "Numerical characterization of electronic states in disordered systems"

[18] P. Oetking, J. Geophys. Res. **71**, 2505 (1966)

[19] B. L. Al'tshuler and P. A. Lee, Physics Today, **41**, December, 1988, "Disordered Electronic Systems"

[20] T. C. Spencer, in *Critical Phenomena, Random Systems, Gauge Theories,* ed. K. Osterwalder and R. Stora, Elsevier, New York, 1986, "The Schrodinger equation with a random potential"

[21] Shanjin He and J. D. Maynard, Phys. Rev. Lett. **57**, 3171 (1986), "Direct Measurements of Inelastic Effects in Anderson Localization"

[22] R. E. Collin, *Foundations for Microwave Engineering* , (McGraw-Hill, New York, 1992); The application to acoustic waves is due to C. Bradley.

[23] R. Burridge, G. S. Papanicolaou, and B. S. White, Wave Motion **10**, 19 (1988), "One-dimensional wave propagation in a highly discontinuous medium"

[24] D. Schechtman et al., Phys. Rev. Lett. **53**, 1951 (1984), "Metallic phase with long range orientational order and no translational symmetry"

[25] B. Simon, Adv. Appl. Math. **3,** 463-490 (1982), "Almost Periodic Schrodinger Operators: A Review",

[26] M. Kohmoto, L. P. Kadanoff, and C. Tang, Phys. Rev. Lett. **50**, 1870-1872 (1983), "Localization Problem in One dimension: Mapping and Escape"

[27] S. Ostlund et al, Phys. Rev. Lett. **50**, 1873-1876 (1983), "One dimensional Schrodinger Equation with an Almost Periodic Potential" and S. Ostlund and R. Pandit, Phys. Rev. **B29**, 1394 (1984), "Renormalization group analysis of the discrete quasiperiodic Schrodinger equation"

[28] J. M. Luck and D. Petritis, J. Stat. Phys. **42**, 289-310 (1986), "Phonon Spectra in One-dimensional Quasicrystals"; M. Kohmoto and J. R. Banavar, Phys. Rev. **B34**, 563-566 (1986), "Quasiperiodic Lattice: Electronic Properies, Phonon Properties and Diffusion"; Q. Niu and F. Nori, Phys. Rev. Lett. **57**, 2057-2060 (1986), "Renormalization group Study of One dimensional Quasiperiodic Systems"; M. Kohmoto, B. Sutherland, and C. Tang, Phys. Rev. **B35**, 1020-1033 (1987), "Critical wave functions and a Cantor-Set Spectrum of a one-dimensional quasiperiodic model"; G. Gumbs and M. K. Ali, Phys. Rev. Lett. **60**, 1081-1084 (1988), "Dynamical Maps, Cantor Spectra and Localization for Fibonacci and Related Quasiperiodic Lattices"

[29] M. Luban, J. H. Luscombe, and S. Kim, Phys. Rev. Lett. **60**, 2689-2692 (1988), "Incipient Infinite Degeneracy and the Delocalization Transition in a Quasiperiodic Potential"

[30] R. Merlin, K. Bajema, R. Clarke. F.-Y. Juang, and P. K. Bhattacharya, Phys. Rev. Lett. **55**, 1768-1770 (1985), "QuasiPeriodic GaAs-AlAs Heterostructures"

[31] R. Penrose, Bull. Inst. Math. Appl. **10**, 266 (1974).

[32] M. Gardner, Sci. Am. **236**, 110-115 (1977). "Mathematical games: Extraordinary nonperiodic tiling that enriches the theory of tiles"

[33] P. J. Steinhardt, American Scientist **74**, 586-597 (1986), "Quasicrystals"

[34] T. C. Choy, Phys. Rev. Lett. **55**, 2915 (1985), "Density of states for a two-dimensional Penrose lattice: Evidence of a strong Van Hove singularity"; T. Odagaki and D. Nguyen, Phys. Rev. B**33**, 2184 (1986), "Electronic and vibrational spectra of two-dimensional quasicrystals"

[35] M. Marcus, Phys. Rev. B**34**, 5981 (1986), "Comparison of properties of quasiperiodic and periodic lattices in two and three dimensions"

[36] M. Kohmoto and B. Sutherland, Phys. Rev. B**34**, 3849 (1986), "Electronic and vibrational modes on a Penrose lattice: Localized states and band structure"; M. Kohmoto and B. Sutherland, Phys. Rev. Lett. **56**, 2740 (1986), "Electronic states on a Penrose lattice"; H. Tsunetsugu, T. Fujiwara, K. Ueda, and T. Tokihiro, J. Phys. Soc. Jpn. **55**, 1420 (1986), "Eigenstates in two-dimensional Penrose tiling"

[37] Shanjin He and J. D. Maynard, Phys. Rev. Lett. **62**, 1888 (1989), "Eigenvalue spectrum, density of states, and eigenfunctions of a two-dimensional quasicrystal"

FIG. 1. Dispersion relation for a one-dimensional system of harmonically coupled particles.

FIG. 2. Graphical representation of the transcendental equation giving the relationship between the frequency parameter $q = \omega/c$, and the Bloch wave parameter k.

FIG. 3. Actual plot of the frequency $\omega = cq$ versus Block wave parameter k, which is the conventional band structure found in solid state physics. The dashed line indicates the dispersion for the wave medium without the scatterers, $\omega = ck$.

FIG. 4. Eigenfunctions corresponding to various points in the band structure. Letters correspond to the letters labeling points in Fig. 2.

FIG. 5. Illustration of a typical bloch wave function. a) The actual wavefunction. b) The modulus of the periodic part. c) A trig function of the phase of the wave function, which is clearly not a simple sinusoid.

FIG. 6. Real part of the driving point impedance as a function of frequency for a periodic mass-loaded string, driven at one end and clamped at the other. The Q of the system is 1500.

FIG. 7. Frequency spectra of power delivered to a periodic system of 32 scatterers with reflection coefficients of .99. a) Power delivered by a velocity drive to the system with no damping, but with the system terminated with an impedance μc. The first three peaks are clipped in the figure; the first peak is actually 50 times higher than the fourth. b) Power delivered by a velocity drive to the system with damping (with a Q of 2000) and fixed at the end. c) Same as b, but with a force drive. Note that the peaks do not occur at the same frequencies as in b.

FIG. 8. Numerical calculations for a periodic system with 64 scatterers with reflection coefficients of .99. a) Transmission spectrum showing dispersion near band edges. b) Typical wavefunction for a state with qa near the band edge. c) Typical wavefunction for a state with qa near the band center. d) A pulse formed with states in the middle third of the band, at an early time. e) Same as d at a later time.

FIG. 9. Same as Fig. 8, but with 2% disorder in the spacing between the scatterers. a) Transmission spectrum, showing significant reduction in total transmission. b) and c) Anderson localized wavefunctions. d) A pulse at an early time. e) A pulse at a later time; the early local oscillators do not give up all of their energy as for the periodic system in Fig. 8.

FIG. 10. Transmission spectra and narrow pulse behavior in a Goupillaud medium. a) Transmission spectrum for a periodic sequence of layers. b) Pulse behavior for a. c) Transmission spectrum for a disordered sequence of layers, but with equal transit times. Note the symmetry in the spectrum. d) Pulse behavior for c. That the response looks like a spreading Gaussian is an accidental artifact of the Goupillaud medium.

FIG. 11. Transmission spectrum and pulse behavior for a system of scatterers whose strength decreases with frequency. a) Transmission spectrum for a periodic system. b) Pulse behavior for a periodic system. c) Transmission spectra for a disordered system, showing the high pass filtering due to Anderson localization. d) Pulse behavior for the disordered system; with high pass filtering, the pulse appears narrower.

FIG. 12. Transmission spectrum and pulse behavior for a system of scatterers whose strength increases with frequency. a) Transmission spectrum for a periodic system. b) Pulse behavior for a periodic system. c) Transmission spectra for a disordered system, showing the low pass filtering due to Anderson localization. d) Pulse behavior for the disordered system, with obvious consequences of the low pass filtering.

FIG. 13. Typical Penrose tiling pattern. Two rhombuses, whose areas are in the ratio of the Golden Mean, are used to tile the plane with five-fold rotational symmetry.

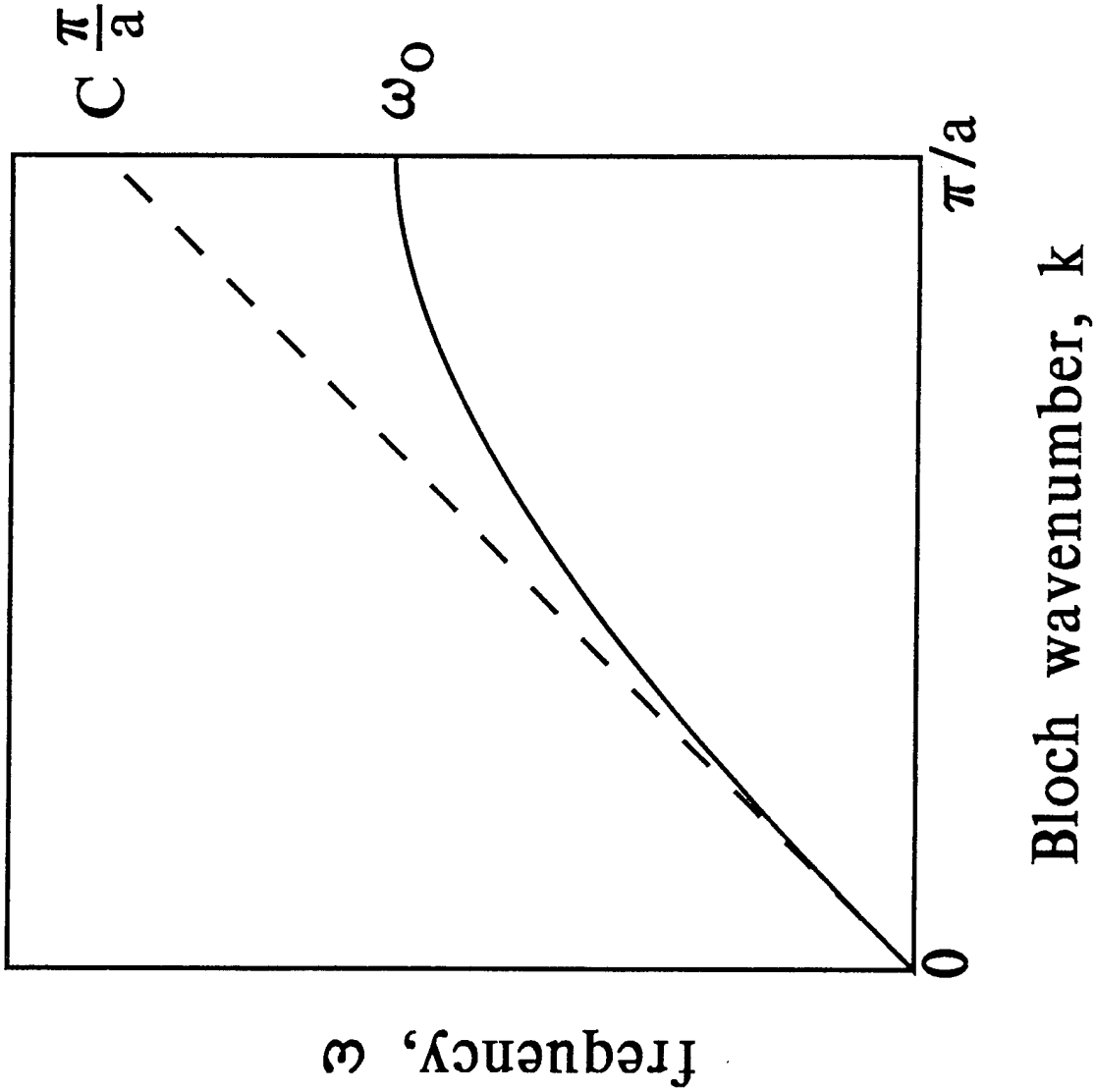


Fig. 1

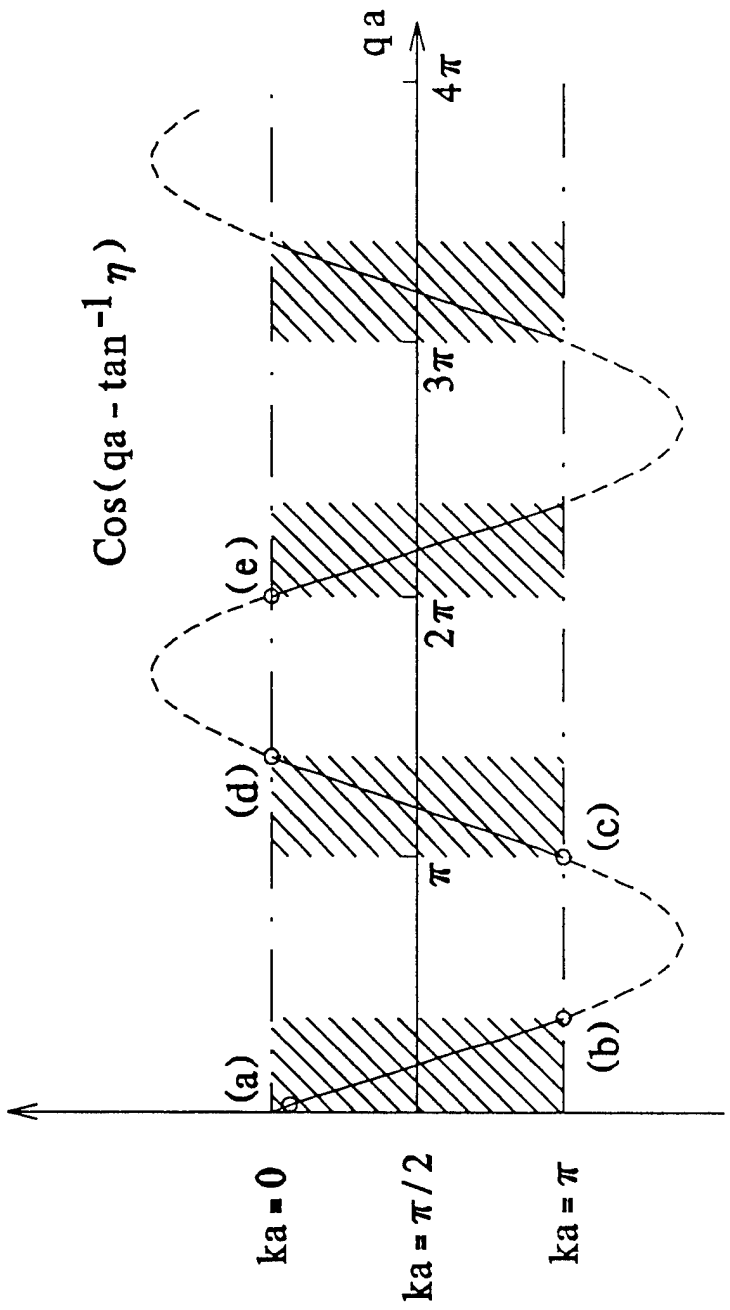


Fig. 2

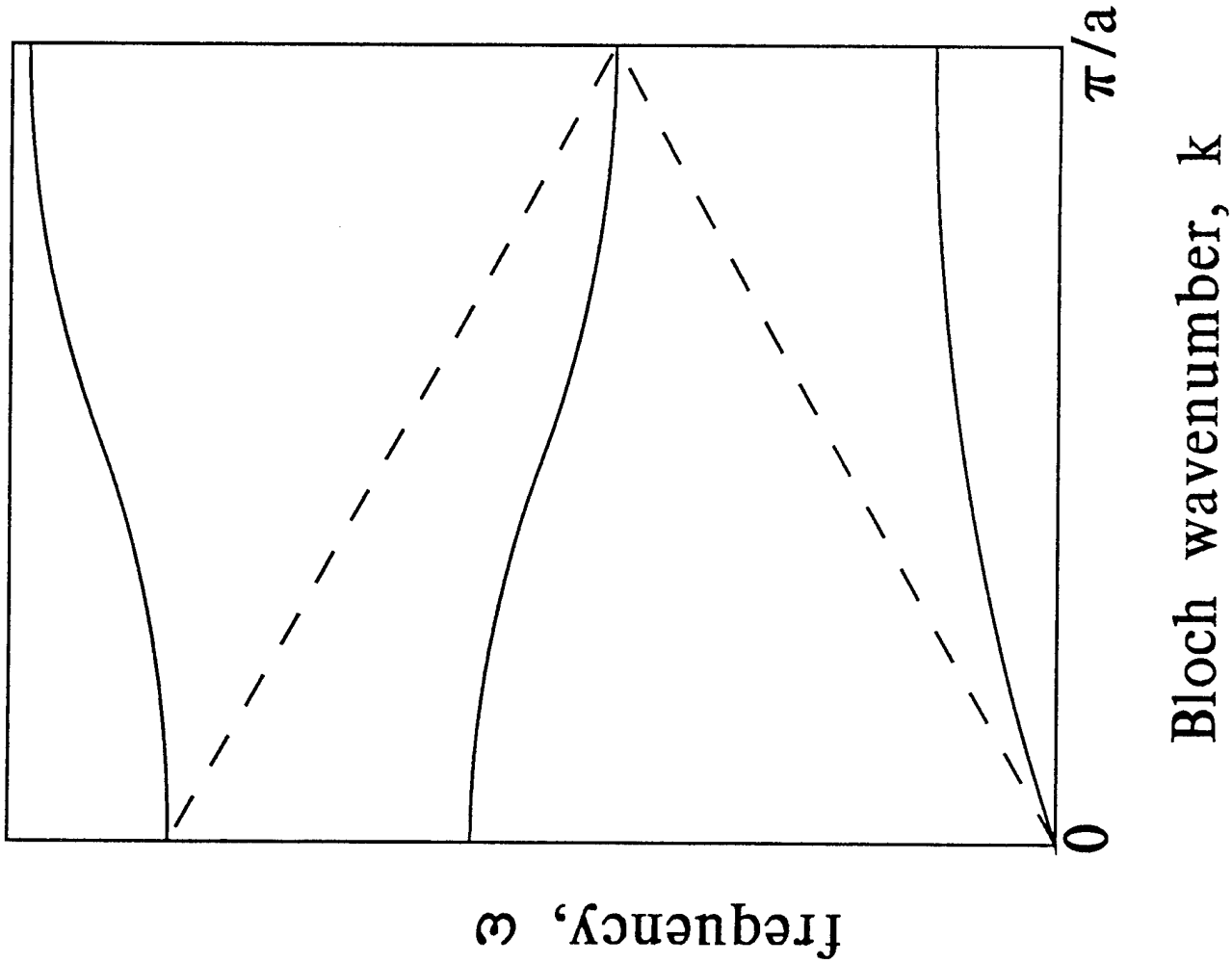


Fig. 3

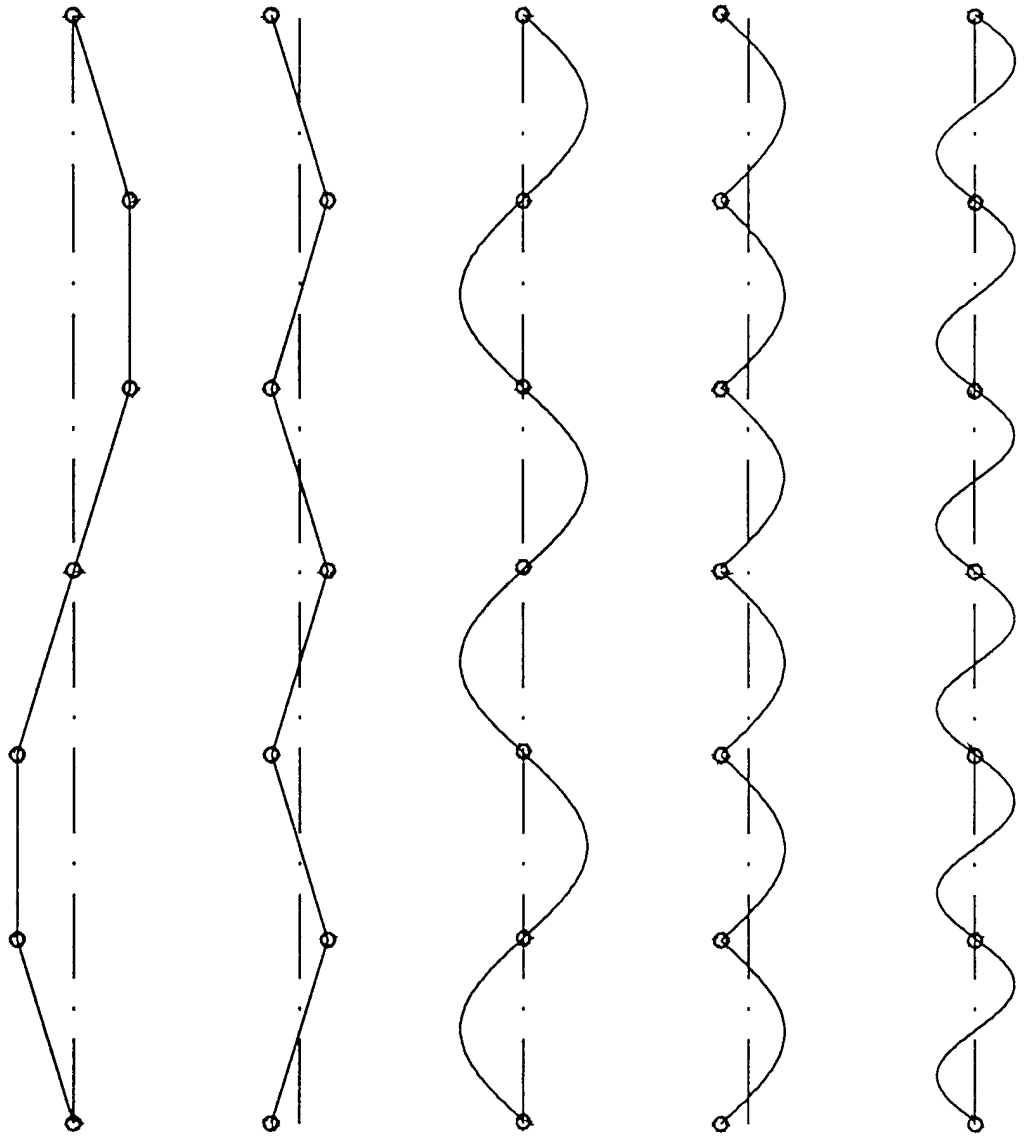

(a)
ka > 0
(b)
ka = π
low freq.
(c)
ka = π
high freq.
unshifted
(d)
ka = 0
high freq.
(e)
ka = 0

Fig. 4

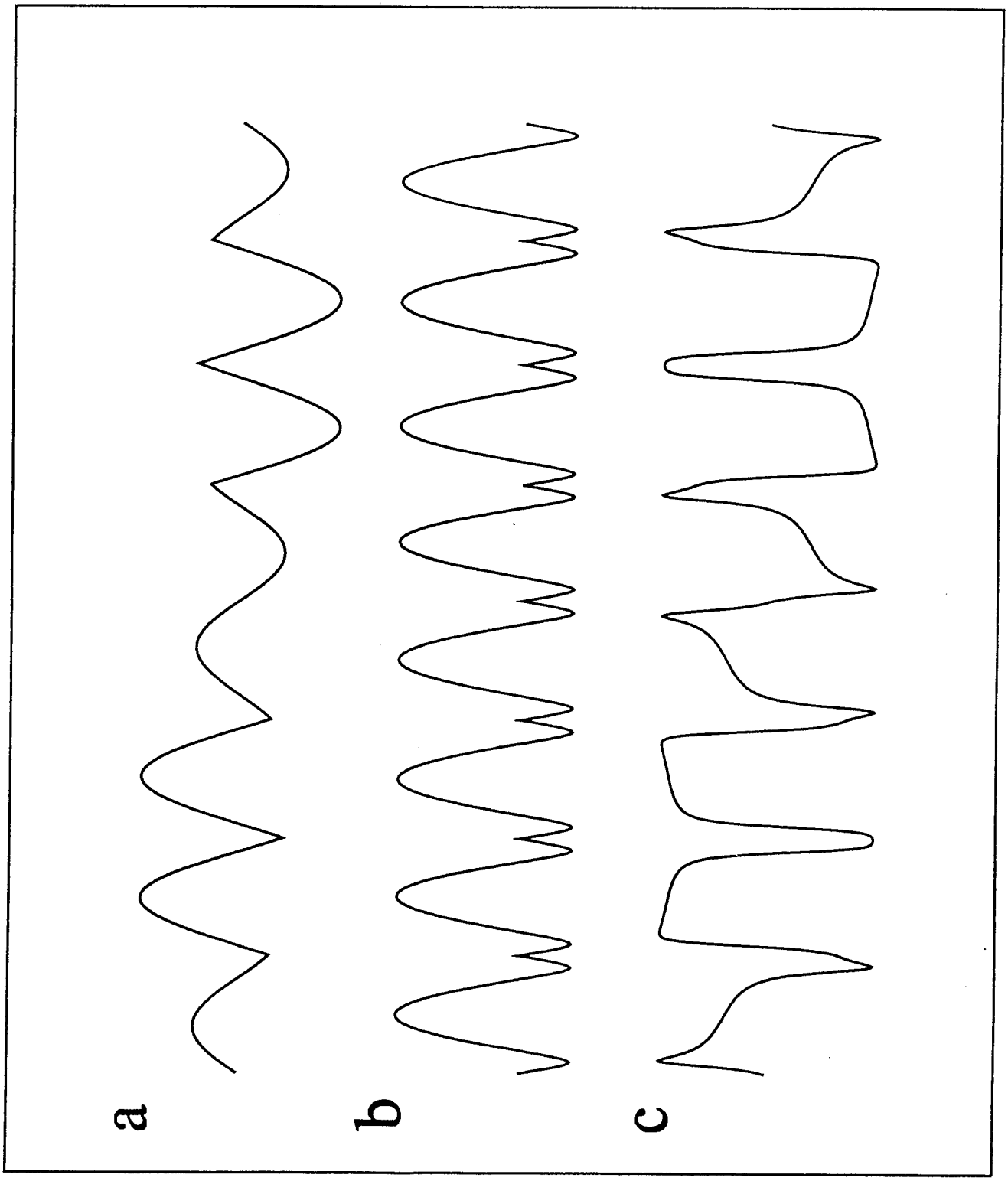


Fig. 5

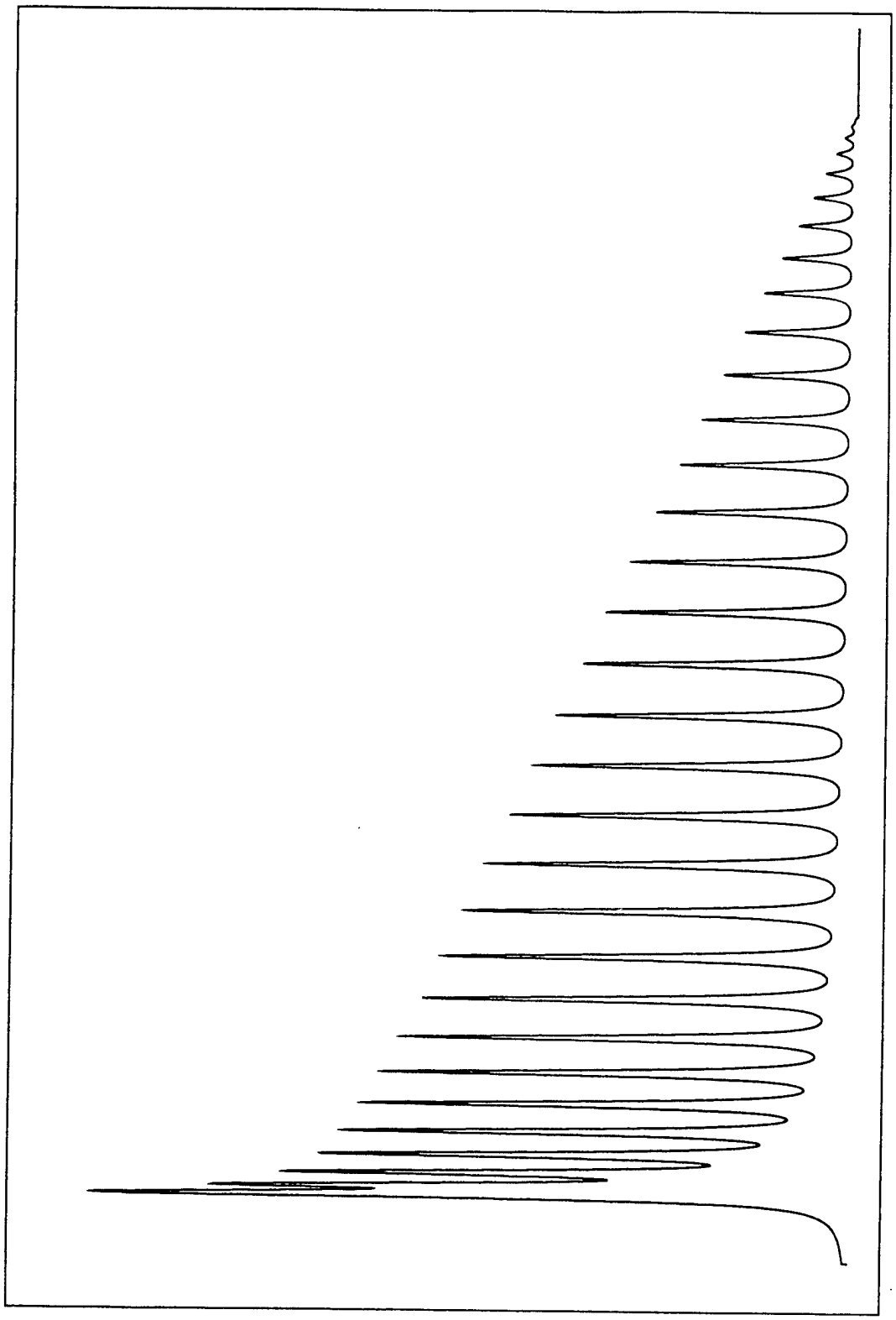


Fig. 6

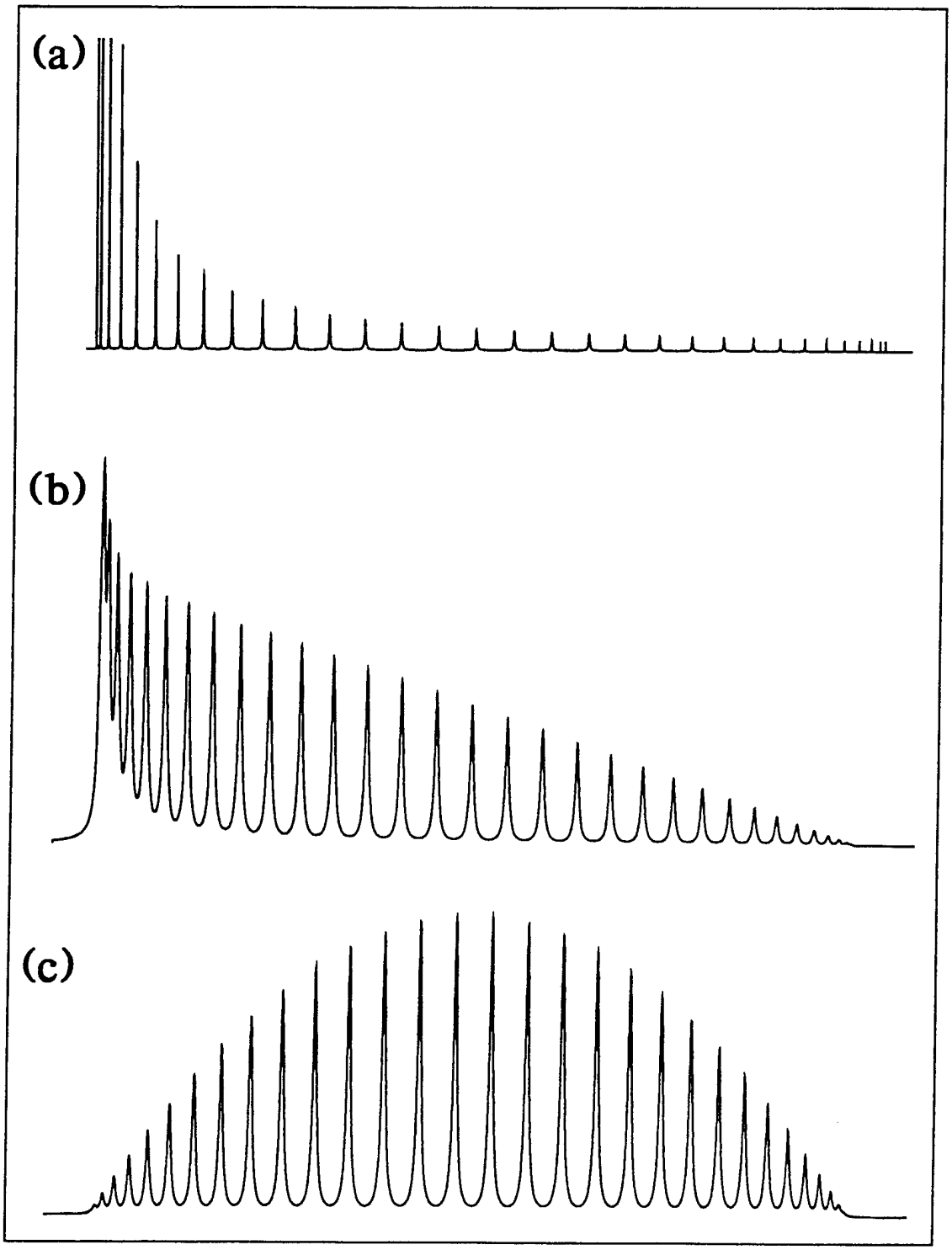


Fig. 7

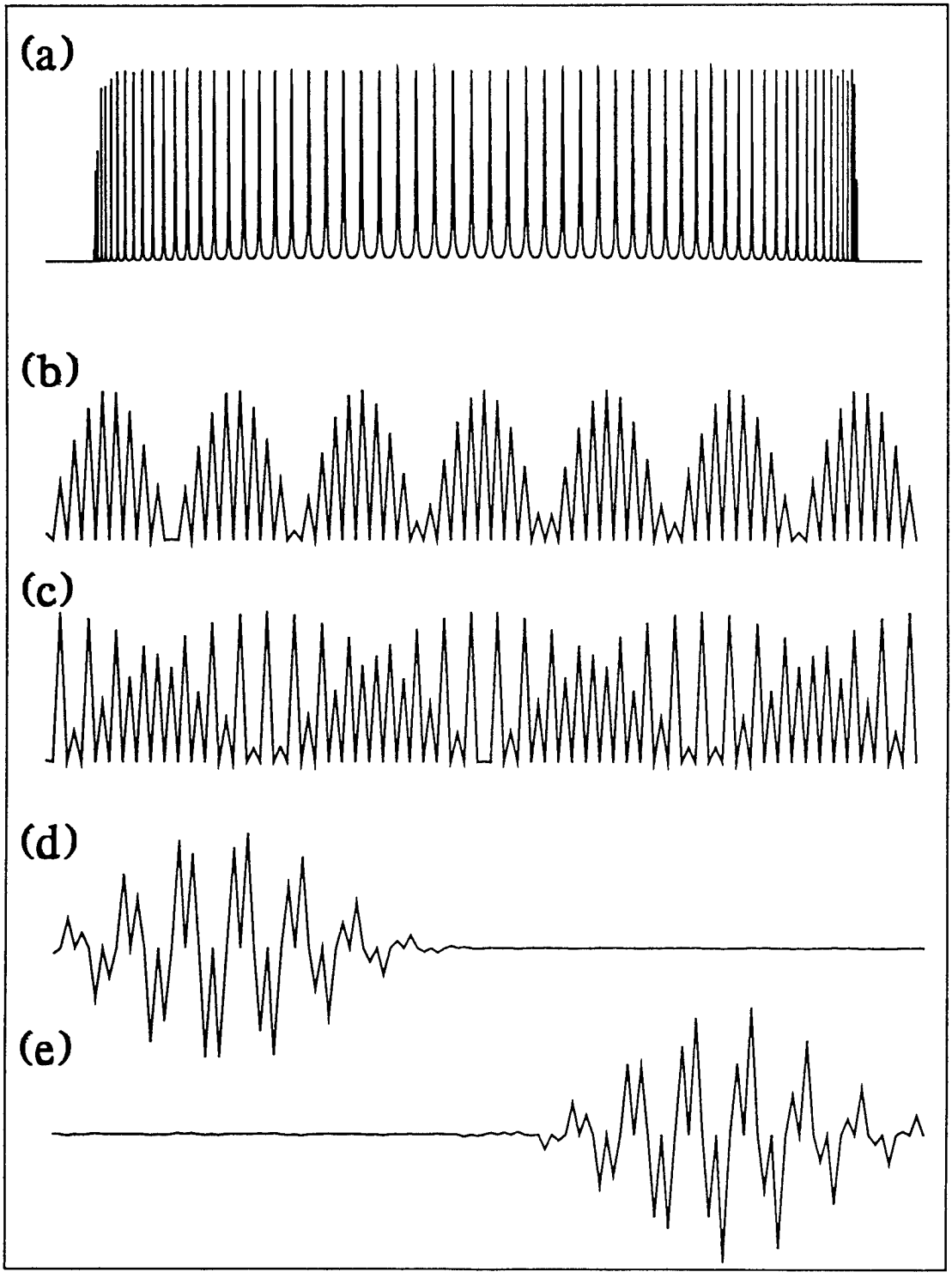


Fig. 8

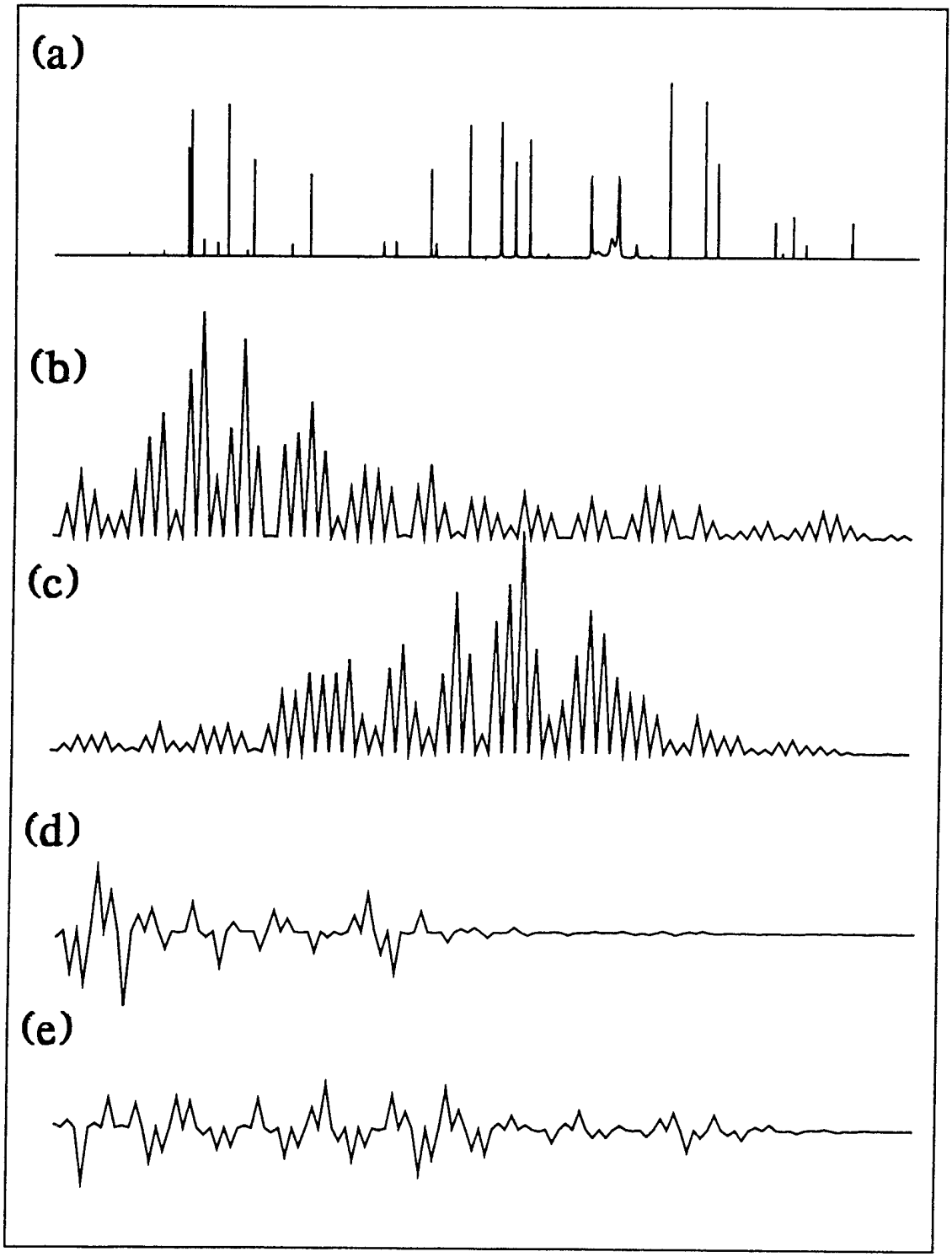


Fig. 9

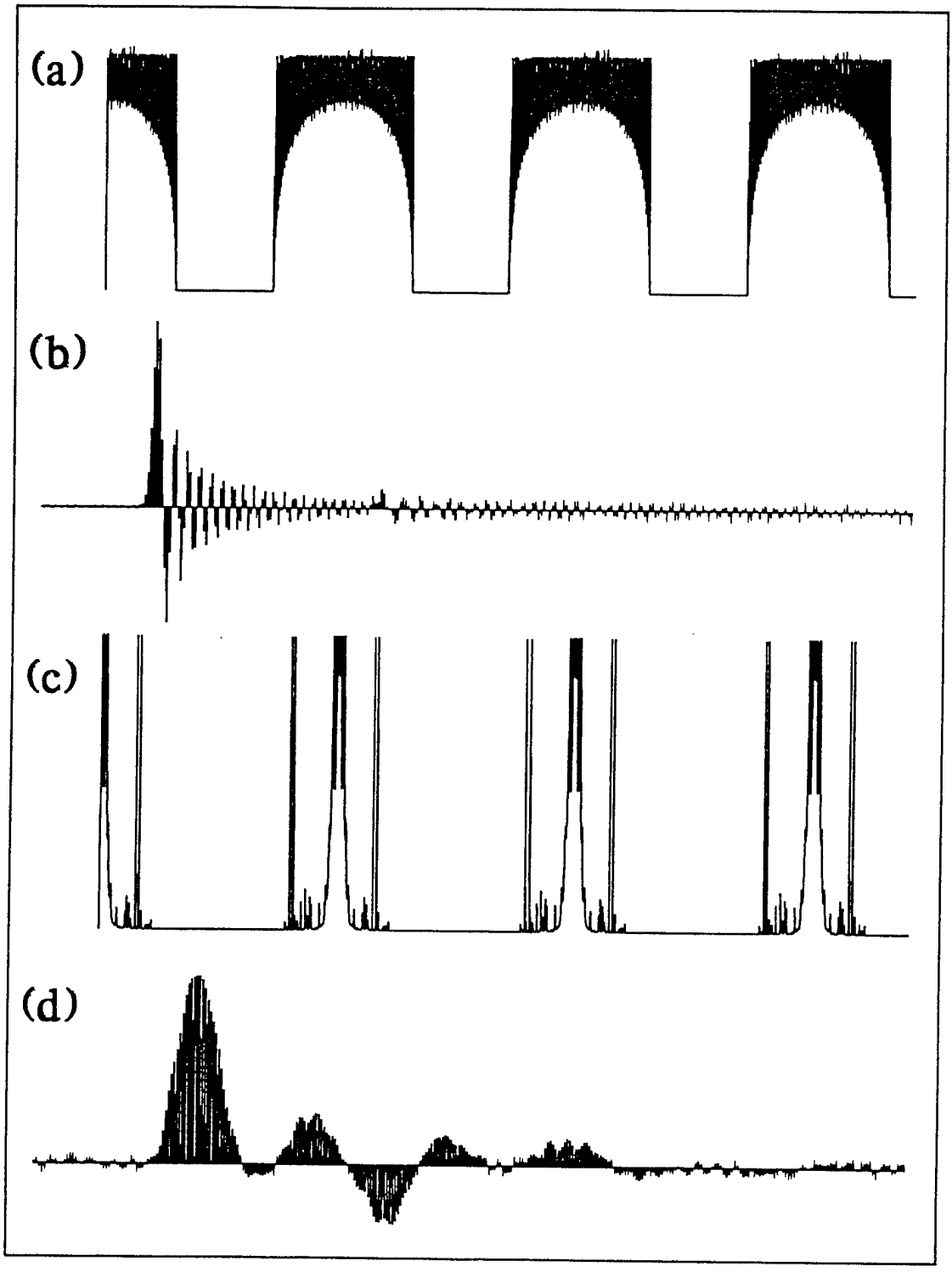


Fig. 10

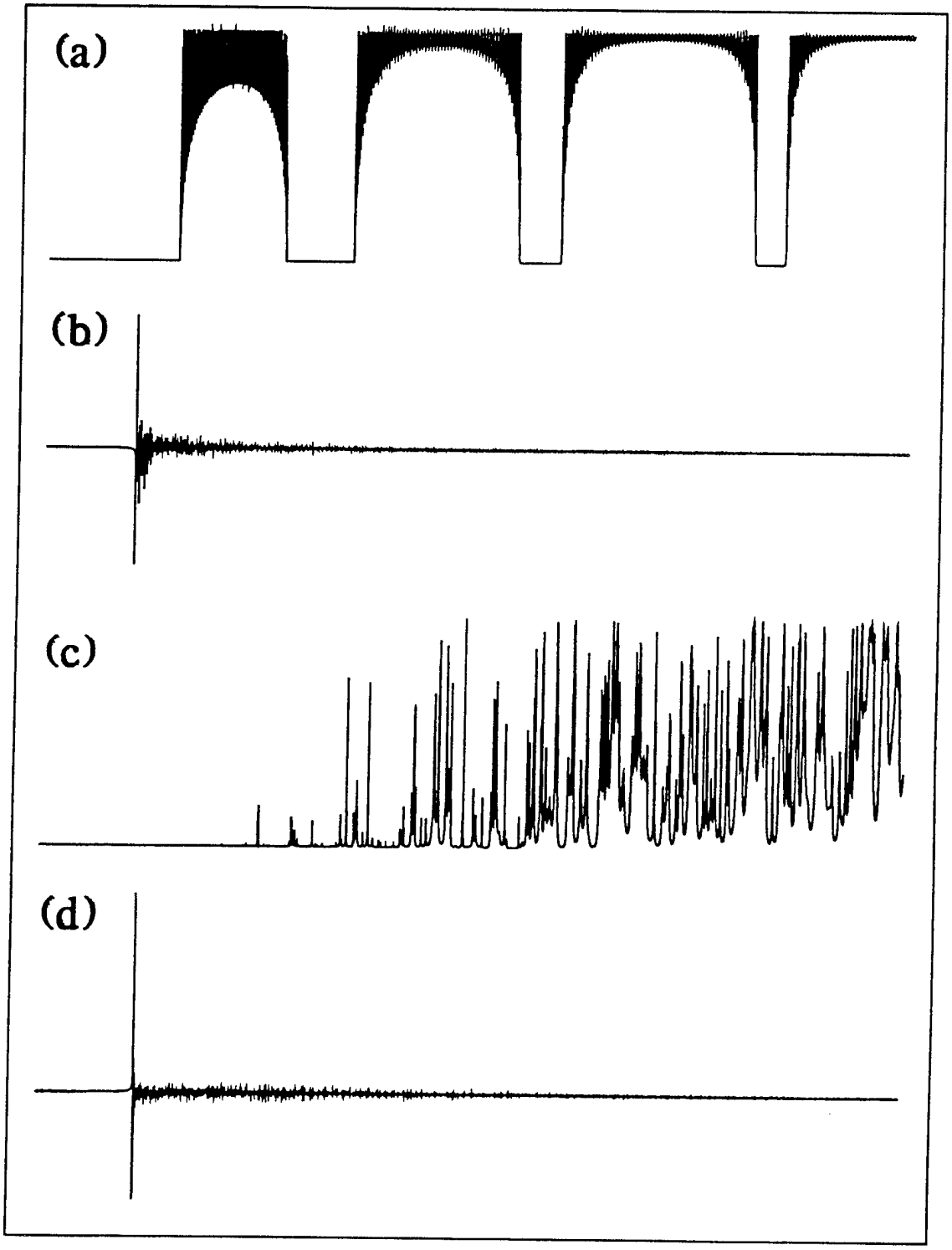


Fig. 11

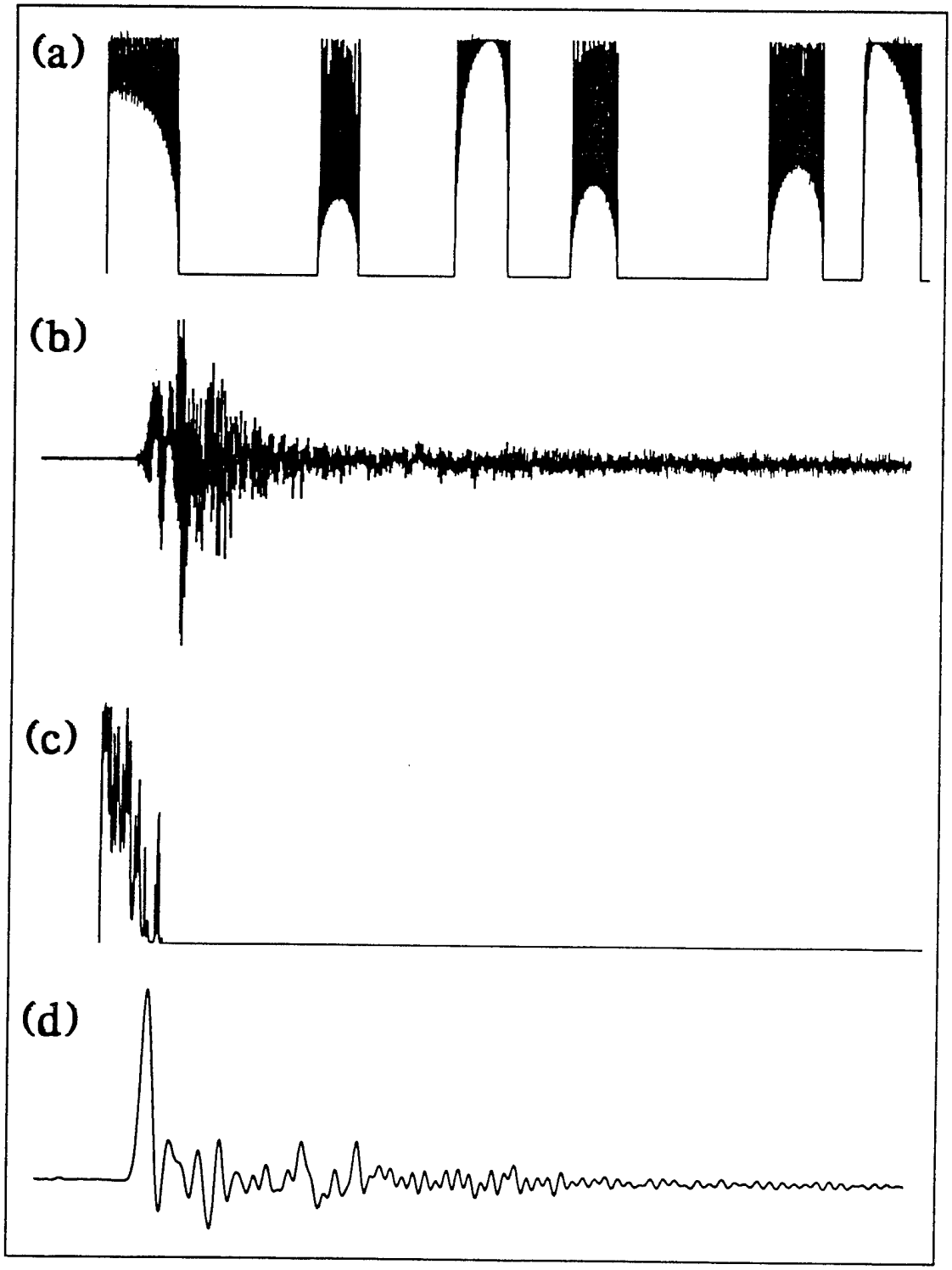


Fig. 12

SINGLE BUBBLE SONOLUMINESCENCE

ANTHONY A. ATCHLEY
Physics Department
Naval Postgraduate School
Monterey, CA

A single gas bubble, acoustically levitated in a standing wave field and oscillating under the action of that field, can emit a single pulse of blue-white light. This emission, called sonoluminescence (SL), has a number of remarkable features. It is easily visible to the unaided eye. The duration of the SL pulse is probably less than 50 ps. The emission can be extremely stable and repetitive, lasting for hours, or it can be chaotic. The spectrum is broadband and increases in intensity in the ultraviolet. Slight cooling of the host liquid significantly increases the intensity. The intensity of the luminescence is sensitive to changes in the composition of gases within the bubble of less than 1%. Single SL pulses are isotropic and unpolarized. Several mechanisms, ranging from classical to quantum mechanical, have been proposed to account for this phenomenon. Yet, there is no completely satisfactory explanation. The purpose of this lecture is to provide a review of this field.

Because single bubble sonoluminescence is intimately tied to the response of a bubble driven by an acoustic field, the lecture begins with a discussion of bubble dynamics. Topics include acoustic levitation and the Rayleigh-Plesset (R-P) equation. The R-P equation is a simple differential equation that describes the motion of the bubble wall under ordinary circumstances. Results of calculations of the pressures and temperatures within the bubble using the R-P equation are used as an oversimplified lead-in to a review of the astonishing properties of SL that have been determined experimentally. Comparisons to a related phenomenon, multiple-bubble sonoluminescence, will be made. Next attention is turned to theoretical explanations of this phenomena. Limitations of the R-P equation are discussed and why more sophisticated methods of prediction the dynamics of the bubble are needed. Other topics include evolution of shock waves inside the bubble, and radiation mechanisms such as black-body, bremsstrahlung, collision-induced emission, and zero-point fluctuations.

Almost from the outset, people have conjectured about the limits of SL. One extreme limit is SL as a source of nuclear fusion. The viability of this proposition will be addressed.

Finally, because seeing is believing, there will be a demonstration of SL (with any luck).

The following references are suggested. Additional material may be found in the references provided in these articles.

1 L. A. Crum, "Sonoluminescence," Physics Today, 22-29, September 1994.

2. S. J. Putterman, "Sonoluminescence: Sound into Light," Scientific American, 46-51, February 1995.

3. B. P. Barber and S. J. Putterman, "Observations of Synchronous Picosecond Sonoluminescence," Nature, Vol. 352, 318-323, 1991.

4. B. P. Barber, R. Hiller, K. Arisaka, H. Fetterman, and S. J. Putterman, "Resolving the picosecond characteristics of synchronous sonoluminescence," J. Acoust. Soc. Am., Vol. 91, 3061- 3063, 1992.

5. R. Hiller, S. J. Putterman and B. P. Barber, "Spectrum of Synchronous Picosecond Sonoluminescence," Phys. Rev. Lett., Vol. 69, 1182-1184, 1992.

FUNDAMENTALS AND APPLICATIONS OF NONLINEAR ACOUSTICS

Mark F. Hamilton
Department of Mechanical Engineering
The University of Texas at Austin
Austin, Texas 78712-1063

Abstract

In recent years, sessions on nonlinear acoustics at meetings of the Acoustical Society of America have focused on topics that include modeling of sonic boom rise times to predict noise due to supersonic aircraft, extracorporeal lithotripsy (the disintegration of kidney stones with focused shock waves), nonlinear losses in thermoacoustic engines, use of acoustic radiation pressure to levitate and position objects in the microgravity environment provided by spacecraft, and acoustical measurement of third-order elastic properties of rock. Each of these processes is associated with nonlinear terms that are discarded in the small-signal approximation (sometimes called, regrettably, the acoustic approximation) of the equations of motion and state.

The purpose of this lecture is to provide an introduction to the basic theory of nonlinear acoustics, and to illustrate the principal nonlinear effects with results from a variety of experiments. Topics of discussion will include: distortion and shock formation in finite amplitude waves; harmonic generation and spectral interactions; effects of absorption and dispersion in thermoviscous fluids, relaxing media, and bubbly liquids; diffraction of intense sound beams; parametric arrays; radiation pressure; acoustic streaming. While the emphasis is on finite amplitude sound in gases and liquids, similarities and differences with nonlinear properties of elastic waves in solids will be noted.

The road map for this lecture is the accompanying overview article having the same title (M. F. Hamilton, "Fundamentals and Applications of Nonlinear Acoustics," ASME, 1986). To prepare for the lecture, students may benefit from perusal of the following sections: Introduction (history); I. Basic Concepts (equations for distortion and shock formation in plane waves); II. Acoustic Saturation (the upper limit on how much sound power can be transmitted through a fluid); III. Suppression of Sound by Sound (use of one sound wave to pump energy out of another); VII. Parametric Arrays (a novel method of beamforming); VIII. Self-Demodulation (nonlinear distortion of a pulse).

The second review article (D. T. Blackstock, "Nonlinear Acoustics—Theoretical," AIP, 1972) is included primarily for future reference. It contains the main equations used to describe waveform distortion and shock formation in one-dimensional progressive sound waves.

FUNDAMENTALS AND APPLICATIONS OF NONLINEAR ACOUSTICS

M. F. Hamilton
Department of Mechanical Engineering
The University of Texas at Austin
Austin, Texas

ABSTRACT

The fundamental theory of nonlinear acoustics in fluids is reviewed in the context of practical applications. First the classical theory of nonlinear sound wave propagation is presented. Examples of acoustic saturation and the suppression of sound by sound are given. Spreading and thermoviscous losses are then taken into account. A spectral analysis is presented that is useful for numerical computation of finite amplitude wave propagation. Classical models of parametric transmitting and receiving arrays are discussed, including the self-demodulation property of pulsed parametric arrays. Considerable attention is given to the combined effects of nonlinearity, absorption, and diffraction in directive sound beams. The distinction between linear and nonlinear farfield regions is emphasized. Analytically simple Gaussian beam models are presented with transformations that permit their application to radiation from circular sources.

INTRODUCTION

Only during the last few decades has nonlinear acoustics evolved into its own distinctive area of research. The development of high power sound sources and improved signal processing techniques has supported a growing number of practical applications that take advantage of finite amplitude sound. Nonlinear sonars referred to as parametric arrays [1] have found widespread use in underwater applications. An increasingly important parameter for characterizing body tissue with medical ultrasonics is the acoustical nonlinearity of the specimen [2]. Nonlinear effects are responsible for the improved resolution observed when acoustic microscopes are operated at high sound power levels [3]. The propagation of noise from underwater explosions [4] and jet aircraft engines [5] is an inherently nonlinear phenomenon. Just recently, the Space Shuttle carried aboard an apparatus for levitating and positioning objects with high intensity sound. The list of these and other applications is expanding as advances continue to be made in the science and technology of nonlinear acoustics.

Although problems in nonlinear acoustics have been pondered since the 18th century, successful contributions to the science have been intermittent, the early history of which has been chronicled by Blackstock [6]. Perhaps the first (although slightly incorrect) nonlinear wave equation governing finite amplitude sound was published by Euler [7] in 1759. Shortly thereafter, Lagrange [8] derived a (once again slightly incorrect) general solution which yielded a propagation speed that depends on the local amplitude of the sound wave. Lagrange concluded that his solution was in error because he felt that a nonconstant propagation speed, the basic mechanism in nonlinear distortion, violated the observed behavior of sound. Not until 1808 did Poisson [9] derive an exact solution for progressive plane waves in an isothermal gas, although he too did not understand the consequences of his solution. The implication of the nonconstant propagation speed that bothered Lagrange and Poisson was explained by Stokes [10], who in 1848 published the first illustration of waveform steepening. Stokes also produced the first analysis of shock waves and suggested that viscosity and heat conduction are responsible for the inevitable losses at the shock front. The beginning era of research in nonlinear acoustics concluded with two important papers published around 1860. Riemann [11] obtained a solution for plane waves that

travel in opposite directions by what is now referred to as the method of characteristics, and Earnshaw [12] tackled the difficult problem of describing the sound radiated by the arbitrary finite displacement of a piston. A number of the early papers have been compiled recently by Beyer [13] in a collection of benchmark papers in nonlinear acoustics.

Apart from the analyses of shock waves by Rankine, Hugoniot, Rayleigh, and Taylor, the next significant advance in the understanding of finite amplitude sound awaited the publication of three papers in the 1930's. Two Fourier series solutions were derived for plane waves that start out as pure tones and subsequently undergo harmonic distortion as they propagate. One by Fubini [14] applies to waves in a lossless fluid prior to shock formation, and the other by Fay [15] is an asymptotic solution for strong waves subjected to viscous losses. While these two papers provide the first explicit models of harmonic generation in sound waves, the third paper by Thuras, Jenkins, and O'Neil [16] gives an account of the first experimental investigation of this phenomenon.

Wave equations derived around 1950 by Eckart [17], Lighthill [18], and Mendousse [19] opened the modern era of nonlinear acoustics. The equations of Eckart and Lighthill permitted the study of nonplanar finite amplitude waves, while Mendousse demonstrated that the Burgers equation may be used to model plane waves in viscous fluids. A considerable number of papers then followed both in the United States and the Soviet Union. Generalizations of the Burgers equation were shown by Khokhlov and coworkers to model spherical [20] and cylindrical [21] waves, and detailed solutions of the plane wave Burgers equation were investigated in detail by Blackstock [22]. By far the most noteworthy contribution in terms of far reaching practical applications was the work by Westervelt on the scattering of sound by sound [23,24] which ultimately led to his theory for the parametric array [1] in 1960. In the parametric array, the nonlinear interaction of two high frequency sound beams produces a narrow beam of low frequency sound having virtually no sidelobes. The process permits the radiation of highly directive sound from a relatively small transducer, with the added benefit that wide frequency bands may be transmitted. Experimental verification of the parametric array by Bellin and Beyer [25] was reported in the paper following that of Westervelt [1] at a meeting of the Acoustical Society of America.

Unlike any other single application of nonlinear acoustics, the parametric array was responsible for a renaissance of research activity, the history of which is discussed in a review by Muir [26]. Even though the parametric array was first conceived and tested in the United States, subsequent work on the device before 1968 appeared only in England and Norway. Berktay [27-29] worked out examples for a variety of possible applications afforded by the parametric array, although a few of his theoretical predictions were later found [30,31] to be somewhat overoptimistic. Nevertheless, as Muir [26] points out, the lack of activity in the United States ended when it became evident from Berktay's work that the attractive features of the parametric array often outweigh the inherent inefficiency of the device. By the early 1970's, parametric sonars were in use in both civilian and military applications, and papers on parametric arrays presented at acoustical society meetings and symposia numbered more that 100 per year. Recently an entire textbook devoted to parametric arrays was published by Novikov, Rudenko, and Timoshenko [32].

The next turning point for the practical application of nonlinear effects in sound beams resulted from the theoretical work performed in the Soviet Union around 1970 by Zabolotskaya and Khokhlov [33] and by Kuznetsov [34]. The result was a parabolic nonlinear wave equation, referred to in the Soviet Union as the KZK (Khokhlov-Zabolotskaya-Kuznetsov) equation, which provides an excellent model for the combined effects of nonlinearity, diffraction, and dissipation in directive sound beams. Similar parabolic equations had previously been used to describe the propagation of laser beams, and consequently the model is sometimes referred to as the quasioptical approximation. Throughout the 1970's there appeared in the Soviet literature a large volume of work based on the KZK equation, much of which was devoted to Gaussian beams.

A review of the numerical work was published recently in a book by Bakhvalov, Zhileikin, and Zabolotskaya [35]. However, the most definitive research on nonlinear effects in sound beams was published this decade in a series of papers by the Norwegian mathematicians Naze Tjøtta, Tjøtta, and coworkers [36-42]. Their work is important in that it accounts for the intricate diffraction effects that exist in high intensity sound fields radiated from circular sources, a geometry which characterizes most ultrasonic transducers.

The purpose of this review is to highlight the fundamental theory of nonlinear acoustics while aiming toward results that have found recent application. The review begins with the classical theory of nonlinear effects in plane wave propagation. Distortion leading up to shock formation is discussed, with the role of thermoviscous and spreading losses taken into account. Examples of acoustic saturation and the suppression of sound by sound are given. Various model equations of nonlinear acoustics are presented, and methods for obtaining both analytical and numerical solutions are pointed out. The classical model for the parametric array is discussed. Finally, the recently well understood role of nonlinear effects in directive sound beams is stressed because of its many practical applications.

Any review of an area of research inevitably reflects the interests of the author, and in any event it cannot do justice to the wide variety of topics usually involved. For example, strong shock waves, radiation pressure, and acoustic streaming are not considered. Two books on nonlinear acoustics may be referred to for more complete discussions. One by Beyer [43] highlights the many achievements in the field, including much of the experimental work. The other by Rudenko and Soluyan [44] is more of a textbook, but whose treatment of recent developments is restricted to work done in the Soviet Union.

I. BASIC CONCEPTS

A particularly simple equation may be derived for a progressive plane wave that propagates in a perfect isentropic gas (see Ref. 45):

$$\frac{\partial u}{\partial t} + (c_0 + \beta u)\frac{\partial u}{\partial x} = 0, \tag{1}$$

where $u(x,t)$ is the particle velocity of the gas, $c_0 = \sqrt{\gamma P_0/\rho_0}$ is the small signal sound speed, P_0 and ρ_0 are ambient values of the pressure and density, respectively, γ is the specific heat ratio, and $\beta = (\gamma + 1)/2$ is the coefficient of nonlinearity. Equation (1) is exact, from which we obtain for the propagation speed of a finite amplitude acoustical disturbance

$$\frac{dx}{dt} = c_0 + \beta u. \tag{2}$$

Points (wavelets) of the waveform having different particle velocity amplitudes thus propagate at different speeds. There are two physical mechanisms responsible for this phenomenon. First, the passage of a sound wave is accompanied by pressure, density, and temperature fluctuations in the medium. As a fluid is compressed, its stiffness increases in such a way as to cause a corresponding increase in the sound speed. A compression phase of a wave therefore sees a slightly stiffer medium with a higher sound speed than that for an expansion phase. For a perfect isentropic gas the sound speed is

$$c = c_0 + \left(\frac{\gamma - 1}{2}\right)u, \tag{3}$$

where the small signal sound speed c_0 is the limiting value for infinitesimal wave amplitudes. The second mechanism that contributes to the nonconstant propagation speed is convection. The motion of the fluid represented by the particle velocity actually constitutes a local flow field.

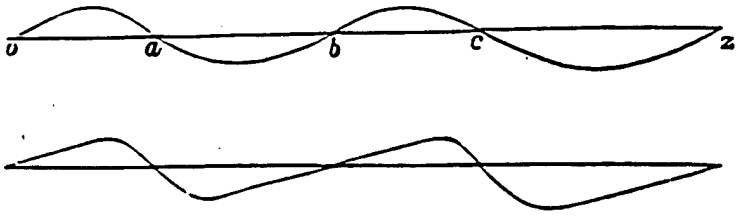


Fig. 1—Sketch of waveform distortion.

To account for the convection of a wavelet by this flow, the local velocity u must be added to Eq. (3) in order to recover the propagation speed in Eq. (2).

The nonconstant propagation speed causes peaks of a sound wave to travel faster than the troughs, a process which can eventually lead to shock formation. Shown in Fig. 1 is the first illustration of waveform distortion from the paper by Stokes [10]. The upper curve represents the spatial profile of a waveform at some instant in time, with the letters a, b, and c marking the zero crossings of the wave. According to Eq. (2), all zero crossings propagate at the same speed c_0. In the lower portion of the figure is a second snapshot of the same wave but at a later time and at a distance $c_0 t$ farther away. In other words, we are in a coordinate system that moves with the zero crossings. Note that the peaks have advanced and the troughs have receded in relation to the zero crossings. Left unchecked by other physical mechanisms, the peaks will catch up with and eventually overtake the troughs, culminating in the physical impossibility of a multivalued waveform. Such an implication was of no small concern to the physicists of the 19th century. In fact, it prompted one [46] to conclude that "plane waves are...impossible" (see Ref. 6 for the ensuing debate), a remark that motivated the paper by Stokes. The key issue here is that thermal and viscous losses, not accounted for in Eq. (1), offset the tendency of a wave to form a discontinuity. More will be said later about the relative effects of nonlinearity and dissipation on waveform distortion.

Poisson [9] obtained a general solution of Eq. (1) (although for an isothermal rather than an isentropic gas) which may be written as either

$$u = f\left(t - \frac{x}{c_0 + \beta u}\right) \qquad \text{or} \qquad u = g\big(x - (c_0 + \beta u)t\big), \tag{4}$$

where f and g are arbitrary functions. The first solution is suitable for boundary value problems, the second for initial value problems. It is difficult to apply either in its present implicit form, but the first yields useful explicit results if we make an approximation. Suppose that $u/c_0 \ll 1$. Values of u/c_0 less than 0.1 are usually sufficient, which in air correspond to sound pressure levels up to 174 dB *re* 20 μPa and encompass levels that are generated by jet engines. The expansion $(c_0 + \beta u)^{-1} \approx c_0^{-1} - \beta u/c_0^2$ may then be used to write the first of Eqs. (4) as

$$u = f\left(\tau + \frac{\beta u x}{c_0^2}\right), \qquad u/c_0 \ll 1, \tag{5}$$

where $\tau = t - x/c_0$ is the retarded time. Now assume that the source excitation is time harmonic, and let

$$u(x = 0) = u_0 \sin \omega t. \tag{6}$$

It may then be shown that the range where shock formation occurs, i.e., where an infinite slope first appears on the waveform, is given by [47]

$$\overline{x} = \frac{1}{\beta \epsilon k}, \tag{7}$$

where $k = \omega/c_0$ is the wavenumber and $\epsilon = u_0/c_0$ the peak acoustic Mach number at the source. Shock formation occurs closer to the source as either the amplitude or frequency of the wave is

increased. As the frequency increases the wavelength decreases, and a peak need not travel as far to catch up with a trough.

Since Eq. (1) does not account for any losses, shock formation is inevitable and the theory eventually breaks down. The dilemma is resolved by incorporating the weak shock theory outlined by Blackstock [45,48]. From the Rankine-Hugoniot relations it can be shown that the propagation speed of a weak shock is

$$\left(\frac{dx}{dt}\right)_{shock} = c_0 + \frac{\beta}{2}(u_a + u_b), \tag{8}$$

where u_a and u_b are the values of the particle velocity immediately ahead of behind the shock, respectively. Theory based on Eq. (1) is valid for all continuous segments of the wave, and it may be used to evaluate u_a and u_b at the shock. The combined application of Eqs. (2) and (8) thus describes the behavior of all parts of the wave not only prior to but also following shock formation. The foregoing procedure constitutes the weak shock method, which has been used to model the propagation of N waves [45] and exponential pulses [4] from explosive sources.

We now return to the problem of pure tone radiation. The solution of Eq. (5) subject to Eq. (6) may be written in the form of the Fourier series

$$u = u_0 \sum_{n=1}^{\infty} B_n \sin n\omega\tau. \tag{9}$$

Prior to shock formation the Fourier coefficients are given by

$$B_n = \frac{2J_n(n\sigma)}{n\sigma}, \qquad \sigma < 1, \tag{10}$$

where J_n is the Bessel function of order n, and the dimensionless parameter $\sigma = \beta\epsilon kx$ measures range in terms of the shock formation distance $\bar{x}$. Equation (10) is known as the Fubini [14] solution, one of the most well known explicit solutions in nonlinear acoustics. The general solution for $\sigma > 1$ is rather involved [48], but for $\sigma > 3$ a very simple asymptotic form is obtained:

$$B_n = \frac{2}{n(1+\sigma)}, \qquad \sigma > 3. \tag{11}$$

Equations (9)–(11) are valid for waves of sufficient strength that thermal and viscous losses do not prevent the formation of well defined shocks. The case history of such a wave is shown

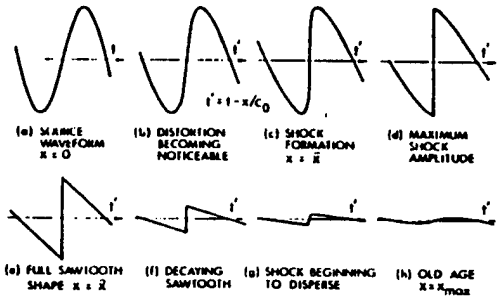


Fig. 2—Case history of a high intensity sound wave.

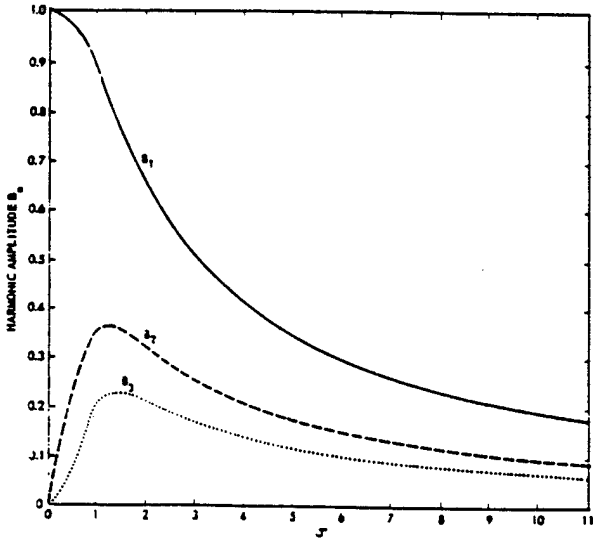


Fig. 3—Harmonic amplitudes of an originally sinusoidal wave.

in Fig. 2 [49], with the first three harmonic amplitudes graphed in Fig. 3 [48]. In Fig. 2*a* is shown one cycle of a sinusoidal source waveform. The subsequent distortion process leading up to shock formation (Fig. 2*c*) is accompanied by an upshift of energy from the spectral component at the fundamental source frequency to the higher harmonics (Fig. 3). Note that in Fig. 2 the abscissa represents time whereas in Fig. 1 it represents range, which accounts for the two figures showing waves that steepen in opposite "directions." At $\sigma = 3$ (Fig. 2*e*), where Eq. (11) is valid, the waveform has acquired a stable sawtooth shape. Energy in the wave is now continuously dissipated at the shock, and the amplitude of the fundamental component is already down by 50% (Fig. 3). For a while the wave maintains its sawtooth shape but, according to Eq. (11), it decays as $(1+\sigma)^{-1}$ (Fig. 2*f*). Eventually the amplitude is sufficiently reduced that nonlinear effects can no longer maintain a shock in the presence of the inevitable thermal and viscous losses (Fig. 2*g*). Thermoviscous losses effectively low-pass filter the wave, which ultimately yields a wave dominated by the fundamental source frequency component (Fig. 2*h*). However, the amplitude of the wave in this *old age* region is much lower than what would have existed in the absence of nonlinear effects.

Some general comments on the nature of *cumulative* versus noncumulative, or *local* nonlinear effects are in order. It can be seen from Eq. (2) and Fig. 2 that distortion due to the nonconstant propagation speed produces effects that accumulate with distance. For example, in the case of pure tone radiation in an inviscid and thermally nonconducting fluid, a shock must eventually form no matter how small the source amplitude (see Eq. (7)). Now recall the source condition in Eq. (6). A real physical source would involve displacement of the radiating surface about the point $x = 0$. Taking into account the finite displacement of the source introduces new distortion components which grow less rapidly with range than those resulting from the nonconstant propagation speed [47]. The new components represent local effects which are significant only on the order of one wavelength from the source, and they are usually ignored. Local effects also appear when the exact nonlinear impedance relation is used to express either the source condition or final solution in terms of the acoustic pressure p rather than the particle velocity u. For example, beyond one wavelength from the source, the linear plane wave impedance relation $p = \rho_0 c_0 u$ may be used to write Eq. (9) in terms of pressure, as is often done.

Finally, the theory in this section is easily generalized to liquids. Recall that the nonconstant propagation speed is due in part to the finite amplitude sound speed given in Eq. (3). Consistent with the approximation $u/c_0 \ll 1$ used in Eq. (5) is the expression

$$c = c_0 + \left(\frac{B}{2A}\right) u \tag{12}$$

for the speed of sound in an arbitrary liquid, where B/A is the ratio of coefficients in a Taylor series expansion of the isentropic equation of state [50] (see also Ref. 43). The generalized coefficient of nonlinearity thus becomes $\beta = 1 + B/2A$, where for a perfect isentropic gas B/A is replaced by $(\gamma - 1)$. Typical values are $\beta = 1.2$ for air and $\beta = 3.5$ for water. Since the "1" in β represents convection, the dominant cause of waveform distortion in air is convection, whereas in water it is the nonlinearity in the equation of state.

II. ACOUSTIC SATURATION

The irreversible energy loss at shock fronts imposes an upper bound on how much sound power can be transmitted beyond a certain range. Suppose x is fixed at a location beyond $\overline{x}$ and either the source amplitude or frequency is increased. The shock formation distance $\overline{x} = (\beta \epsilon k)^{-1}$ moves closer to the source, which allows more and more energy to be dissipated at the shocks before the wave arrives at the observation point. A limit is eventually reached, say by increasing the source amplitude, where $\sigma \gg 1$ and Eqs. (9) and (11) reduce to

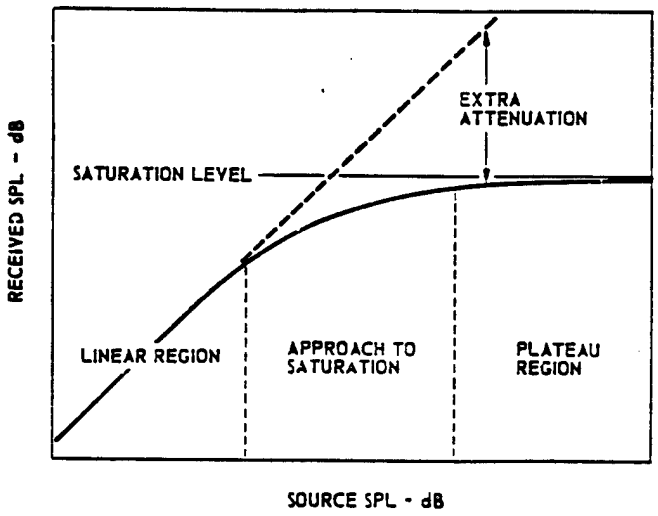


Fig. 4—Amplitude response curve showing the development of saturation.

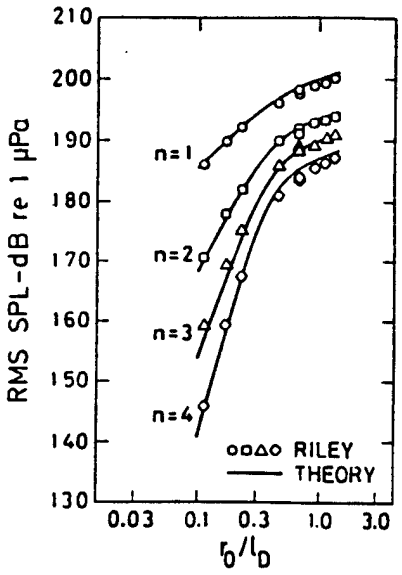


Fig. 5—Saturation of harmonic components in a directive sound beam.

$$u = \frac{2c_0}{\beta k x} \sum_{n=1}^{\infty} \frac{1}{n} \sin n\omega\tau, \qquad \sigma \gg 1. \tag{13}$$

Note that the wave no longer depends on the source amplitude u_0. All additional energy pumped into the wave by the source is lost at the shock fronts, and acoustic saturation is said to have occurred.

In Fig. 4 [51] is an amplitude response curve showing the development of saturation. The distance between the source and receiver is held constant, and the received sound pressure level is monitored as the source level is varied. At sufficiently low source levels the response of the fluid to the passage of the sound wave is linear, and any change in the source level causes an identical change in the received level. As the source level is increased, saturation eventually sets in. The energy lost at the shocks is responsible for the extra attenuation noted in Fig. 4. The extra attenuation of the fundamental component in decibels, often referred to as EXDB, has been calculated by Blackstock [22] for plane waves in thermoviscous fluids (see also Eq. (18)).

Acoustic saturation is not restricted to plane waves. The first correctly interpreted observation of the phenomenon was made in 1950 by Allen [52] in an experimental investigation of high intensity spherical waves in air. Subsequent experimental work on saturation involved not only plane waves [51] but also directive sound beams [49,53]. In Fig. 5 [42] are shown data reported by Riley [53] for the saturation of a 470 kHz sound beam radiated in water by a circular source of radius 3.8 cm. Measurements of the fundamental through fourth harmonic component were taken in the center of the beam at a distance of 11.2 m from the source. Along the abscissa is the normalized source amplitude in terms of the Rayleigh distance r_0 (1.4 m) and the plane wave shock formation distance l_D (the same as given by Eq. (7)). The dimensionless amplitude may therefore be rewritten $r_0/l_D = \beta\epsilon k r_0$. It can be seen that each harmonic component saturates at approximately the same source level. Note that in the linear region ($r_0/l_D < 0.2$), the rate of growth of each harmonic component increases with n. Expansion of Eq. (10) reveals that at low amplitudes $B_n \propto \epsilon^n$, which is the same trend that is displayed in Fig. 5. The bottom line on acoustic saturation is that large amounts of power may be wasted by overdesigned sound sources. Muir [26] points out that the concept of saturation was not appreciated by sonar engi-

neers until the 1970's, when many existing systems could have operated with much less power and yet produced the same sound pressure level at the target.

III. SUPPRESSION OF SOUND BY SOUND

The suppression of sound by sound refers to the use of one wave to pump energy *out* of another wave. Consider the bifrequency source condition

$$u(x=0) = u_p \sin \omega_p t + u_w \sin \omega_w t. \tag{14}$$

Fenlon [54] obtained for the preshock region an exact Fubini-type solution of Eq. (5) for the boundary condition in Eq. (14). Of interest here is the interaction of a strong, low frequency (pump) wave (ω_p) with a weak, high frequency wave (ω_w). The weak wave has little effect on the propagation of the pump wave, and harmonic distortion of the latter is adequately described by Eq. (10). However, the pump wave will modulate the weak wave and generate sidebands at $\omega_w \pm n\omega_p$. An approximation of Fenlon's bifrequency solution provides an expression for the wave formed by ω_w and its sidebands [55]. Here we are concerned only with the source frequency component $u(\omega_w)$ which is given by

$$u(\omega_w) = u_w J_0(\beta \epsilon_p k_w x) \sin \omega_w \tau, \qquad u_p \gg u_w, \qquad \omega_p \ll \omega_w, \tag{15}$$

where $\epsilon_p = u_p/c_0$ and $k_w = \omega_w/c_0$. The nonlinear effect of the pump wave on the weak wave, subject to the restrictions indicated in Eq. (15), does not depend on the amplitude of the weak wave or the frequency of the pump wave. Note that $u(\omega_w)$ vanishes at zeros of the Bessel function. Whenever $u(\omega_w)$ vanishes, all of its energy has been pumped into adjacent sidebands. Thus the energy in $u(\omega_w)$ is not absorbed but merely redistributed at other frequencies. Experimental studies of the suppression of sound by sound have been performed in both air [55] and water [56]. Suppression has also been observed when the two interacting waves propagate in different directions [57].

Shown in Fig. 6 [58] are results from a suppression experiment performed by Gong, Zhu, and Du [59] in an air filled plane wave tube. The frequencies used were 899 Hz for the pump wave ($\Omega = \omega_p$) and 4023 Hz for the weak wave ($\omega = \omega_w$). As the amplitude of the pump wave ($U_0/c_0 = \epsilon_p$) was increased from 0 to 150 dB, a point was reached where 47 dB suppression of the high frequency weak wave was observed at a fixed location downstream from the source. Along

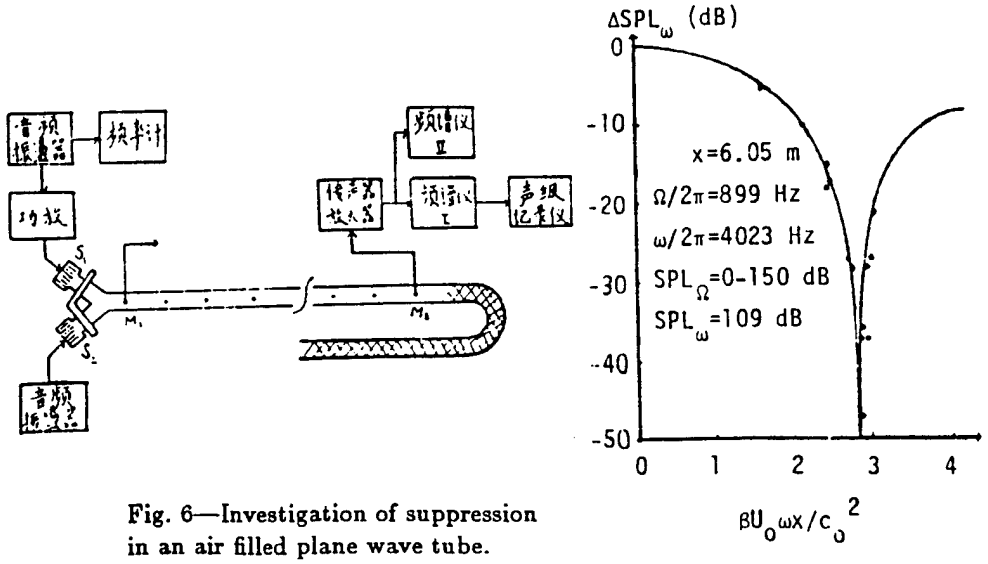


Fig. 6—Investigation of suppression in an air filled plane wave tube.

the abscissa in Fig. 6 is the argument of the Bessel function in Eq. (15). Although the first zero of the Bessel function occurs when its argument has the value 2.4, the maximum suppression shown in Fig. 6 occurs when $\beta\epsilon_p k_w x = 2.8$. Such discrepancies have in the past been attributed to the effects of absorption [55,58].

So far, interest in the suppression of sound by sound has been primarily academic, since practical application of the phenomenon, in particular to noise control, is apparently thwarted by various inherent difficulties [58]. Energy pumped out of the weak wave ends up in sidebands which, by virtue of the restriction $\omega_p \ll \omega_w$, are located very close to the frequency where suppression is desired. A possible remedy would seem to result from choosing $\omega_p \gg \omega_w$, for then the sidebands would cluster around the pump frequency rather than around the frequency of the weak wave. Unfortunately, however, high frequency pump waves are very ineffective in suppressing low frequency waves [60]. An alternative application was considered by Westervelt [61], who developed a theory that predicts the amount of energy removed from a pure tone as it propagates through an isotropic noise field. Westervelt's theory, which was subsequently verified by experiment [62,63], may account for additional losses suffered by underwater sound that propagates long distances through ambient noise fields.

IV. THERMOVISCOUS LOSSES

The nonlinear propagation of plane waves in a viscous fluid was shown by Mendousse [19] to be governed by the Burgers equation. Subsequent generalizations by Lighthill [64] to thermoviscous gases and later by Tjøtta [65] to arbitrary thermoviscous fluids permit the Burgers equation to be written

$$\frac{\partial u}{\partial x} - \frac{b}{2\rho_0 c_0^3}\frac{\partial^2 u}{\partial \tau^2} = \frac{\beta}{c_0^2} u \frac{\partial u}{\partial \tau}. \tag{16}$$

Thermoviscous effects are accounted for by $b = \zeta + (4/3)\eta + \kappa(1/c_v - 1/c_p)$, where ζ is the bulk viscosity, η the shear viscosity, κ the thermal conductivity, and c_v and c_p the specific heats at constant volume and pressure, respectively. As implied by the retarded time τ, the Burgers equation applies only to progressive wave motion. The local nonlinear effects discussed in Sec. I are not accounted for in Eq. (16).

Both exact and asymptotic solutions of Eq. (16) that satisfy the pure tone source condition in Eq. (6) have been investigated by Blackstock [22]. Exact solutions are obtained by employing the nonlinear Hopf-Cole transformation, which allows Eq. (16) to be written as a linear diffusion equation in terms of the new dependent variable. However, the exact analytical solutions are unwieldy and seldom used in practice. Here we shall consider only a farfield asymptotic solution for the pure tone source condition. The solution for strong plane waves that have propagated beyond the shock formation distance was derived by Fay [15]:

$$u = u_0 \frac{2}{\Gamma} \sum_{n=1}^{\infty} \frac{\sin n\omega\tau}{\sinh n(1+\sigma)/\Gamma}, \qquad \sigma > 3, \qquad \Gamma \gg 1. \tag{17}$$

The quantity $\Gamma = \beta\epsilon k/\alpha$ is called the Gol'dberg [66] number (see Ref. 22), where $\alpha = b\omega^2/2\rho_0 c_0^3$ is the thermoviscous attenuation coefficient at the source frequency. The criterion $\Gamma \gg 1$ describes a plane wave of sufficient strength, relative to thermoviscous effects, that a shock will form around $\sigma = 1$. In the absence of thermoviscous effects ($\Gamma = \infty$), Eq. (17) reduces to the sawtooth solution given by Eqs. (9) and (11). Now suppose that we move sufficiently far away from the source that the old age region shown in Fig. $2h$ is reached, where

$$u = \frac{4c_0\alpha}{\beta k} \sum_{n=1}^{\infty} e^{-n\alpha x} \sin n\omega\tau, \qquad \sigma \gg \Gamma \gg 1. \tag{18}$$

Saturation has again occurred, but the amplitude here is different from that in Eq. (13) because of thermal and viscous losses. Now suppose that the nonlinear term on the right hand side of Eq. (16) is neglected, and linear solutions for time harmonic waves are derived. It would be found that the attenuation coefficient of a wave at frequency $n\omega$ is $n^2\alpha$, where α is the coefficient at frequency ω. The linear attenuation coefficient thus depends quadratically on frequency. In the old age region, however, where one might suppose the wave to be sufficiently weak as to propagate according to linear theory, the attenuation coefficient of the nth harmonic component is $n\alpha$ and not $n^2\alpha$. The decay rates of the spectral components are therefore indicative of whether the signal has *ever* experienced strong nonlinear effects. Or to put it another way, a wave that was once finite amplitude is always finite amplitude.

A simple approximate expression was derived by Merklinger [67] for the source frequency component of a plane wave in a thermoviscous fluid:

$$u = \frac{u_0 e^{-\alpha x} \sin \omega\tau}{\sqrt{1 + (\Gamma/4)^2 (1 - e^{-2\alpha x})^2}}. \tag{19}$$

Equation (19) differs from the exact solution of Eq. (16) for the fundamental component by at most 1 dB. The numerator of Eq. (19) is the linear solution, and the denominator accounts for the nonlinear losses incurred when Γ is sufficiently large. A suitable criterion for which nonlinear effects may be ignored altogether is evidently $\Gamma < 1$. A comparison of Merklinger's model with other approximate solutions is made in Appendix C of Ref. 41. The perturbation methods that lead to Eq. (19) have also been applied to spherical waves [68].

V. SPREADING LOSSES

Spreading rather than thermoviscous losses are taken into account by rewriting Eq. (16) (see Ref. 69)

$$\frac{\partial u}{\partial r} + \frac{a}{r} u = \frac{\beta}{c_0^2} u \frac{\partial u}{\partial \tau}, \tag{20}$$

where $a = 0, 1/2, 1$ for plane, cylindrical, or spherical waves. A restriction on Eq. (20) for nonplanar waves ($a \neq 0$) is that $kr \gg 1$, where k is a characteristic wavenumber and r is range. The restriction is relatively weak in practice, since it merely requires a wave to possess a radius of curvature that exceeds roughly one wavelength. For example, a sound source whose radius is smaller than one wavelength is very inefficient and therefore unlikely to radiate finite amplitude sound.

An exact solution of Eq. (20) for plane waves ($a = 0$) is given by Eq. (5) (with $x = r$), and a simple transformation permits generalization of the solution to cylindrical and spherical waves [69]. By defining the new dependent variable $w = (r/r_0)^a u$, where r_0 is a reference distance, we may eliminate the spreading loss term in Eq. (20). A new independent variable z may then be defined that reduces the equation for $w(z, \tau)$ to one identical in form to that for $u(r, \tau)$ when $a = 0$. The necessary transformation is given by:

$$\text{cylindrical waves } (a = 1/2): \qquad z = 2\sqrt{r_0}(\sqrt{r} - \sqrt{r_0}); \tag{21a}$$

$$\text{spherical waves } (a = 1): \qquad z = r_0 \ln (r/r_0). \tag{21b}$$

All results derived from Eq. (5) may now be used for cylindrical and spherical waves by replacing x everywhere with z and multiplying the final solution by $(r_0/r)^a$. The reference range r_0 can be either the source radius or any other point in space where the waveform is known. For example, to obtain the shock formation distance for a spherical wave whose time dependence at r_0 is $u_0 \sin \omega t$, set z in Eq. (21b) equal to $(\beta\epsilon k)^{-1}$ and solve for $r = \bar{r}$ to obtain

$$\bar{r} = r_0 e^{1/\beta\epsilon k r_0}. \tag{22}$$

Note that the shock formation distance can be considerably farther from a spherical source than from a planar source that radiates at the same amplitude and frequency. Shock formation is nevertheless inevitable in the absence of thermoviscous losses, regardless of spreading losses.

The above analysis was used by Lockwood, Muir, and Blackstock [70] to investigate the propagation of directive spherical waves. Sufficiently far from any directive sound source, diffraction effects are negligible and the radiated waves obey spherical propagation laws. If the thermoviscous attenuation is not too high, most of the nonlinear distortion occurs in the spherical farfield of the source. Now choose for the reference distance r_0 the range where the spherical wave region begins, and let

$$u(r = r_0) = u_0 D(\theta) \sin \omega t, \tag{23}$$

where $D(\theta)$ is the farfield directivity function associated with linear radiation from the source at frequency ω. In writing Eq. (23) it is assumed that nonlinear effects due to propagation up to the range r_0 may be ignored. Beyond r_0 the results for spherical waves may be used, with the acoustic Mach number at r_0 given by $\epsilon D(\theta)$. The solution is thus found by replacing σ with $\beta \epsilon D(\theta) k r_0 \ln(r/r_0)$ in Eq. (10). Expansion of Eq. (10) for weak nonlinearity then yields for the directivity function $D_n(\theta)$ of the nth harmonic component

$$D_n(\theta) = D^n(\theta). \tag{24}$$

The power law dependence on the directivity function of the fundamental component shows that beam patterns of the nonlinearly generated higher harmonics become increasingly narrower as n increases. For example, the first sidelobe of radiation from a circular source is around 18 dB below the main lobe. According to Eq. (24), the second harmonic radiation has not only a narrower main lobe, but its first sidelobe is down by almost 36 dB. Although the power law dependence is valid only at large distances from the source (see Sec. X), it provides a very convenient estimate of directive harmonic radiation.

Shown in Fig. 7 are results from an experiment performed by Muir [71] that demonstrate how the properties governed by Eq. (24) may be used to enhance the resolution of acoustic imaging systems. A sound source radiating at 100 kHz was used underwater to image a barge located 100 m away. In the first column are beam patterns for the fundamental component and the first few nonlinearly generated harmonics. Note both the narrowing of the beam pattern and improved sidelobe suppression with increasing harmonic number. The second column shows the images obtained by monitoring the associated frequencies. Resolution improves as the harmonic number increases, with the higher harmonics clearly able to define pilings that are 15 cm in diameter. Although the higher frequencies can usually be generated directly by

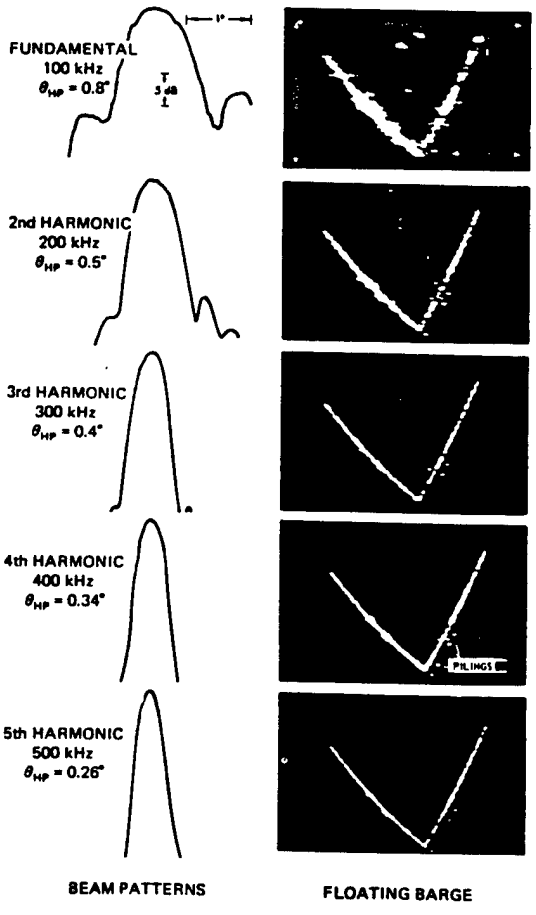


Fig. 7—Use of nonlinearly generated harmonics for acoustic imaging.

the sound source, lower sidelobes and therefore less reverberation problems result from generating the higher frequencies nonlinearly.

VI. NUMERICAL SOLUTION

The combined effects of nonlinearity, thermoviscous dissipation, and spreading losses are taken into account by single equation

$$\frac{\partial u}{\partial r} + \frac{a}{r}u - \frac{b}{2\rho_0 c_0^3}\frac{\partial^2 u}{\partial \tau^2} = \frac{\beta}{c_0^2}u\frac{\partial u}{\partial \tau}. \tag{25}$$

Attempts have been made to rewrite Eq. (25) in a form for which exact analytical solutions are known. Khokhlov and coworkers [20,21] showed by using the transformations $w = (r/r_0)^a u$ and Eqs. (21) that although the spreading loss term in Eq. (25) can be eliminated, the coefficient of the term that accounts for thermoviscous effects becomes a function of range for nonplanar waves. So far, no exact solutions of Eq. (25) have been found for nonplanar waves, and analytical approaches are usually abandoned in favor of numerical techniques. A number of the more popular numerical methods (e.g., see Refs. 72 and 73) are based on the spectral analysis of Fenlon [74]. The particle velocity is first expanded in terms of the complex Fourier series

$$u(r,\tau) = \sum_{n=-\infty}^{\infty} u_n(r)e^{jn\omega\tau}, \qquad \text{where} \qquad u_n(r) = \frac{\omega}{2\pi}\int_{-\pi/\omega}^{\pi/\omega} u(r,\tau)e^{-jn\omega\tau}d\tau. \tag{26}$$

Substitution of the first of Eqs. (26) in Eq. (25) yields the following coupled ordinary differential equations for the harmonic amplitudes u_n:

$$\frac{du_n}{dr} + \left(\frac{a}{r} + n^2\alpha\right)u_n = j\frac{n\omega\beta}{2c_0^2}\left(\sum_{m=1}^{n-1} u_m u_{n-m} + 2\sum_{m=n+1}^{\infty} u_m u_{m-n}^*\right), \tag{27}$$

where $\alpha = b\omega^2/2\rho_0 c_0^3$. In writing the summations it was assumed that u is real; therefore $u_{-n} = u_n^*$, where u_n^* is the complex conjugate of u_n.

Korpel [75] noted that the spectral form of Eq. (27) permits generalization to fluids that possess arbitrary absorption and dispersion. The thermoviscous attenuation coefficient $n^2\alpha$ may be replaced by a coefficient α_n that depends in an arbitrary fashion on n. Dispersion is included by making α_n complex and writing the imaginary part as $j\omega(c_n^{-1} - c_0^{-1})$, where c_n is the frequency dependent sound speed and c_0 the reference sound speed. Equation (27) can also be modified for application to finite amplitude waves in horns (see Ref. 45). The cross sectional area $A(x)$ of the horn, where x replaces the range variable r, must vary sufficiently slowly that the waves remain quasiplanar. Then the spreading loss coefficient a/r may be replaced by $A'/2A$, where A' is the derivative of A. The nonlinear horn equation reduces to the equation for spherical waves with conical horns ($A \propto x^2$) and to the equation for cylindrical waves with parabolic horns ($A \propto x$).

In numerical computations, the number of harmonics retained in Eq. (26) must be limited to some finite number N. Consequently, the upper limit on the second summation in Eq. (27) is also replaced by N. The coupled equations may then be solved by standard techniques (see Ref. 73). Calculations for the distortion of waves from monofrequency sources can be performed very rapidly and are conveniently done on personal computers. In fact, when there is no dispersion, for a monofrequency source only the imaginary parts of u_n need to be calculated, and the efficiency of the algorithm may be further increased. However, the algorithm is not particularly suitable for pulses or noise. Prohibitive numbers of harmonics must be retained in order to account for the harmonics generated by the highest frequencies in the initial waveform. Time domain algorithms such as those described in Refs. 5 and 76 are preferable for multifrequency sources. The amplitude dependent propagation speed is used to modify the waveform as it propagates step by step. Fourier transforms must be performed at each step in order to include the frequency dependent absorption.

VII. PARAMETRIC ARRAYS

As pointed out in the Introduction, the parametric array is probably the most well known practical application of nonlinear acoustics. The concept was originated by Westervelt [1] in 1960, and the last two decades have witnessed numerous scientific and technological advances made in a number of different countries. A large volume of papers has been written on the subject, many of which are discussed in reviews by leading researchers in the field [26,32,43,77–80]. Here we shall discuss only the fundamental properties of parametric arrays.

A parametric array is formed when two collimated, finite amplitude sound beams of different yet neighboring frequencies (ω_1 and ω_2, $\omega_1 \approx \omega_2$) are transmitted in the same direction. The nonlinear interaction of the two primary waves gives rise to a component at the difference frequency ($\omega_- = \omega_1 - \omega_2$). Additional components are also generated at other intermodulation and harmonic frequencies. Since the tendency of fluids is to low-pass filter sound waves, the difference frequency component, the lowest in the spectrum, is eventually the lone survivor. In Westervelt's model of the parametric array, it is assumed that linear absorption mechanisms limit the region of nonlinear interaction to the nearfields of the two primary waves. The amplitudes of the primary waves are attenuated as $e^{-\alpha_i r}$, where α_i is the attenuation coefficient at frequency ω_i. To a first approximation, the two primary beams may be modeled as plane waves that propagate in an imaginary tube whose cross section is constant and equal to that of the sound source. Now suppose the source is circular with radius a, and on its surface the pressure is $p_{01}\sin\omega_1 t + p_{02}\sin\omega_2 t$. The quasilinear solution for the complex difference frequency pressure p_- in the farfield of the nonlinear interaction region is then [1,81]

$$p_- = -\frac{j\beta k_-^2 a^2 p_{01} p_{02}}{4\alpha_T \rho_0 c_0^2}\frac{e^{-jk_- r}}{r} D_W(\theta) D_A(\theta), \qquad r \gg \alpha_T^{-1}, k_- a^2/2, \tag{28}$$

where

$$D_W(\theta) = \frac{1}{1 + j2(k_-/\alpha_T)\sin^2(\theta/2)}, \qquad D_A(\theta) = \frac{2J_1(k_- a\sin\theta)}{k_- a\sin\theta}.$$

The physical pressure is given by $\mathrm{Im}(p_- e^{j\omega_- t})$, with $\chi_- = k_- - j\alpha_-$ the complex wavenumber, $k_- = \omega_-/c_0$ the real wavenumber, and α_- the attenuation coefficient at the difference frequency; $\alpha_T = \alpha_1 + \alpha_2 - \alpha_-$ is a combined attenuation coefficient. If $r_0 = k_0 a^2/2$ is defined to be the collimation length (or Rayleigh distance) of the primary waves, where k_0 is the average primary wavenumber, then $\alpha_T r_0 > 1$ is a suitable criterion for the validity of Eq. (28). An additional restriction on Eq. (28) is that the primary waves suffer only relatively small finite amplitude losses.

The most remarkable property of the parametric array is its beam pattern. When $k_- a < 1$, as is often the case, the aperture factor $D_A(\theta)$ is approximately equal to unity and may be ignored [81]. The angular distribution of difference frequency radiation is then governed by the Westervelt directivity function $D_W(\theta)$. Note that $D_W(\theta)$ does not predict the appearance of any sidelobes. The reason for the absence of sidelobes is that p_- is not radiated from the physical source of radius a, but instead it is generated from the volume of fluid occupied by the primary waves. The volume where the nonlinear interaction occurs extends out from the circular source to a range given approximately by $L_a = \alpha_T^{-1}$. The absorption length L_a, and not the source radius a, is the characteristic dimension of the parametric array. In addition, the parametric array exhibits an exponential amplitude taper, and it is phased in such a way as to create an end-fire configuration. The difference frequency radiation from this exponentially tapered end-fire array forms a beam pattern devoid of a sidelobe structure. From the Westervelt directivity function, the half power angle of the radiation is found to be $\theta_{HP} = \sqrt{2\alpha_T/k_-}$. Since the beamwidth is independent of the physical source dimension a, the parametric array is capable of generating narrow low frequency radiation fields from relatively small ultrasonic sources. In fact, direct radiation at the difference frequency from a circular source is described

by the aperture factor $D_A(\theta)$, which typically predicts a much broader beam than does $D_W(\theta)$. Finally, the parametric array possesses unusual bandwidth characteristics. Since $\omega_1 \approx \omega_2$, a small percentage change in either of the primary frequencies causes a relatively large percentage change in the difference frequency. Although a narrowband transducer may generate the primary waves, a relatively wide frequency band exists at the downshifted difference frequency.

The model conceived by Westervelt is often referred to as an absorption limited array, because absorption is assumed to be the only loss mechanism responsible for terminating the nonlinear interaction region. However, when the absorption length is large compared to the Rayleigh distance ($\alpha_T r_0 \ll 1$), spherical spreading can be the principal mechanism that limits the length of the interaction region. The latter type of parametric interaction is said to be diffraction limited. Taking into account nonlinear interaction in the spherical wave region leads to different conclusions regarding sum and difference frequency generation in an absorbing fluid [82]. Sufficiently far from the parametric array, the difference frequency wave is governed by linear theory and decays as $r^{-1}e^{-\alpha - r}$. On the other hand, the sum frequency wave depends asymptotically on the product of the primary waves and decays as $r^{-2}e^{-(\alpha_1+\alpha_2)r}$ (see also Ref. 83). In effect, the sum frequency wave never truly escapes from the nonlinear interaction region. Difference frequency beamwidths measured independently by researchers working with different parameter ranges have been compiled and analyzed by Hobæk and Tjøtta [79] (see also Ref. 84). It was found that for diffraction limited arrays, the half power angle of the difference frequency wave is close to that of the primary waves.

Results from a parametric array experiment performed in water by Muir and Willette [85] are shown in Figs. 8 and 9. Primary waves at 418 and 482 kHz were radiated by a single 3.8 cm radius circular transducer, and sound pressure levels at the sum (900 kHz), difference (64 kHz), and primary frequencies were measured as functions of range and angle. Shown in Fig. 8 are the axial propagation curves. The collimated nearfield of the primary waves ends just beyond

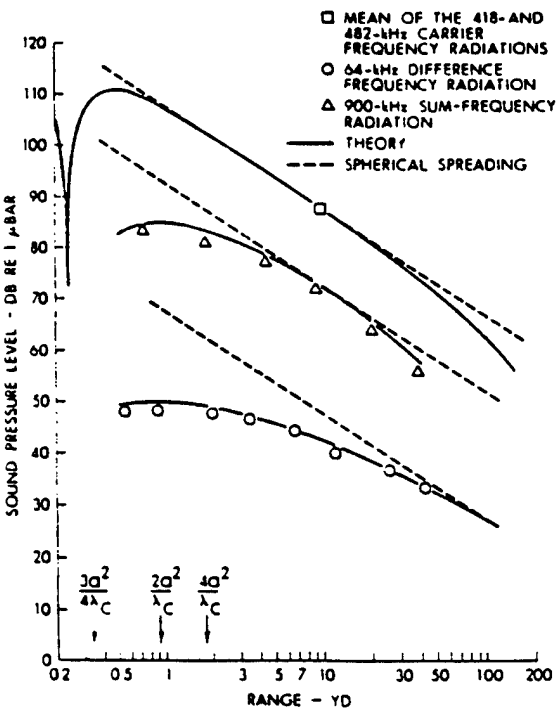


Fig. 8—Propagation curves for a parametric array.

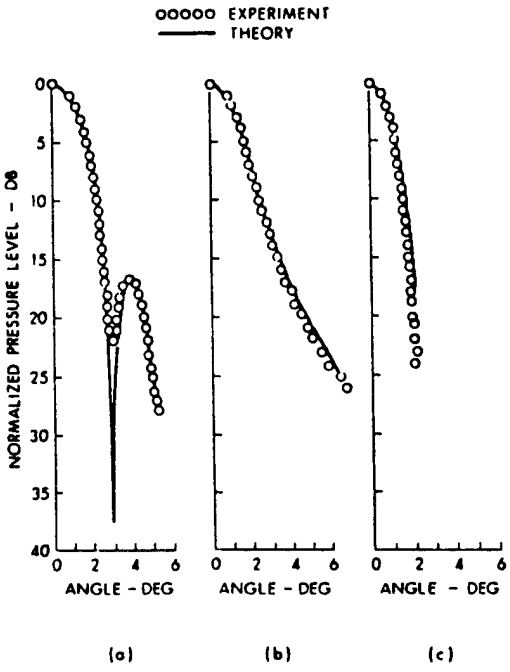


Fig. 9—Beam patterns for the (a) 482 kHz primary, (b) 64 kHz difference frequency, and (c) 900 kHz sum frequency waves.

1 yd from the source, as seen by where the propagation curve begins to follow the spherical spreading law $p \propto r^{-1}$. The primary waves eventually decay more rapidly than r^{-1} on account of absorption. Note that the difference frequency wave does not experience spherical spreading until it has traveled 50–100 yd. Significant nonlinear interaction therefore occurs well into the farfield of the primary waves. The sum frequency wave falls off much sooner because of the much higher absorption at that frequency. As pointed out above, the difference frequency wave is eventually the lone survivor, despite the fact that it starts out more than 30 dB below the sum frequency wave. Shown in Fig. 9 are the corresponding farfield beam patterns. Between the -10 dB points, the beam pattern of the upper primary wave and that of the difference frequency wave are very similar, despite their frequencies differing by a factor of 7.5. Note also the absence of sidelobes in the difference frequency beam pattern.

By far the most widespread use of parametric arrays has been in underwater applications, many of which are discussed by Novikov, Rudenko, and Timoshenko [32]. The absence of sidelobes makes the parametric array particularly useful in highly reverberant environments created by shallow water. For example, fish detection and location in shallow water is sometimes more easily performed with a parametric array then with conventional sonar. Parametric arrays have been used in tank experiments where the propagation of single modes in water channels is investigated. The narrow, low frequency beams have also proved useful in bottom and sub-bottom profiling of the ocean floor. Although only a small percentage of the energy in the primary waves is transferred to the difference frequency component, signal processing advantages are gained by exploiting the wide bandwidth properties of the parametric array. Moreover, the beam pattern of the parametric array is relatively unaffected by changes in frequency.

VIII. SELF-DEMODULATION

When pulses rather than continuous signals are used to drive the parametric array, the time signal received in the farfield is dramatically different from what is supplied to the source. For the time waveform at the source let

$$p(x=0) = p_0 f(t) \sin \omega t, \tag{29}$$

where $f(t)$ is an envelope function that is normalized to have a maximum value of unity. It is assumed that the envelope varies slowly with time compared to $\sin \omega t$. Berktay [27] showed that the farfield pressure on the axis of the parametric array is (see also Ref. 86)

$$p(\theta=0) = \frac{\beta p_0^2 S}{16\pi\rho_0 c_0^4 \alpha r} \frac{\partial^2}{\partial t^2} f^2\left(t - \frac{r}{c_0}\right), \qquad r \gg \alpha^{-1}, S/\lambda, \tag{30}$$

where α is the absorption coefficient of the carrier, S is the area of the source, and λ is the characteristic wavelength of the farfield pressure p. The carrier wave is of much higher frequency than the envelope function and is therefore more rapidly absorbed by the fluid. Eventually all that is left on axis is a squared and differentiated version of the original envelope function. The process is referred to as self-demodulation.

Graphic experimental verification of the self-demodulation phenomenon appeared in a series of papers by Moffett and coworkers [86–88]. Shown in Fig. 10 [86] are results from an experiment performed with 10 MHz pulses in carbon tetrachloride. Note the simultaneous absorption of the carrier and distortion of the envelope. The experimental results are in excellent agreement with Eq. (30) [87]. Research currently underway in Japan [89,90] is directed toward the use of parametric arrays in air for the transmission of audio signals in narrow sound beams. Problems associated with construction of such a parametric loudspeaker center around power requirements and suitable electrical predistortion networks.

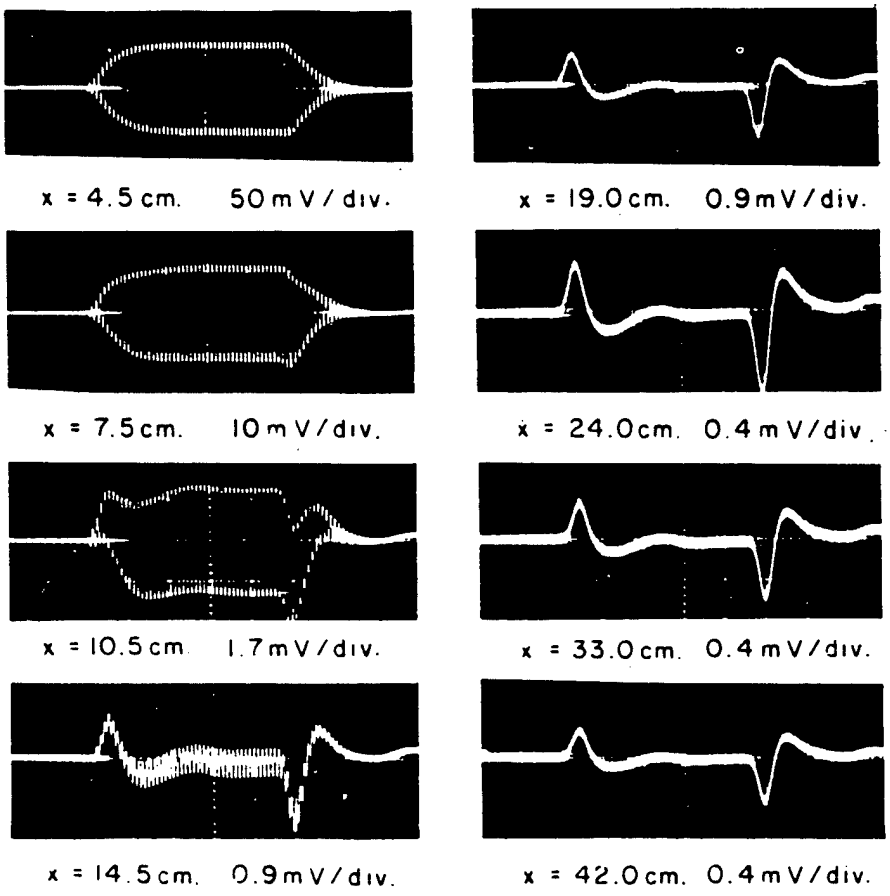


Fig. 10—Self-demodulation of a pulsed 10 MHz carrier wave in carbon tetrachloride.

IX. PARAMETRIC RECEIVERS

Westervelt [1] pointed out that highly directional receivers as well as transmitters can be formed by the interaction of finite amplitude sound waves. A parametric receiver consists of only two transducers (see Fig. 11 [91]). One is a source, referred to as the pump, capable of radiating a collimated, high intensity sound beam. The other is a hydrophone located along the acoustic axis of the pump. Because of its high intensity, the pump wave interacts nonlinearly with, and thus modulates, any acoustic signal that propagates through its path. The resulting intermodulation components detected by the hydrophone provide information about the incoming signal. When the pump wave and incoming signal are each a pure tone of different frequency, the intermodulation components of interest are the sum and difference frequency signals. The amplitudes of the sum and difference frequency signals received by the hydrophone are maximized when the parametric receiver is aligned

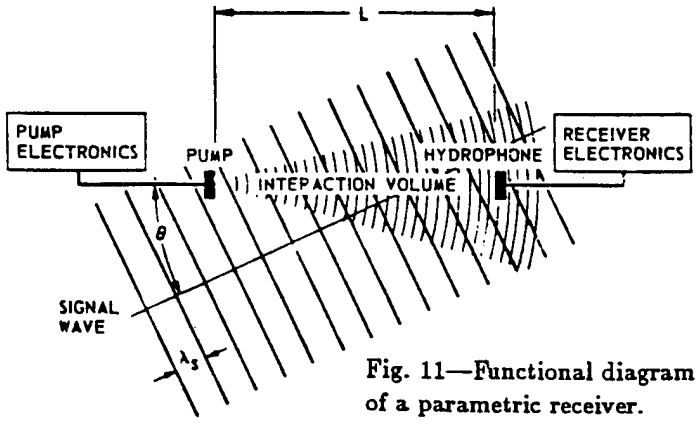


Fig. 11—Functional diagram of a parametric receiver.

with the propagation direction of the incoming signal. Departure from such a collinear orientation causes the amplitudes to vary according to the farfield directivity function of an end-fire array having the same length as the distance L between the pump and hydrophone. If the pump generates a collimated plane wave of frequency f_p and the incoming signal is a plane wave of frequency f_s, and if $f_p \gg f_s$, $k_s a < 1$, $\alpha_p^{-1} \gg L$, where a is the pump radius and α_p is the attenuation coefficient of the pump wave, then the dependence of the received sum and difference frequency pressure amplitudes on the angle of the incoming wave is [92,93]

$$D(\theta) = \frac{\sin\left[k_s L \sin^2(\theta/2)\right]}{k_s L \sin^2(\theta/2)}. \tag{31}$$

The differential elements that form a continuous end-fire array are thus synthesized by the nonlinear response of the fluid to the pump wave.

The first experiments with parametric receivers were reported around 1969 in England by Berktay and Al-Temini [92] and in the Soviet Union by Zverev and Kalachev [93]. Subsequent work improved on the basic theoretical model by allowing for spherically spreading [94] and diffracting [95] pump waves, but the directivity function was always found to closely resemble Eq. (31). Truchard [96,97] investigated both theoretically and experimentally the effect of pump geometry on parametric reception. The novel concept of an array of parametric receivers was considered by Berktay and Muir [98]. Use of more than one receiver not only allows more directive reception but in addition provides a means for beam steering. Goldsberry and coworkers [91,99] considered some of the engineering problems associated with parametric reception.

The utility of the parametric receiver lies in the advantage of being able to construct large apertures for directional reception with only two relatively small transducers. Since the half power angle is $\theta_{HP} = 0.94\sqrt{\lambda_s/L}$, the beamwidth is determined primarily by the pump to hydrophone separation. An upper bound on the length of a parametric receiver is determined by the maximum range at which the intensity of the pump wave is capable of maintaining sufficiently strong nonlinear interaction with the noise signal. Maximum receiver lengths are typically on the order of hundreds of meters, although design parameters for receivers tens of kilometers in length have been discussed by McDonough [100]. A particularly useful application of the parametric receiver is the directional measurement of low frequency noise in the ocean. The construction of conventional low frequency receivers that are sufficiently large to provide high directivity can be impractical for either financial or logistical reasons.

X. DIFFRACTION

The combined effects of nonlinearity, dissipation, and diffraction of directive sound beams are accounted for consistently to lowest order by the KZK (Khokhlov-Zabolotskaya-Kuznetsov) equation [33,34] (see also Ref. 36)

$$\frac{\partial^2 p}{\partial z \partial \tau} - \frac{c_0}{2}\nabla_\perp^2 p - \frac{b}{2\rho_0 c_0^3}\frac{\partial^3 p}{\partial \tau^3} = \frac{\beta}{2\rho_0 c_0^3}\frac{\partial^2 p^2}{\partial \tau^2}. \tag{32}$$

In the derivation of Eq. (32) it is assumed that the sound waves are sufficiently well collimated ($ka \gg 1$) that the wavefronts are quasiplanar. A coordinate system is used where the z axis coincides with the acoustic axis of the sound beam, and the retarded time associated with the propagating quasiplane waves is $\tau = t - z/c_0$. The transverse laplacian operator $\nabla_\perp^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$ accounts for the wavefront curvature that results from diffraction. Equation (32) may be written in terms of particle velocity by using the linear plane wave impedance relation $p = \rho_0 c_0 u$. The substitution is valid because (1) $ka \gg 1$ [36], and (2) the local effects discussed in Sec. I are not accounted for in Eq. (32). Without the diffraction term $\nabla_\perp^2 p$, Eq. (32) thus reduces to the form of the Burgers equation in Eq. (16).

The first investigations based on the KZK equation were performed almost exclusively in the Soviet Union. Much of the Soviet work on nonlinear effects in sound beams is reviewed in books by a number of the leading researchers [32,35,44]. Early work on second harmonic generation [101] and parametric radiation [102] was often based on Gaussian beam theory because of the resulting analytical simplifications in quasilinear analyses (see Sec. XI). Numerical solutions presented in a series of papers by Bakhvalov and coworkers [103–109] and reviewed in Ref. [35] demonstrated the significance of diffraction in nonlinear waveform distortion. However, apart from the quasilinear analysis by Kunitsyn and Rudenko [110], the intricate diffraction effects in the practical case of finite amplitude radiation from a circular source were all but ignored.

The most detailed analysis of the combined effects of nonlinearity, diffraction, and absorption in directive sound beams appeared in a series of papers by Naze Tjøtta, Tjøtta, and coworkers [36–42]. Their theoretical work is based on the KZK equation, and it is in excellent agreement with experimental results. Although their analysis applies to arbitrary sources, here we shall restrict attention to single frequency excitation of a uniform circular source of radius a. Specifically, let

$$\begin{aligned} p(z=0) &= p_0 \sin \omega t, \qquad r \le a, \\ &= 0, \qquad r > a, \end{aligned} \tag{33}$$

where $r = \sqrt{x^2 + y^2}$ is the radial coordinate in the plane that is perpendicular to the acoustic axis. Since the relation $p = \rho_0 c_0 u$ is valid for Eq. (32), Eq. (33) is the correct boundary condition for a circular source that vibrates with uniform velocity. The basic analytical models for the radiated sound field are derived via the method of successive approximations.

As a first approximation, it is assumed that the source amplitude is sufficiently low that linear theory provides an adequate description of the wave at frequency ω. This approximation holds as long as nonlinear losses can be ignored. The right hand side of Eq. (32) is then set equal to zero, and the solution that satisfies Eq. (33) is obtained. The linear, or first-order solution p_1 is found using integral transforms to be [37,111]

$$p_1 = p_0 e^{-\alpha z}\left(1 - e^{-jz_0/z} + 2e^{-jz_0/z}\int_0^{z_0 r/az} e^{-jzu^2/z_0} J_1(2u)du\right), \tag{34}$$

where $z_0 = ka^2/2$ is the Rayleigh distance, $k = \omega/c_0$, $\alpha = b\omega^2/2\rho_0 c_0^3$, and the physical pressure is $\mathrm{Im}(p_1 e^{j\omega t})$. Equation (34) is written in a form that is particularly convenient for numerical computations. On axis ($r = 0$), the integral vanishes and the solution reduces to $p_1 = p_0 e^{-\alpha z}(1 - e^{-jz_0/z})$. In the absence of absorption the axial amplitude is proportional to $\sin(z_0/2z)$, which for $z/a > 0.5(ka)^{1/3}$ [37,110,111] is in good agreement with the exact axial solution given by King [112]. In other words, Eq. (34) breaks down on axis only within a few radii from the source. Off axis, Eq. (34) is valid within the paraxial region, away from the edges of the beam. For example, the farfield directivity function

$$D(\theta) = \frac{2J_1(ka\tan\theta)}{ka\tan\theta} \tag{35}$$

obtained from Eq. (34) is the same as that which is derived in the Fresnel approximation. In summary, Eq. (34) provides an accurate description of those regions of the sound field where most of the energy is located.

The second approximation consists of adding to the linear solution the quasilinear solution obtained from Eq. (32) when $p = \mathrm{Im}(p_1 e^{j\omega t})$ is substituted in the nonlinear term on the right hand side. The resulting inhomogeneous wave equation has a forcing function at frequency 2ω. According to Eq. (33) there is no second harmonic radiation by the source, so the quasilinear solution p_2 must vanish at $z = 0$. The general solution for p_2 involves a quadruple integral that

must be evaluated numerically [40]. However, a relatively simple farfield result is obtained for the region beyond the Rayleigh distance, but before the range where absorption is important:

$$p_2 \sim -\frac{\beta p_0^2 k^3 a^4}{8\rho_0 c_0^2}\frac{e^{-jkr^2/z}}{z}\left[\ln\left(\frac{2z}{z_0}\right) - j\frac{\pi}{2} - \gamma\right]D^2(\theta) + O(z^{-1}), \qquad z_0/2 \ll z \ll 1/2\alpha. \quad (36)$$

The physical pressure is $\mathrm{Im}(p_2 e^{j2\omega\tau})$, and $\gamma = 0.577\ldots$ is the Euler number.

Analysis of Eq. (36) reveals that the farfield of the second harmonic component begins at a significantly greater range than that of the fundamental component. The farfield is defined as the region where the beam pattern no longer depends on range. Although the first term in Eq. (36) may be separated into independent functions of z and θ, the second term represented by $O(z^{-1})$ may not. The farfield of the second harmonic component begins when the first term in Eq. (36) dominates the second. When the first term is of the same order as the second, the beam pattern is still a function of range. The second term accounts for nonlinear effects generated very near the source, as indicated by the z^{-1} decay rate. Cumulative effects are represented by the first term, which decays at the slower rate of $z^{-1}\ln z$ because energy is continuously pumped into the second harmonic wave from the fundamental component. Nearfield effects are dominated by cumulative effects only when the slowly varying logarithm in the first term becomes large compared to unity. The *nonlinear* farfield is therefore reached only when z exceeds z_0 by several orders of magnitude, in contrast to the usual *linear* farfield criterion $z \gg z_0$. Only in the nonlinear farfield does the product directivity $D^2(\theta)$ apply to the second harmonic, as was derived previously (see Eq. (24)) by ignoring diffraction effects. Similar results apply to all higher harmonics [42].

Numerical solutions of Eq. (32) for the full set of nonlinearly generated harmonics may be obtained by a method similar to that described in Sec. VI. Such an approach was developed by Aanonsen [113] (see also Refs. 41 and 42). A Fourier expansion where the spectral amplitudes are functions of x, y, and z is substituted in Eq. (32). The resulting set of coupled equations is similar to Eq. (27), except now the differential operators on the left hand side are of the diffusion type. Solutions can then be obtained by standard finite difference methods. Finite difference solutions of the KZK equation were used to generate the theory shown in Fig. 5.

Beam patterns calculated numerically for the fundamental through the fourth harmonic components that result from the boundary condition in Eq. (33) are shown in Fig. 12 [42]. The range is fixed at $z = 10z_0$, for which the dimensionless coordinate along the abscissa is $u' = 0.45ka\tan\theta$. The source amplitude is characterized by $\beta\epsilon kz_0 = 1.5$, where $\epsilon = p_0/\rho_0 c_0^2$. A plane wave having the same initial amplitude would form a shock at $0.67z_0$, so the sound beam is strongly nonlinear. The three sets of beam patterns correspond to different amounts of absorption: (*a*) $\alpha z_0 = 1.0$, (*b*) $\alpha z_0 = 0.1$, and (*c*) $\alpha z_0 = 0.01$. Consider first the case with high absorption ($\alpha z_0 = 1.0$). In this case the absorption prevents any significant nonlinear losses at the fundamental frequency. The beam pattern of the fundamental component ($n = 1$) in Fig. 12*a* is described by Eq. (35). Absorption has also eliminated the nonlinear nearfield effects, and the directivity functions of the higher harmonics are given by Eq. (24). Figure 12*a* therefore depicts a situation where the nonlinear farfield is reached within ten Rayleigh distances from the source, and there is agreement with the conventional farfield criterion $z \gg z_0$ for linear radiation. Now consider the case of low absorption in Fig. 12*c*, where $\alpha z_0 = 0.01$. Because of nonlinear losses, the beam pattern of the fundamental component is no longer described by Eq. (35). Since nonlinear losses are most significant where the amplitude of a wave is highest, in a directive sound beam the losses are most pronounced on axis. Note that the main lobe is considerably flatter in Fig. 12*c* than in Fig. 12*a*. Nonlinear losses are manifested as an erosion of the main lobe, which corresponds to an increase in the relative level of the sidelobes. Whereas the first sidelobe of the fundamental component is down by almost 18 dB in Fig. 12*a*, the same lobe is down by only 11 dB in Fig. 12*c* (see Ref. 49 for related experimental work). However, the most striking feature of Fig. 12*c* is the

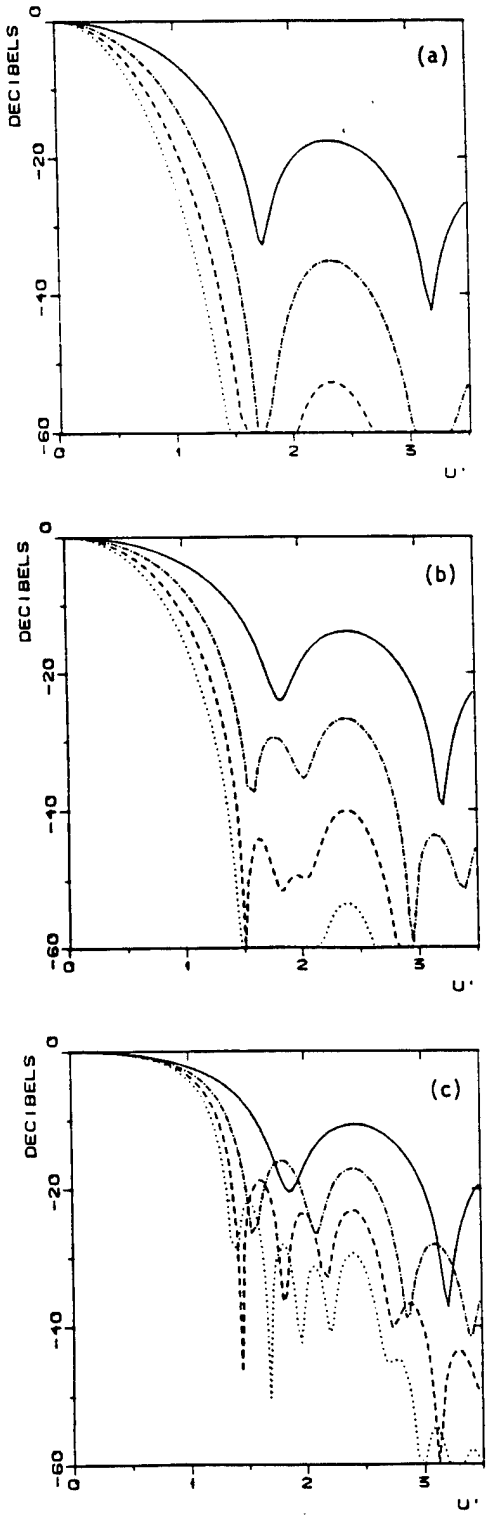


Fig. 12—Beam patterns calculated at $z = 10z_0$ for $n = 1$ (—), $n = 2$ (– –), $n = 3$ (– – –), and $n = 4$ (· · ·), where $u' = 0.45ka\tan\theta$, and (a) $\alpha z_0 = 1.0$, (b) $\alpha z_0 = 0.1$, (c) $\alpha z_0 = 0.01$.

appearance of nearfield effects. In the second harmonic beam pattern there are twice as many sidelobes as in Fig. 12*a*, three times as many for the third harmonic, and so on. The additional sidelobes, referred to as *fingers* [40], are higher in amplitude than the sidelobes predicted by the power law for the directivity pattern. The fingers are the nearfield effects represented by $O(z^{-1})$ in Eq. (36). Not until the fingers disappear is the nonlinear farfield reached, which for low absorption may not occur for hundreds of Rayleigh distances. Figure 12*b* shows a case of medium absorption, where at $z = 10z_0$ the fingers still exist but do not dominate the field structure.

Fingers have been observed in experimental investigations (see Refs. 40 and 42), but in the past they have been attributed to direct radiation from the source at the various harmonic frequencies. Such a conclusion is not suprising, since the fingers appear precisely where one would expect to find sidelobes if the source were radiating directly at that particular frequency. Shown in Fig. 13 [42] are experimental results obtained by Lockwood [114] (see also Ref. 70). The fundamental, second, and third harmonic beam patterns were measured 10 m ($z = 7.1z_0$) away from a circular source of radius 3.8 cm driven at 450 kHz in water. The dimensionless source amplitude was $\beta\epsilon kz_0 = 0.88$, just over half that modeled in Fig. 12, and the dimensionless absorption was $\alpha z_0 = 0.01$. If the measurements were taken in the farfield, Eq. (24) would be valid, and sidelobes would appear only at $\theta = 4°$. However, not only are nearfield effects quite pronounced, but agreement between experiment and theory based on Eq. (32) is very good.

The use of finite amplitude sound for acoustic imaging requires a basic understanding of the relative effects of nonlinearity, diffraction, and absorption in directive radiation. Nonlinear losses at the source frequency can result in wider beams and lower received levels than are predicted by linear theory. When an imaging system uses information from nonlinearly generated waves, it is important to take into account the sensitive dependence of diffraction effects on the geometry and amplitude of the source and the absorption of the medium. The theoretical model discussed in this section has played an important role in advancing the understanding of parametric arrays [38], nonlinear focusing systems [115–117], and the transmission [118] and reflection [119] of finite amplitude sound beams incident on an interface between different media.

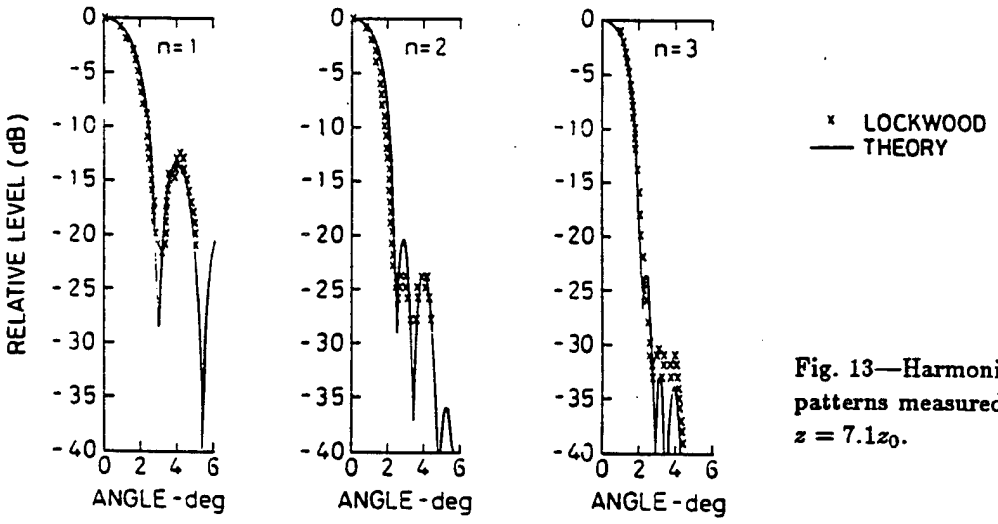


Fig. 13—Harmonic beam patterns measured at $z = 7.1z_0$.

XI. GAUSSIAN BEAMS

Gaussian beams are often used to model directive radiation when the basic phenomena to be investigated do not require the complexity of the analysis outlined in Sec. X. The extensive research based on Gaussian beams in the Soviet Union is reviewed in Refs. 32, 35 and 44. When appropriately scaled, the results from Gaussian beam theory often compare favorably with experiments that involve radiation from circular sources. For the Gaussian source condition let

$$p(z=0) = p_g e^{-(r/r_0)^2} \sin \omega t, \tag{37}$$

where r is again the transverse radial coordinate, and r_0 is the spot size of the source excitation. As in Sec. X, the method of successive approximations is used to derive expressions for the fundamental and second harmonic components. Ignoring the nonlinear term in Eq. (32), we obtain for the fundamental component

$$p_1 = \frac{p_g}{1 - jz/z_0} \exp\left(-\alpha z - \frac{(r/r_0)^2}{1 - jz/z_0}\right), \tag{38}$$

where $z_0 = kr_0^2/2$ is the collimation length of the source, and the physical pressure is $\mathrm{Im}(p_1 e^{j\omega t})$. In the nearfield ($z \ll z_0$) the factors of z/z_0 may be ignored and Eq. (38) describes a plane wave, with Gaussian amplitude shading, that propagates in an absorbing fluid. None of the oscillatory nearfield structure that characterizes radiation from circular sources appears in Gaussian beams. Note that the transverse amplitude distribution is Gaussian at all ranges. In the farfield, for example, the directivity function is $D(\theta) = \exp[-(\frac{1}{2}kr_0 \tan\theta)^2]$.

The second harmonic component is calculated by solving Eq. (32) with the linear solution substituted in the nonlinear term [83]:

$$p_2 = \frac{j\beta p_g^2 k^2 r_0^2}{4\rho_0 c_0^2 (1 - jz/z_0)} \exp\left(-4\alpha z - \frac{2(r/r_0)^2}{1 - jz/z_0} - j2\alpha z_0\right)\left(E_1[-j2\alpha] - E_1[-j2\alpha(1 - jz/z_0)]\right), \tag{39}$$

where $E_1(z) = \int_z^\infty t^{-1} e^{-t} dt$ is the exponential integral, which is easily evaluated with various series expansions [120,121]. The physical pressure of the second harmonic component is given by $\mathrm{Im}(p_2 e^{j2\omega t})$. First consider the limiting case of no absorption, for which a particularly simple result is obtained. As $\alpha \to 0$, the two exponential integrals in Eq.(39) combine to yield $\ln(1 - jz/z_0)$. In the nearfield $p_2 \propto z$, which agrees with the result for plane waves in the quasilinear approximation. In the farfield, where z exceeds z_0 by several orders of magnitude, the

dependence becomes $p_2 \propto z^{-1} \ln z$, the same as was found in Sec. X for radiation from a circular source. When absorption is taken into account, the farfield solution yields $p_2 \propto z^{-2} e^{-2\alpha z}$, and the directivity function (whether or not there is absorption) is $D_2(\theta) = D^2(\theta)$. Since the farfield dependence of the linear solution is $p_1 \propto z^{-1} e^{-\alpha z} D(\theta)$, one obtains in the nonlinear farfield $p_2 \propto p_1^2$. The farfield behavior of the second harmonic component in an absorbing fluid thus depends only on the local behavior of the fundamental component [82,83,122].

Fenlon and coworkers [83,123,124] have employed a transformation that modifies Gaussian beam solutions for application to sound radiated from circular sources. A farfield matching technique is used that equates Eq. (38) near the acoustic axis to the linear solution that satisfies the source condition in Eq. (33). The matching procedure yields $p_g = 2p_0$ and $r_0 = a/\sqrt{2}$ for the Gaussian beam parameters. Use of the transformation in Eq. (39) yields a solution for the second harmonic component which for $\alpha = 0$ matches Eq. (36) in the paraxial farfield (note that z_0 is defined differently in the two equations). Modified Gaussian beam solutions have been compared with results for parametric arrays generated by circular sources [39,83,123]. The solutions are shown to be accurate on axis beyond about one half Rayleigh distance from the source. In fact, modified Gaussian beam solutions for both the sum and difference frequency waves generated with a parametric array were shown to be virtually indistinguishable from the theory presented in Fig. 8 [83].

Equations (38) and (39) are also easily modified for the study of focused finite amplitude sound. A source with focal length d is modeled by multiplying the complex source condition by the phase term $e^{jkr^2/2d}$. Equivalently, one could just replace the spot size r_0 by the modified complex spot size $\tilde{r}_0 = r_0(1 - jz_0/d)^{-1/2}$. Substitution of $\tilde{r}_0$ for r_0 in Eqs. (38) and (39) thus yields results for focused sound beams. For example, making the substitution in Eq. (38) yields for the linear solution

$$p_1 = \frac{p_g}{1 - z/d - jz/z_0} \exp\left(-\alpha z - \frac{1 - jz_0/d}{1 - z/d - jz/z_0}\left(\frac{r}{r_0}\right)^2\right). \tag{40}$$

If the effects of absorption are ignored, the pressure amplitude at the focus is found to be $(kr_0^2/2d)p_g$. To model focused sound from a circular source we may again let $p_g = 2p_0$ and $r_0 = a/\sqrt{2}$. The pressure amplitude at the focus is then found to be $(ka^2/2d)p_0$, which is in agreement with the exact linear solution of Eq. (32) for a focused circular source [125]. Focused finite amplitude sound from a circular source can thus be modeled fairly accurately in the paraxial focal region with Gaussian beam theory.

ACKNOWLEDGMENTS

The author would like to thank D. T. Blackstock, J. Naze Tjøtta, and S. Tjøtta for helpful discussions during the preparation of this review. Support from the Office of Naval Research is also gratefully acknowledged.

REFERENCES

[1] P. J. Westervelt, "Parametric end-fire array," *J. Acoust. Soc. Am.* **32**, 934–935 (A) (1960); and "Parametric acoustic array," *J. Acoust. Soc. Am.* **35**, 535–537 (1963).

[2] N. Ichida, T. Sato, H. Miwa, and K. Murakami, "Real-time nonlinear parameter tomography using impulsive pumping waves," *IEEE Trans. Sonics and Ultrasonics* **SU-31**, 635–641 (1984).

[3] D. Rugar, "Resolution beyond the diffraction limit in the acoustic microscope: A nonlinear effect," *J. Appl. Phys.* **56**, 1338–1346 (1984).

[4] P. H. Rogers, "Weak-shock solution for underwater explosive shock waves," *J. Acoust. Soc. Am.* **62**, 1412–1419 (1977).

[5] D. A. Webster and D. T. Blackstock, "Experimental investigation of outdoor propagation of finite-amplitude noise," NASA Contractor Rep. 2992, Langley Research Center (1978).

[6] D. T. Blackstock, "History of nonlinear acoustics and a survey of Burgers' and related equations," in *Nonlinear Acoustics*, Proceedings of the Symposium held at Applied Research Laboratories, The University of Texas at Austin, 1969, edited by T. G. Muir (AD 719 936), pp. 1–27.

[7] L. Euler, *1759 Mém. Acad. Sci. Berlin* **15**, 185–209 (1766).

[8] J. Lagrange, "New researches on the nature and propagation of sound," *Misc. Taur.* **2**, 11-172 (1760–61).

[9] S. D. Poisson "Memoir on the theory of sound," *J. L'Ecole Polytech.* **7**, 364–370 (1808).

[10] G. G. Stokes, "On a difficulty in the theory of sound," *Philos. Mag.*, . Ser. 3, **33**, 349–356 (1848).

[11] B. Riemann, "The propagation of sound waves of finite amplitude," *Abhandl. Ges. Wiss., Göttingen, Math-Physik-Kl.* **8**, 43–65 (1858–59).

[12] S. Earnshaw, "On the mathematical theory of sound," *Philos. Trans. R. Soc. Lond.* **150**, 133–148 (1860).

[13] R. T. Beyer, *Nonlinear Acoustics in Fluids*, Benchmark Papers in Acoustics, Vol. 18 (Van Nostrand Reinhold Co., New York, 1984).

[14] E. Fubini, "Anomalies in the propagation of an acoustic wave of large amplitude," *Alta. Freq.* **4**, 173–180 (1935).

[15] R. D. Fay, "Plane sound waves of finite amplitude," *J. Acoust. Soc. Am.* **3**, 222–241 (1931).

[16] A. L. Thuras, R. T. Jenkins, and H. T. O'Neil, "Extraneous frequencies generated in air carrying intense sound waves," *J. Acoust. Soc. Am.* **6**, 173–180 (1935).

[17] C. Eckart, "Vortices and streams caused by sound waves," *Phys. Rev.* **73**, 68–76 (1948).

[18] M. J. Lighthill, "On sound generated aerodynamically," *Proc. Roy. Soc.* **A222**, 564–587 (1952).

[19] J. S. Mendousse, "Nonlinear dissipative distortion of progressive sound waves at moderate amplitudes," *J. Acoust. Soc. Am.* **25**, 51–54 (1953).

[20] K. A. Naugol'nykh, S. I. Soluyan, and R. V. Khokhlov, "Spherical waves of finite amplitude in a viscous thermally conducting medium," *Sov. Phys. Acoust.* **9**, 42–46 (1963).

[21] R. V. Khokhlov, K. A. Naugol'nykh, and S. I. Soluyan, "Waves of moderate amplitude in absorbing media," *Acustica* **14**, 248–253 (1964).

[22] D. T. Blackstock, "Thermoviscous attenuation of plane, periodic, finite-amplitude sound waves," *J. Acoust. Soc. Am.* **36**, 534–542 (1964).

[23] P. J. Westervelt, "Scattering of sound by sound," *J. Acoust. Soc. Am.* **29**, 199–203 (1957).

[24] P. J. Westervelt, "Scattering of sound by sound," *J. Acoust. Soc. Am.* **29**, 934–935 (1957).

[25] J. L. S. Bellin and R. T. Beyer, "Experimental investigation of an end-fire array," *J. Acoust. Soc. Am.* **32**, 935 (A) (1960); and *J. Acoust. Soc. Am.* **34**, 1051–1054 (1962).

[26] T. G. Muir, "Nonlinear acoustics: A new dimension in underwater sound," Tech. Rep. ARL-TP-76-25, Applied Research Laboratories, The University of Texas at Austin (1976).

[27] H. O. Berktay, "Possible exploitation of non-linear acoustics in underwater transmitting applications," *J. Sound Vib.* **2**, 435–461 (1965).

[28] H. O. Berktay, "Parametric amplification by the use of acoustic non-linearities and some possible applications," *J. Sound Vib.* **2**, 462–470 (1965).

[29] H. O. Berktay, "Proposals for underwater transmitting applications of non-linear acoustics," *J. Sound Vib.* **6**, 244–254 (1967).

[30] S. Tjøtta, "Some non-linear effects in sound fields," *J. Sound Vib.* **6**, 255–267 (1967).

[31] V. A. Zverev, A. I. Kalachev, and N. S. Stepanov, "Utilization of nonlinear effects in underwater acoustics," *Sov. Phys. Acoust.* **13**, 324–326 (1968).

[32] B. K. Novikov, O. V. Rudenko, and V. I. Timoshenko, *Nonlinear Underwater Acoustics* (Sudostroenie, Leningrad, 1981); (in Russian, to appear in English through the Acoustical Society of America).

[33] E. A. Zabolotskaya and R. V. Khokhlov, "Quasi-plane waves in the nonlinear acoustics of confined beams," *Sov. Phys. Acoust.* **15**, 35–40 (1969).

[34] V. P. Kuznetsov, "Equations of nonlinear acoustics," *Sqv. Phys. Acoust.* **16**, 467-470 (1971).

[35] N. S. Bakhvalov, Ya. M. Zhileikin, and E. A. Zabolotskaya, *Nonlinear Theory of Sound Beams* (Nauka, Moscow, 1982); (in Russian, to appear in English through the American Institute of Physics).

[36] J. Naze Tjøtta and S. Tjøtta, "Nonlinear equations of acoustics, with application to parametric acoustic arrays," *J. Acoust. Soc. Am.* **69**, 1644–1652 (1981).

[37] G. S. Garrett, J. Naze Tjøtta, and S. Tjøtta, "Nearfield of a large acoustic transducer. Part I: Linear radiation," *J. Acoust. Soc. Am.* **72**, 1056–1061 (1982).

[38] G. S. Garrett, J. Naze Tjøtta, and S. Tjøtta, "Nearfield of a large acoustic transducer. Part II: Parametric radiation," *J. Acoust. Soc. Am.* **74**, 1013–1020 (1983).

[39] G. S. Garrett, J. Naze Tjøtta, and S. Tjøtta, "Nearfield of a large acoustic transducer. Part III: General results," *J. Acoust. Soc. Am.* **75**, 769–779 (1984).

[40] J. Berntsen, J. Naze Tjøtta, and S. Tjøtta, "Nearfield of a large acoustic transducer. Part IV: Second harmonic and sum frequency radiation," *J. Acoust. Soc. Am.* **75**, 1383–1391 (1984).

[41] S. I. Aanonsen, T. Barkve, J. Naze Tjøtta, and S. Tjøtta, "Distortion and harmonic generation in the nearfield of a finite amplitude sound beam," *J. Acoust. Soc. Am.* **75**, 749–768 (1984).

[42] M. F. Hamilton, J. Naze Tjøtta, and S. Tjøtta, "Nonlinear effects in the farfield of a directive sound source," *J. Acoust. Soc. Am.* **78**, 202–216 (1985).

[43] R. T. Beyer, *Nonlinear Acoustics* (Naval Sea Systems Command, Washington, DC, 1974).

[44] O. V. Rudenko and S. I. Soluyan, *Theoretical Foundations of Nonlinear Acoustics* (Plenum, New York, 1977).

[45] D. T. Blackstock, "Nonlinear acoustics (theoretical)," in *American Institute of Physics Handbook*, 3rd ed., edited by D. E. Gray (McGraw, New York, 1972), pp. 3-183 to 3-205.

[46] J. Challis, "On the velocity of sound," *Philos. Mag.*, Ser. 3, **32**, 494–499 (1848).

[47] D. T. Blackstock, "Propagation of plane sound waves in nondissipative fluids," *J. Acoust. Soc. Am.* **34**, 9–30 (1962).

[48] D. T. Blackstock, "Connection between the Fay and Fubini solutions for plane sound waves of finite amplitude," *J. Acoust. Soc. Am.* **39**, 1019–1026 (1966).

[49] J. A. Shooter, T. G. Muir, and D. T. Blackstock, "Acoustic saturation of spherical waves in water," *J. Acoust. Soc. Am.* **55**, 54–62 (1974).

[50] R. T. Beyer, "Parameter of nonlinearity in fluids," *J. Acoust. Soc. Am.* **32**, 719–721 (1960).

[51] D. A. Webster and D. T. Blackstock, "Finite-amplitude saturation of plane sound waves in air," *J. Acoust. Soc. Am.* **62**, 518–523 (1977).

[52] C. H. Allen, "Finite amplitude distortion in a spherically diverging sound wave in air," Ph.D. Dissertation, Pennsylvania State University (1950).

[53] T. L. Riley, "Generation of harmonics in finite amplitude sound radiated in water by a circular piston," M.S. Thesis, The University of Texas at Austin (1983).

[54] F. H. Fenlon, "An extension of the Bessel-Fubini series for a multiple-frequency cw acoustic source of finite amplitude," *J. Acoust. Soc. Am.* **51**, 284–289 (1972).

[55] M. E. Schaffer, "The suppression of sound with sound," Tech. Rep. ARL-TR-75-64, Applied Research Laboratories, The University of Texas at Austin (1975) (ADA 023 128).

[56] M. B. Moffett, W. L. Konrad, and L. F. Carlton, "Experimental demonstration of the absorption of sound by sound in water," *J. Acoust. Soc. Am.* **63**, 1048–1051 (1978).

[57] J. A. TenCate, "Nonlinear interaction of two noncollinear sound waves in a rectangular waveguide," Tech. Rep. ARL-TR-84-16, Applied Research Laboratories, The University of Texas at Austin (1984) (ADA 144 440).

[58] D. T. Blackstock, "Spectral interactions in nonlinear acoustics—A review," in *Nonlinear Deformation Waves*, Proceedings of the IUTAM Symposium in Tallinn, USSR, 1982, edited by U. Nigul and J. Engelbrecht (Springer, Berlin, 1983), pp. 301–315.

[59] X. Gong, Z. Zhu, and G. Du, "Nonlinear interaction of a finite amplitude wave with a small-signal wave in air," *J. Nanjing University* **76**, 19–28 (1979) (in Chinese).

[60] W. L. Willshire, "The suppression of sound by sound of higher frequency," Tech. Rep. ARL-TR-77-22, Applied Research Laboratories, The University of Texas at Austin (1977) (ADA 040 008).

[61] P. J. Westervelt, "Absorption of sound by sound," *J. Acoust. Soc. Am.* **59**, 760–764 (1976).

[62] T. K. Stanton and R. T. Beyer, "The interaction of sound with noise in water," *J. Acoust. Soc. Am.* **64**, 1667–1670 (1978).

[63] T. K. Stanton and R. T. Beyer, "Interaction of sound with noise in water II," *J. Acoust. Soc. Am.* **69**, 989–992 (1981).

[64] M. J. Lighthill, *Surveys in Mechanics*, edited by G. K. Batchelor and R. M. Davies (Cambridge University Press, Cambridge, England, 1956), pp. 250–351.

[65] S. Tjøtta, "On some nonlinear effects in sound fields, with special emphasis on the generation of vorticity and the formation of streaming patterns," *Arch. Math. Naturv.* **55**, 1–68 (1959).

[66] Z. A. Gol'dberg, "On the propagation of plane waves of finite amplitude," *Sov. Phys. Acoust.* **3**, 340–347 (1957).

[67] H. M. Merklinger, "Fundamental-frequency component of a finite-amplitude plane wave," *J. Acoust. Soc. Am.* **54**, 1760–1761 (1973).

[68] H. M. Merklinger, R. H. Mellen, and M. B. Moffett, "Finite-amplitude losses in spherical sound waves," *J. Acoust. Soc. Am.* **59**, 755–759 (1976).

[69] D. T. Blackstock, "On plane, spherical, and cylindrical sound waves of finite amplitude in lossless fluids," *J. Acoust. Soc. Am.* **36**, 217–219 (1964).

[70] J. C. Lockwood, T. G. Muir, and D. T. Blackstock, "Directive harmonic generation in the radiation field of a circular piston," *J. Acoust. Soc. Am.* **53**, 1148–1153 (1973).

[71] T. G. Muir, "Nonlinear effects in acoustic imaging," in *Acoustical Imaging*, Vol. 9, edited by K. Y. Wang (Plenum, New York, 1980), pp. 93–109.

[72] P. Y. Hennion and M. La Grève, "Evaluation of the nonlinear attenuation of a finite amplitude acoustic plane wave," *Acustica* **48**, 44–49 (1981).

[73] D. H. Trivett and A. L. Van Buren, "Propagation of plane, cylindrical, and spherical finite amplitude waves," *J. Acoust. Soc. Am.* **69**, 943–949 (1981).

[74] F. H. Fenlon, "A recursive procedure for computing the nonlinear spectral interactions of progressive finite-amplitude waves in nondispersive fluids," *J. Acoust. Soc. Am.* **50**, 1299–1312 (1971).

[75] A. Korpel, "Frequency approach to nonlinear dispersive waves," *J. Acoust. Soc. Am.* **67**, 1954–1958 (1980).

[76] F. M. Pestorius, "Propagation of plane acoustic noise of finite amplitude," Tech. Rep. ARL-TR-73-23, Applied Research Laboratories, The University of Texas at Austin (1975) (AD 778 868).

[77] F. H. Fenlon, "Nonlinear acoustics," Applied Research Laboratory, The Pennsylvania State University (1976).

[78] H. Hobæk, "Parametric acoustic transmitting arrays—A survey of theories and experiments," Sci./Tech. Rep. No. 99, Department of Physics, University of Bergen, Bergen, Norway (1977).

[79] H. Hobæk and S. Tjøtta, "Theory of parametric acoustic arrays," *J. Phys.* (Paris) **40**, Colloque C8, 101–110 (1979) (Proceedings of the 8th International Symposium on Nonlinear Acoustics, Paris, 3–6 July 1978).

[80] Anonymous, *Scientific and Engineering Studies: Nonlinear Acoustics 1954 to 1983* (Naval Underwater Systems Center, New London and Newport) (available on request).

[81] J. Naze and S. Tjøtta, "Nonlinear interaction of two sound beams," *J. Acoust. Soc. Am.* **37**, 174–175 (1965).

[82] J. Naze Tjøtta and S. Tjøtta, "Nonlinear interaction of two collinear, spherically spreading sound beams," *J. Acoust. Soc. Am.* **67**, 484–490 (1980).

[83] M. F. Hamilton and F. H. Fenlon, "Parametric acoustic array formation in dispersive fluids," *J. Acoust. Soc. Am.* **76**, 1474–1492 (1984).

[84] M. Vestrheim, "A parameter representation of the parametric acoustic array," in *Finite-Amplitude Wave Effects in Fluids*, Proceedings of the 1973 Symposium, Copenhagen, edited by L. Bjørnø(IPC Science and Technology, Guildford, England, 1974), pp. 140–144.

[85] T. G. Muir and J. G. Willette, "Parametric acoustic transmitting arrays," *J. Acoust. Soc. Am.* **52**, 1481–1486 (1972).

[86] M. B. Moffett, P. J. Westervelt, and R. T. Beyer, "Large-amplitude pulse propagation—A transient effect," *J. Acoust. Soc. Am.* **47**, 1473–1474 (1970).

[87] M. B. Moffett, P. J. Westervelt, and R. T. Beyer, "Large-amplitude pulse propagation—A transient effect. II," *J. Acoust. Soc. Am.* **49**, 339–343 (1971).

[88] M. B. Moffett and P. Mello, "Parametric acoustic sources of transient signals," *J. Acoust. Soc. Am.* **66**, 1182–1187 (1979).

[89] M. Yoneyama, J. Fujimoto, Y. Kawamo, and S. Sasabe, "The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design," *J. Acoust. Soc. Am.* **73**, 1532–1536 (1983).

[90] T. Kamakura, M. Yoneyama, and K. Ikegaya, "Developments of parametric loudspeaker for practical use," in *Proceedings of the 10th International Symposium on Nonlinear Acoustics*, edited by A. Nakamura (Teikohsha, Kadoma, Japan, 1984), pp. 147–150.

[91] D. F. Rhode, T. G. Goldsberry, W. S. Olsen, and C. R. Reeves, "Band elimination processor for an experimental parametric receiving array," *J. Acoust. Soc. Am.* **66**, 484–487 (1979).

[92] H. O. Berktay and C. A. Al-Temini, "Virtual arrays for underwater reception," *J. Sound Vib.* **9**, 295–307 (1969).

[93] V. A. Zverev and A. I. Kalachev, "Modulation of sound by sound in the intersection of sound waves," *Sov. Phys. Acoust.* **16**, 204–208 (1970).

[94] H. O. Berktay and J. A. Shooter, "Parametric receivers with spherically spreading pump waves," *J. Acoust. Soc. Am.* **54**, 1056–1061 (1973).

[95] P. H. Rogers, A. L. Van Buren, A. O. Williams Jr., and J. M. Barber, "Parametric detection of low-frequency acoustic waves in the nearfield of an arbitrary directional pump transducer," *J. Acoust. Soc. Am.* **55**, 528–534 (1974).

[96] J. J. Truchard, "Parametric acoustic receiving array. I. Theory," *J. Acoust. Soc. Am.* **58**, 1141–1145 (1975).

[97] J. J. Truchard, "Parametric acoustic receiving array. II. Experiment," *J. Acoust. Soc. Am.* **58**, 1146–1150 (1975).

[98] H. O. Berktay and T. G. Muir, "Arrays of parametric receiving arrays," *J. Acoust. Soc. Am.* **53**, 1377–1383 (1973).

[99] C. R. Reeves, T. G. Goldsberry, D. F. Rhode, and V. E. Maki Jr., "Parametric acoustic receiving array response to transducer vibration," *J. Acoust. Soc. Am.* **67**, 1495–1501 (1980).

[100] R. N. McDonough, "Long-aperture parametric receiving arrays," *J. Acoust. Soc. Am.* **57**, 1150–1155 (1975).

[101] O. V. Rudenko, S. I. Soluyan, and R. V. Khokhlov, "Confinement of a quasiplane beam of periodic perturbations in a nonlinear medium," *Sov. Phys. Acoust.* **19**, 556–559 (1974).

[102] B. K. Novikov, M. S. Rybachek, and V. I. Timoshenko, "Interaction of diffracting sound beams and the theory of highly directional ultrasonic radiators," *Sov. Phys. Acoust.* **23**, 354–357 (1977).

[103] N. S. Bakhvalov, Ya. M. Zhileikin, E. A. Zabolotskaya, and R. V. Khokhlov, "Nonlinear propagation of a sound beam in a nondissipative medium," *Sov. Phys. Acoust.* **22**, 272–274 (1976).

[104] N. S. Bakhvalov, Ya. M. Zhileikin, E. A. Zabolotskaya, and R. V. Khokhlov, "Focused high-amplitude sound beams," *Sov. Phys. Acoust.* **24**, 10–15 (1978).

[105] N. S. Bakhvalov, Ya. M. Zhileikin, E. A. Zabolotskaya, and R. V. Khokhlov, "Propagation of finite-amplitude sound beams in a dissipative medium," *Sov. Phys. Acoust.* **24**, 271–275 (1978).

[106] N. S. Bakhvalov, Ya. M. Zhileikin, E. A. Zabolotskaya, and R. V. Khokhlov, "Harmonic generation in sound beams," *Sov. Phys. Acoust.* **25**, 101–106 (1979).

[107] N. S. Bakhvalov, Ya. M. Zhileikin, and E. A. Zabolotskaya, "Parametric interaction of sound beams," *Sov. Phys. Acoust.* **25**, 280–283 (1979).

[108] N. S. Bakhvalov, Ya. M. Zhileikin, and E. A. Zabolotskaya, "Nonlinear propagation of Gaussian beams," *Sov. Phys. Acoust.* **25**, 458–460 (1979).

[109] N. S. Bakhvalov, Ya. M. Zhileikin, and E. A. Zabolotskaya, "Nonlinear propagation of sound beams with a uniform amplitude distribution," *Sov. Phys. Acoust.* **26**, 95–100 (1980).

[110] V. E. Kunitsyn and O. V. Rudenko, "Second-harmonic generation in the field of a piston radiator," *Sov. Phys. Acoust.* **24**, 310–313 (1978).

[111] J. Naze Tjøtta and S. Tjøtta, "An analytical model for the nearfield of a baffled piston transducer," *J. Acoust. Soc. Am.* **68**, 334–339 (1980).

[112] L. V. King, "On the acoustic radiation field of the piezoelectric oscillator and the effect of viscosity on the transmission," *Can. J. Res.* **11**, 135–155 (1934).

[113] S. I. Aanonsen, "Numerical computation of the nearfield of a finite amplitude sound beam," Rep. No. 73, Department of Mathematics, University of Bergen, Bergen, Norway (1983).

[114] J. C. Lockwood, "Two problems in high intensity sound," Tech. Rep. ARL-TR-71-26, Applied Research Laboratories, The University of Texas at Austin (1971) (AD 740 498).

[115] B. G. Lucas, J. Naze Tjøtta, and T. G. Muir, "Field of a parametric focusing source," *J. Acoust. Soc. Am.* **73**, 1966–1971 (1983).

[116] B. G. Lucas and T. G. Muir, "Field of a finite-amplitude focusing source," *J. Acoust. Soc. Am.* **74**, 1522–1528 (1983).

[117] J. Naze Tjøtta and S. Tjøtta, "Sound field of a parametric focusing source," *J. Acoust. Soc. Am.* **75**, 1392–1394 (1984).

[118] T. Mannseth, "A model for acoustic transmission from water to sediment: Analytical and numerical results," thesis for the Cand. Real. Degree, Department of Mathematics, University of Bergen, Norway (1982) (in Norwegian).

[119] M. F. Hamilton, J. Naze Tjøtta, and S. Tjøtta, "Propagation and reflection of finite amplitude sound beams," in *Proceedings of the 12th International Congress on Acoustics*, Toronto, Canada (1986).

[120] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972), 9th ed.

[121] M. F. Hamilton, "Parametric acoustic array formation via weak collinear and noncollinear interaction in dispersive fluids," Tech. Rep. ARL-TR-83-19, Applied Research Laboratories, The University of Texas at Austin (1983) (ADA 130 533).

[122] D. A. Webster and D. T. Blackstock, "Asymptotic decay of periodic spherical waves in dissipative media," *J. Acoust. Soc. Am.* **64**, S33 (A) (1978).

[123] F. H. Fenlon, "A weak interaction model for the axial difference-frequency field of symmetric and asymmetric parametric acoustic transmitting arrays," *J. Sound Vib.* **64**, 17–30 (1979).

[124] F. H. Fenlon and F. S. McKendree, "Axisymmetric parametric radiation—A weak interaction model," *J. Acoust. Soc. Am.* **66**, 534–547 (1979).

[125] B. G. Lucas and T. G. Muir, "The field of a focusing source," *J. Acoust. Soc. Am.* **72**, 1289–1296 (1982).

reprinted from

Nonlinear Wave Propagation in Mechanics – AMD-Vol. 77
Editor: T.W. Wright
(Book Number H00346)
1986

published by

THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS
345 East 47th Street, New York, N.Y. 10017
Printed in U.S.A.

American Institute of Physics Handbook, 3rd edition
McGraw-Hill, New York, 1972
Chapter 3n

3n. Nonlinear Acoustics (Theoretical)

DAVID T. BLACKSTOCK

University of Texas

Until the early 1950s most of what was known about sound waves of finite amplitude was confined to propagation, and to a lesser extent reflection, of plane waves in lossless gases. Since that time a great deal has been learned about propagation in other media, about nonplanar propagation (still chiefly in one dimension), about the effect of losses, and about standing waves. Inroads have been made on problems of refraction. Diffraction is still relatively untouched.

In this section the exact equations of motion for thermoviscous fluids will first be stated. Various retreats from the full generality of these equations will then be discussed. No attempt will be made to cover streaming and radiation pressure. See Secs. 3c-7 and 3c-8 for a discussion of those topics.

GENERAL EQUATIONS FOR FLUIDS

The basic conservation equations will be stated briefly for viscous fluids with heat flow. Other compressible media, such as solids and relaxing fluids, are discussed later in the section.

3n-1. Conservation of Mass, Momentum, and Energy. In Eulerian (spatial) coordinates the continuity and momentum equations are respectively

$$\frac{D\rho}{Dt} + \rho \frac{\partial u_i}{\partial x_i} = 0 \tag{3n-1}$$

$$\rho \frac{Du_i}{Dt} + \frac{\partial p}{\partial x_i} = \frac{\partial}{\partial x_j} (\eta' d_{kk}\delta_{ij} + 2\eta d_{ij}) \tag{3n-2}$$

An entropy equation is stated here in place of the usual energy equation:

$$\rho T \frac{DS}{Dt} = C_v \left[\rho \frac{D\mathfrak{I}}{Dt} - \frac{\gamma - 1}{\beta_\epsilon} \frac{D\rho}{Dt} \right] = \psi^{(\eta)} - \frac{\partial Q_i}{\partial x_i} \tag{3n-3}$$

Here ρ is the density, u_i is the ith (cartesian) component of particle velocity, p is pressure, δ_{ij} is the Kronecker delta, $d_{ij} = \frac{1}{2}(\partial u_i/\partial x_j + \partial u_j/\partial x_i)$ is the rate-of-deformation tensor, η and η' are the shear and dilatational coefficients of viscosity, C_v and C_p are the specific heats at constant volume and pressure, $\mathfrak{I}$ is absolute temperature, S is entropy per unit mass, $\gamma = C_p/C_v$ is the ratio of specific heats, $\beta_\epsilon = -\rho^{-1}(\partial\rho/\partial\mathfrak{I})_p$ is the coefficient of thermal expansion, $\psi^{(\eta)} = 2\eta d_{ij}d_{ji} + \eta' d_{kk}d_{ii}$ is the viscous energy dissipation function, and Q_i is the ith component of the total heat flux. The material derivative $D(\ )/Dt$ stands for $\partial(\ )/\partial t + u_i\partial(\ )/\partial x_i$. If the flow of heat is due to conduction,

$$Q_i = -\kappa \frac{\partial \mathfrak{I}}{\partial x_i} \tag{3n-4}$$

where κ is the coefficient of thermal conduction. For heat radiation the relation between q and $\mathfrak{I}$ is generally quite complicated; see, for example, Vincenti and Baldwin (ref. 1). The model used by Stokes (ref. 2) amounts to Newton's law of cooling and may be expressed by

$$\frac{\partial Q_i}{\partial x_i} = \rho C_v q(\mathfrak{I} - \mathfrak{I}_0) \tag{3n-5}$$

where $\mathfrak{I}_0$ is the ambient temperature, and q is the radiation coefficient. Although too simple to describe radiant heat transfer in a fluid adequately, this equation is worth considering because of (1) its analytical simplicity and (2) its application as a convenient model for relaxation processes.

3n-2. Equation of State. To the conservation equations must be added an equation of state.

Perfect Gas. The gas law for a perfect gas is

$$p = R\rho\mathfrak{I} \tag{3n-6}$$

where R is the gas constant. An approximate form of this equation will now be derived. For a perfect gas the small-signal sound speed c_0 is given by $c_0^2 = \gamma R\mathfrak{I}_0 = \gamma p_0/\rho_0$, where p_0 and ρ_0 are the ambient values of p and ρ. Let $\mathfrak{I} = \beta_{\epsilon 0}(1 + \theta)$, $p = p_0 + \rho_0 c_0^2 P$, and $\rho = \rho_0(1 + s)$, where $\beta_{\epsilon 0}$ is the ambient value of β_ϵ (for perfect gases $\beta_{\epsilon 0}\mathfrak{I}_0 = 1$). Assume that θ, P, and s are small quantities of first order. Expansion of Eq. (3n-6) to second order yields

$$\theta = \gamma P - s + s^2 - \gamma P s \tag{3n-7}$$

First-order relations are now defined to be those that hold in linear, lossless acoustic theory; examples are $\rho_t = -\rho_0 \nabla \cdot u$ and $p - p_0 = c_0^2(\rho - \rho_0)$. At this point we assert that any factor in a second-order term in Eq. (3n-7) may be replaced by its first-order equivalent. The justification is that any more precise substitution would result in the appearance of third- or higher-order terms, and such terms have already been excluded from Eq. (3n-7). Thus in the last second-order term in Eq. (3n-7) P may be replaced by s to give

$$\theta = \gamma P - s - (\gamma - 1)s^2 \tag{3n-8}$$

correct to second order. This is a useful approximate form of the perfect gas law.

One of the most fruitful special cases to consider is the isentropic perfect gas. When a perfect gas is inviscid and there is no heat flow, Eq. (3n-3) can be used to reduce the gas law, Eq. (3n-6), to

$$\frac{p}{p_0} = \left(\frac{\rho}{\rho_0}\right)^{\gamma} \tag{3n-9}$$

The square of the sound speed, which by definition is,

$$c^2 \equiv \left(\frac{\partial p}{\partial \rho}\right)_s \tag{3n-10}$$

becomes

$$c^2 = \frac{\gamma p}{\rho} = c_0^2 \left(\frac{p}{p_0}\right)^{(\gamma-1)/\gamma} \tag{3n-11}$$

An expanded form of Eq. (3n-9) is as follows:

$$P = s + \tfrac{1}{2}(\gamma - 1)s^2 + \cdots \tag{3n-12}$$

Other Fluids. For liquids and for gases that are not perfect, one can start with a general equation of state $\mathfrak{Z} = \mathfrak{Z}(p,\rho)$. Recognizing that $(\partial \mathfrak{Z}/\partial p)_\rho = \gamma(\rho c^2 \beta_s)^{-1}$, one obtains the exact expression

$$\theta_t = \frac{\beta_{s0}}{\beta_s}(1+s)^{-1}\left[\gamma\left(\frac{c_0}{c}\right)^2 P_t - s_t\right] \tag{3n-13}$$

In order to obtain an approximation analogous to Eq. (3n-8), it is first necessary to set down a general isentropic equation of state,

$$p - p_0 = \rho_0 c_0^2 \left(s + \frac{B}{2A}s^2 + \frac{C}{3A}s^3 + \cdots\right) \tag{3n-14}$$

where the coefficients B/A, C/A, etc., are to be determined experimentally (see Sec. 3o). With the help of this expression and some elementary thermodynamic relations, one invokes the approximation procedure described following Eq. (3n-7) and reduces Eq. (3n-13) to (ref. 3)

$$\theta = \gamma P - s - (h-1)s^2 \tag{3n-15}$$

correct to second order, where

$$h = 1 + \frac{\gamma B}{2A} + \tfrac{1}{2}(\gamma - 1)\left(1 - \frac{B}{2A}\right) - (\gamma-1)^2(4\beta_{s0}\mathfrak{Z})^{-1} \tag{3n-16}$$

If Eqs. (3n-14) and (3n-12) are compared, it will be seen that B/A replaces the quantity $\gamma - 1$ in describing second-order nonlinearity of the $p - \rho$ relation. For a perfect gas, therefore, replace B/A by $\gamma - 1$ and β_{s0} by $\mathfrak{Z}_0^{-1}$ in Eq. (3n-16). The quantity h then reduces to γ, and Eq. (3n-7) is recovered.

PROPAGATION IN LOSSLESS FLUIDS

For isentropic flow (taken here to mean that the entropy of every particle is the same and remains so) Eqs. (3n-1) and (3n-2) reduce to

$$\frac{D\rho}{Dt} + \frac{\rho \partial u_i}{\partial x_i} = 0 \tag{3n-17a}$$

$$\frac{\rho Du_i}{Dt} + \frac{\partial p}{\partial x_i} = 0 \tag{3n-17b}$$

and the equation of state may be expressed simply by $p = p(\rho)$. If the new thermodynamic quantity

$$\lambda \equiv \int_{\rho_0}^{\rho} \frac{c}{\rho'}\, d\rho' \tag{3n-18}$$

is introduced, Eqs. (3n-17) take the following symmetric form:

$$\frac{D\lambda}{Dt} + \frac{c\partial u_i}{\partial x_i} = 0 \tag{3n-19a}$$

$$\frac{Du_i}{Dt} + \frac{c\partial \lambda}{\partial x_i} = 0 \tag{3n-19b}$$

Very little has been done in the way of solving these general equations.

3n-3. Plane Waves in Lossless Fluids. For one-dimensional flow in the x direction Eqs. (3n-19) become

$$\lambda_t + u\lambda_x + cu_x = 0 \tag{3n-20a}$$

$$u_t + uu_x + c\lambda_x = 0 \tag{3n-20b}$$

where subscripts x and t now denote partial differentiation, and u represents the particle velocity in the x direction. Hyperbolic equations of this form have been studied in great detail (ref. 4). Their solutions are of two general types: (1) those representing simple waves (waves propagating in one direction only), and (2) those representing compound waves (waves propagating in both directions).

Simple Waves. Simple-wave flow is characterized by the existence of a unique relationship between the particle velocity and the local thermodynamic state of the fluid. For simple waves traveling into a medium at rest, this relationship is (ref. 5)

$$\lambda = \pm u \tag{3n-21}$$

where the (+) sign holds for outgoing waves (waves traveling in the direction of increasing x), and the (−) sign for incoming waves (waves traveling in the direction of decreasing x). Hereinafter when multiple signs are used, the upper sign pertains to outgoing waves. Equations (3n-20) now reduce to the single equation

$$u_t + (u \pm c)u_x = 0 \tag{3n-22}$$

which becomes autonomous once the equation of state is specified, since Eqs. (3n-18) and (3n-21) imply a relationship $c = c(u)$. Note that the linearized version of Eq. (3n-22), $u_t \pm c_0 u_x = 0$, possesses the familiar traveling-wave solution $u = f(x \mp c_0 t)$ of linear acoustics.

The most important nonlinear effect in simple-wave flow can be readily identified directly from Eq. (3n-22). Combine that equation with the differential expression $du = u_x\,dx + u_t\,dt$ to obtain

$$\left(\frac{dx}{dt}\right)_{u=\text{const}} = -\frac{u_t}{u_x} = u \pm c \tag{3n-23}$$

This relation states that the propagation speed of a given point on the waveform (the point being identified by the value of u there) is $u \pm c$. In linear theory the propagation speed of all points is the same, namely, $\pm c_0$. The ramifications of the variable propagation speed are discussed in Sec. 3n-4.

Compound Waves. When waves traveling in both directions are present, there is no fixed relationship between u and λ. A propagation speed can still be defined, however. New dependent variables $\mathfrak{r}$ and $\mathfrak{s}$, called "Riemann invariants," may be defined by

$$2\mathfrak{r} = \lambda + u \qquad 2\mathfrak{s} = \lambda - u \tag{3n-24}$$

If Eqs. (3n-20) are first added and then subtracted, the results are respectively

$$\mathfrak{r}_t + (u + c)\mathfrak{r}_x = 0 \tag{3n-25a}$$

$$\mathfrak{s}_t + (u - c)\mathfrak{s}_x = 0 \tag{3n-25b}$$

Thus, as first found by Riemann (ref. 6),

$$\left(\frac{dx}{dt}\right)_{r=\text{const}} = u + c \tag{3n-26a}$$

$$\left(\frac{dx}{dt}\right)_{s=\text{const}} = u - c \tag{3n-26b}$$

Despite its apparent simplicity, this result is much more complicated to apply than Eq. (3n-23).

3n-4. Plane, Simple Waves in Lossless Gases. For perfect gases the isentropic equation of state is given by Eq. (3n-9). For this case $\lambda = 2(c - c_0)/(\gamma - 1)$, and the simple-wave relation Eq. (3n-21) becomes

$$c = c_0 \pm (\beta - 1)u \tag{3n-27}$$

where $\beta = \frac{1}{2}(\gamma + 1)$. Combination of this equation with Eq. (3n-11) leads to

$$p - p_0 = p_0 \left\{ \left[1 \pm (\beta - 1)\frac{u}{c_0} \right]^{2\gamma/(\gamma-1)} - 1 \right\} \tag{3n-28}$$

which can be used to obtain the characteristic impedance for finite-amplitude waves. For weak waves, i.e., $u/c_0 \ll 1$, this expression reduces to the traditional one,

$$p - p_0 = \pm \rho_0 c_0 u \tag{3n-29}$$

The nonlinear differential equation for simple waves, Eq. (3n-22), becomes

$$u_t + (\beta u \pm c_0)u_x = 0 \tag{3n-30}$$

If we restrict ourselves momentarily to outgoing waves, the propagation speed is

$$\left(\frac{dx}{dt}\right)_{u=\text{const}} = \beta u + c_0 \tag{3n-31a}$$

which shows quite clearly that the peaks of the wave travel fastest, the troughs slowest. Equivalently, as the wave travels from one point to another, the peaks suffer the least delay, the troughs the most. This latter view is illustrated in Fig. 3n-1,

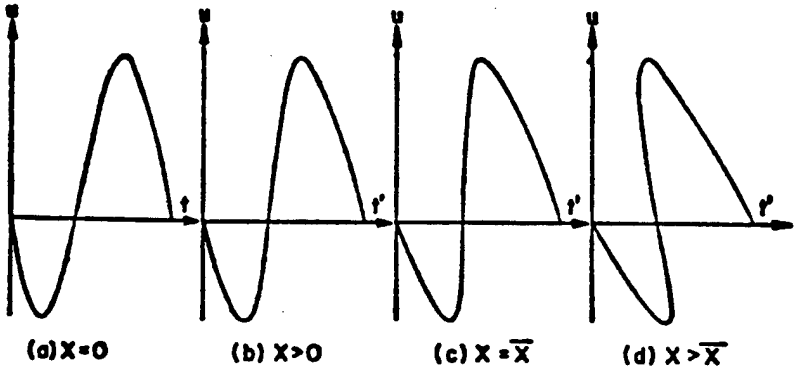


FIG. 3n-1. Progressive distortion of a finite-amplitude wave. Symbols are: u = particle velocity, x = spatial coordinate, t = time, $t' = t - x/c_0$ (delay time), $\bar{x}$ = point at which a shock begins to form.

which shows the time waveform of an outgoing disturbance at various distances from the source. The progressive distortion is quite striking, leading eventually to the curious waveform shown in Fig. 3n-1*d*. The interpretation of Fig. 3n-1*d* will be discussed presently.

Why physically does the exact propagation speed differ from c_0, the accepted value in linear theory? Two effects are at work: one kinematic, the other thermodynamic.

The sound wave travels with speed c with respect to the fluid particles. But these particles are themselves in motion, moving with velocity u. To a fixed observer, therefore, the net speed is $u + c$. This is the kinematic effect and is frequently referred to as *convection* (the fluid particles convect the wave along as a result of their own motion). The thermodynamic effect is the deviation from constancy of the sound speed c. Where the acoustic pressure is positive, the gas is a little hotter. Consequently c is greater. Conversely, in the wave troughs, where the gas is expanded and therefore colder, c is less. The variation of c from point to point along the wave can be traced to nonlinearity of the pressure-density relation. As Eq. (3n-10) shows, c would be constant if p were linearly related to ρ. This would be true, for example, for an isothermal gas.

For an incoming wave the propagation speed is

$$\left(\frac{dx}{dt}\right)_{u=\text{const}} = \beta u - c_0 \tag{3n-31b}$$

Similar arguments apply in this case. A difference is that the troughs of the particle velocity wave travel fastest (in a backward direction), the peaks slowest. Because pressure and particle velocity are out of phase in an incoming wave, however, it is still true that the peaks of the pressure wave proceed most rapidly and the troughs least so.

General Solutions. Three forms of the general solution of Eq. (3n-30) are now given. First is what might be called the "Poisson solution" (ref. 7)

$$u = f[x - (\beta u \pm c_0)t] \tag{3n-32}$$

which is implied by Eq. (3n-31); f is an arbitrary function. This result is most easily interpreted as the solution of an initial-value problem for which the spatial dependence of the particle velocity is prescribed everywhere at $t = 0$, i.e., $u(x,0) = f(x)$. The problem is somewhat artificial, however, because the progressive wave motion must already exist at $t = 0$. Of more practical interest are boundary-value problems involving a source; then simple waves arise quite naturally. If the time history of the particle velocity is known at a particular place, say $u(0,t) = g(t)$, the solution is

$$u = g\left(t - \frac{x}{\beta u \pm c_0}\right) \tag{3n-33}$$

This equation has been used to construct the waveforms in Fig. 3n-1. To make such constructions, it is convenient to use the following "inverted" form of the solution:

$$t' = g^{-1}(u) - \frac{\beta u}{c_0 \pm \beta u}\frac{x}{c_0} \tag{3n-34}$$

where $t' = t \mp x/c_0$ is the delay (for outgoing waves) or advance (for incoming waves) time appropriate for zeros of the waveform, and $g^{-1}(u)$ is the inverse function corresponding to g, i.e., $g^{-1}[g(u)] = u$.

The solution of the classic piston problem, in which a piston at rest begins at time $t = 0$ to move smoothly with a given displacement $X(t)$ in a lossless tube, is more complicated because of the moving boundary condition

$$u[X(t),t] = X'(t)H(t) \tag{3n-35}$$

where $H(t)$ is the unit step function. The solution of this problem may be given in parametric form as follows (refs. 5, 8):

$$u = X'(\phi)H\left(\frac{t \mp x}{c_0}\right) \tag{3n-36a}$$

where

$$\phi = t - \frac{x - X(\phi)}{\beta X'(\phi) \pm c_0} \tag{3n-36b}$$

The parameter ϕ represents the time at which a given signal (i.e., given value of u) left the piston.

It is generally quite difficult to convert any of the three general solutions into an explicit analytical expression $u(x,t)$. One can, however, always obtain a sketch of the waveform through use of the inversion procedure indicated by Eq. (3n-34).

Shock Formation. A more far-reaching limitation, both mathematically and physically, is that these solutions contain the seeds of their own destruction. Except for a wave of pure expansion, the dependence of the propagation speed on u will cause steepening of the waveform. Steepening eventually leads to multivalued shapes like that shown in Fig. 3n-1*d*. But these must be rejected because pressure disturbances in nature cannot be multivalued, either in time or in space. In fact, once any section of the waveform attains a vertical tangent, as in Fig. 3n-1*c*, results cannot in general be continued further (ref. 9). Physically, what happens is that a shock wave begins to form. For reasons discussed in detail in Sec. 3n-8, this formally marks the end of validity of lossless, simple-wave theory. For mathematical analyses of shock formation see, for example, refs. 4 and 8.

Fubini Solution. A problem of special interest in acoustics is the propagation of a finite-amplitude wave that is sinusoidal at its point of origin. Suppose that the wave is produced by sinusoidal vibration of a piston in a lossless tube. Let the piston displacement be given by $X(t) = (u_0/\omega)(1 - \cos \omega t)$ where u_0 is the velocity amplitude of the piston, and ω is the angular frequency. The solution is given by applying Eqs. (3n-36). For the outgoing wave we have

$$\frac{u}{u_0} = \sin \omega\phi H\left(t - \frac{x}{c_0}\right) \tag{3n-37a}$$

where

$$\omega\phi = \omega t - \frac{kx - \epsilon(1 - \cos \omega\phi)}{1 + \beta\epsilon \sin \omega\phi} \tag{3n-37b}$$

Here $k = \omega/c_0$ is the wave number, and $\epsilon = u_0/c_0$ is the velocity amplitude expressed as a Mach number.

An explicit solution is now sought by writing u as a Fourier series,

$$\frac{u}{u_0} = \Sigma A_n \cos n(\omega t - kx) + \Sigma B_n \sin n(\omega t - kx) \tag{3n-38}$$

Although the exact expressions for all the coefficients A_n and B_n have not been obtained, an approximate computation is available. First expand Eq. (3n-37*b*), writing σ for $\beta\epsilon kx$, and t' for $t - x/c_0$, and rearrange as follows:

$$\omega\phi - \omega t' = \sigma \sin \omega\phi + \epsilon(1 - \cos \omega\phi - \beta\sigma \sin^2 \phi) + O(\epsilon^2)$$

If $\sigma \gg \epsilon$ (i.e., $\beta kx \gg 1$), and $\epsilon \ll 1$, this equation reduces to

$$\omega\phi = \omega t' + \sigma \sin \omega\phi \tag{3n-39}$$

Under this approximation the Fourier coefficients A_n vanish, and the B_n can be evaluated in terms of Bessel functions. The final result is (ref. 8)

$$\frac{u}{u_0} = \sum_{n=1}^{\infty} \frac{2}{n\sigma} J_n(n\sigma) \sin n(\omega t - kx) \tag{3n-40}$$

which is generally referred to as the Fubini solution (ref. 10).

The acoustic pressure signal is found by substituting the value of u given by Eq. (3n-40) in the linear impedance relation, Eq. (3n-29). Use of a more accurate

expansion of Eq. (3n-28) for this purpose would not be consistent with the approximations that led to Eq. (3n-39).

The shock formation distance for this problem can be deduced by inspection of Eqs. (3n-39) [or, alternatively, the exact expression Eqs. (3n-37*b*)] and (3n-37*a*). The relationship of u to t' is one-to-one only if $\sigma < 1$. For $\sigma \geq 1$ the waveform curve $u(t')$ is multivalued. Hence a shock starts to form at $\sigma = 1$, i.e., at

$$\bar{x} = (\beta\epsilon k)^{-1} \tag{3n-41}$$

where the overbar signifies shock formation. The physical interpretation of σ is therefore that it is a spatial variable scaled in terms of the shock formation distance. The Fubini solution is not valid beyond the point $\sigma = 1$.

3n-5. An Approximate Theory of Lossless Simple Waves. The approximations leading to the Fubini solution can be used to obtain a general approximate theory of traveling waves of finite amplitude. The mathematical restrictions required are

$$\sigma \gg \epsilon \tag{3n-42a}$$

$$\epsilon \ll 1 \tag{3n-42b}$$

where the definitions of σ and ϵ are generalized to

$$\sigma = \frac{\beta\epsilon x}{x_c} \qquad \epsilon = \frac{u_0}{c_0} \tag{3n-43}$$

Here x_c is a characteristic distance defined so that significant distortion (for example, shock formation) takes place over the range $0 < \sigma < 1$, and u_0 is the maximum particle velocity that occurs in the flow. The physical implications of these restrictions are as follows:

1. The finite displacement of the source can be neglected. In other words, the exact boundary condition given by Eq. (3n-35) can be replaced by

$$u(0,t) = X'(t)H(t) \tag{3n-44}$$

Any error thus committed is made small by inequality (3n-42*a*).

2. The linear impedance relation, Eq. (3n-29), can be used to obtain the acoustic pressure, once the particle velocity waveform is known.

3. The nonlinear effect that *must* be taken into account is the nonconstancy of the propagation speed. But this effect is approximated by writing Eqs. (3n-31) as follows:

$$\left(\frac{dx}{dt}\right)_{u=\text{const}} \doteq \frac{\pm c_0}{1 \mp \beta u/c_0} \tag{3n-45}$$

Retention of nonconstancy of the propagation speed as the only important nonlinear effect gives recognition to the fact that this effect is the only *cumulative* one. It is the cause of the progressive distortion that engulfs the wave. We neglect the other nonlinear effects because they are *noncumulative*, or local. The distortion they cause does not grow with distance.

The formal theory based on these ideas will now be developed. An approximate differential equation may be derived by applying the method used earlier to convert Eq. (3n-7) to (3n-8). For simple waves the appropriate first-order relation is $u_x = \mp c_0^{-1}u_t$. When this is substituted in the nonlinear term in Eq. (3n-30), the result is

$$c_0 u_x \pm u_t - \beta c_0^{-1} u u_t = 0 \tag{3n-46}$$

This differential equation could also have been deduced from Eq. (3n-45).

Next let x and $t' = t \mp x/c_0$ be new independent variables. Equation (3n-46) reduces to

$$c_0^2 u_x - \beta u u_{t'} = 0 \tag{3n-47}$$

For the boundary condition

$$u\Big|_{x=0} = g(t)H(t) = g(t') \tag{3n-48}$$

where it is assumed that $g(t) = 0$ for $t < 0$, the solution is

$$u = g(\phi) \tag{3n-49a}$$
$$\phi = t' + \beta c_0^{-2} x g(\phi) \tag{3n-49b}$$

When the excitation is sinusoidal, i.e. $g(t) = u_0 \sin \omega t$, the Fubini solution follows exactly. It is also worth noting that within the limits of the approximate theory the difference between Lagrangian and Eulerian coordinates is negligible. As a general rule, the approximate theory is useful when $\epsilon < 0.1$ (ref. 8).

3n-6. Plane, Simple Waves in Liquids and Solids. *Liquids.* For lossless fluids whose isentropic equation of state is not given by Eq. (3n-9), we may proceed by using Eq. (3n-14). The propagation speed is (ref. 8)

$$\left(\frac{dx}{dt}\right)_{u=\text{const}} = u \pm c_0(1 + c_1 U + c_2 U^2 + \cdots) \tag{3n-50}$$

where $U = u/c_0$ and $c_1 = B/2A$, $c_2 = C/2A + B/4A - (B/2A)^2$, etc. Thus, in the exact solution of the piston problem [Eqs. (3n-36)], the parameter ϕ is given by

$$\phi = t - \frac{x - X(\phi)}{u \pm c_0(1 + c_1 U + c_2 U^2 \cdots)} \tag{3n-51}$$

where U is to be interpreted as $c_0^{-1} x_t(\phi)$.

Solids. The mathematical formalism for plane, longitudinal elastic waves in solids, either crystalline or isotropic, is very similar to that for liquids and gases (refs. 11-13). The wave equation is given in Lagrangian coordinates as

$$\xi_{tt} = c_0^2 G(\xi_a)\xi_{aa} \tag{3n-52}$$

where

$$G(\xi_a) = 1 + \left(\frac{M_3}{M_2}\right)\xi_a + \left(\frac{M_4}{M_2}\right)\xi_{aa} \cdots \tag{3n-53}$$

Here a represents the rest position of a particle; ξ is partical displacement; and M_2, M_3, M_4, etc., are quantities involving the second-, third-, fourth-, and higher-order elastic coefficients (ref. 12). The quantity $c_0^2 G$ plays the same role that $(\rho c/\rho_0)^2$ does for fluids (ref. 14). By the Lagrangian equation of continuity, $\rho_0/\rho = 1 + \xi_a$; thus replace Eq. (3n-18) by

$$\lambda = -c_0 \int_0^{\xi_a} [G(\xi_a')]^{\frac{1}{2}}\, d\xi_a' \tag{3n-54}$$
$$= -c_0[\xi_a - \tfrac{1}{4}m_3\xi_a^2 + (\tfrac{1}{8} - \tfrac{1}{6}m_4)m_3^2\xi_a^3 \cdots] \tag{3n-55}$$

where $m_3 = -M_3/M_2$, $m_4 = 1 - M_4/M_2 m_3^2$, etc. Riemann invariants are defined as before by Eq. (3n-24). Note that $u = \xi_t$ in Lagrangian coordinates.

Simple-wave fields are again specified by Eq. (3n-21), which when combined with Eq. (3n-5) leads to

$$\xi_a = \mp U + \tfrac{1}{4}m_3 U^2 \mp \tfrac{1}{6}m_4 m_3^2 U^3 \cdots \tag{3n-56}$$

The propagation speed for simple waves is

$$\left(\frac{da}{dt}\right)_{u=\text{const}} = \pm c_0 G^{\frac{1}{2}} \tag{3n-57}$$

The factor u, which appears in Eq. (3n-23), is absent here because the coordinate system is Lagrangian. Equation (3n-57) expanded in series form is

$$\left(\frac{da}{dt}\right)_{u=\text{const}} = \pm c_0[1 \pm \tfrac{1}{2}m_3 U + \tfrac{1}{4}m_3^2(1 - 2m_4)U^2 \cdots] \tag{3n-58}$$

Therefore, the solution of the piston problem, given $u(0,t) = X_t(t)$, is

$$\phi = t \mp \frac{a/c_0}{1 \pm \frac{1}{2}m_3 U + \frac{1}{4}m_3^2(1 - 2m_4)U^2 \cdots} \tag{3n-59}$$

where U is to be interpreted, as in Eq. (3n-51), as $c_0^{-1}X_t(\phi)$. More complete versions of some of the series expansions given above can be found in ref. 12.

Approximate Theory. The approximate theory of simple waves described in Sec. 3n-5 is very easily generalized to apply to liquids and solids. For liquids $\gamma - 1$ is replaced by B/A, as mentioned after Eq. (3n-16). For solids $\gamma + 1$ is replaced by $-M_3/M_2$ (see ref. 12 for other useful associations). Therefore, let

$$\beta = \tfrac{1}{2}(\gamma + 1) \qquad \text{for gases} \tag{3n-60a}$$

$$\beta = 1 + \frac{B}{2A} \qquad \text{for liquids} \tag{3n-60b}$$

$$\beta = \frac{-M_3}{2M_2} \qquad \text{for solids} \tag{3n-60c}$$

and all results stated in Sec. 3n-5 become applicable for a very wide range of continuous media. For many liquids and solids the first "nonlinearity coefficient" (B/A for liquids, M_3/M_2 for solids) is known, but higher-order ones are not. In such cases it is difficult to justify using anything more precise than the approximate theory. But see ref. 12 for a discussion related to this point.

3n-7. Nonplanar Simple Waves. In this section one-dimensional nonplanar waves are considered, namely, spherical and cylindrical waves, and waves in horns. The general theory is not very highly developed. One fundamental difficulty is that simple waves of arbitrary waveform do not generally exist for nonplanar waves (ref. 15). Consider, for example, the wave motion generated by a pulsating sphere in an infinite medium. Most of the wave field consists of outgoing radiation, but there is also some backscatter (ref. 15). In the far field, however, simple waves do occur as an approximation. This is the case treated here. The results represent an extension of the approximate theory developed in Secs. 3n-5 and 3n-6.

Spherical and Cylindrical Waves. For large values of the radial coordinate r (actually large kr, where k is an appropriate wave number of the disturbance), the following approximate equation for simple waves in a fluid can be obtained (ref. 16):

$$c_0^2 w_z - \beta w w_{t'} = 0 \tag{3n-61}$$

where $t' = t \mp (r - r_0)/c_0$, r_0 is a reference distance, and β is given by Eq. (3n-60*a*) or (3n-60*b*). This equation may also apply to longitudinal waves in an isotropic solid, but so far no derivation has been given. The dependent variable w equals $(r/r_0)^{\frac{1}{2}}u$ and $(r/r_0)u$ for cylindrical and spherical waves, respectively. The independent variable z is given for the two cases by

Cylindrical: $$z = 2(\sqrt{r} - \sqrt{r_0})\sqrt{r_0} \tag{3n-62a}$$

Spherical: $$z = r_0 \ln \frac{r}{r_0} \tag{3n-62b}$$

Note that $z > 0$ for diverging waves ($r > r_0$), but $z < 0$ for converging waves ($r < r_0$).

Equation (3n-61) is solved by recognizing that it has the same form as the plane-wave equation (3n-47). For the boundary condition take $u(r_0,t) = g(t)$, which may represent either the motion of a source at r_0 or the measured time signal of a wave as it passes by the point r_0. Since $z = 0$ and $t' = t$ when $r = r_0$, the condition on w is

$$w(0,t') = g(t') \tag{3n-63}$$

Therefore, for the two kinds of waves the solution is

Cylindrical:
$$u = \left(\frac{r_0}{r}\right)^{\frac{1}{2}} g(\phi) \tag{3n-64a}$$
$$\phi = t' + 2\beta c_0^{-2}\sqrt{r_0}(\sqrt{r} - \sqrt{r_0})g(\phi) \tag{3n-64b}$$

Spherical:
$$u = \frac{r_0}{r} g(\phi) \tag{3n-65a}$$
$$\phi = t' + \beta c_0^{-2} r_0 \ln \frac{r}{r_0} g(\phi) \tag{3n-65b}$$

Some applications of these results are given in refs. 16 to 18. It has been shown (ref. 19) that Eq. (3n-65*b*) corresponds to a second-order approximation of results obtained using the Kirkwood-Bethe hypothesis (ref. 20).

Many special solutions for spherical and cylindrical waves have also been found. Most are of the similarity type. The most famous is Taylor's solution for the compression wave generated by a sphere that expands at a constant rate (refs. 21, 22).

Waves in Horns. For waves traveling in ducts whose cross-sectional area $A = A(x)$ does not vary rapidly, the waves may be assumed to be quasi-plane. It is assumed that the effect of variations in the cross section can be accounted for simply by correcting the continuity equation as follows:

$$\frac{D(A\rho)}{Dt} + \rho A u_x = 0 \tag{3n-66}$$

The one-dimensional formalism is thereby retained.

By the same methods used for spherical and cylindrical waves it is possible to derive an equation exactly like Eq. (3n-61). However, w and z are now defined as

$$w = \left(\frac{A}{A_0}\right)^{\frac{1}{2}} u \tag{3n-67a}$$
$$z = \int_{x_0}^{x} \left(\frac{A_0}{A}\right)^{\frac{1}{2}} dx' \tag{3n-67b}$$

where x_0 is a reference distance, $A_0 = A(x_0)$, and $t' = t \pm (x - x_0)/c_0$. The sign of z identifies the wave as outgoing ($x > x_0$) or incoming ($x < x_0$). Note that a conical horn ($A \propto x^2$) gives results identical with those for spherical waves, and a parabolic horn ($A \propto x$) gives results identical with those for cylindrical waves.

The general solution for a boundary condition of the form given by Eq. (3n-63) is (ref. 23)

$$w = \left(\frac{A}{A_0}\right)^{\frac{1}{2}} u = g(\phi) \tag{3n-68a}$$
$$\phi = t' + \beta c_0^{-2} z g(\phi) \tag{3n-68b}$$

For reference the value of the stretched coordinate z for an exponential horn ($A \propto e^{2lx}$) is

$$z = l^{-1}(1 - e^{-l(x-x_0)}) \tag{3n-69a}$$

and for a catenoidal horn ($A \propto \cosh^2 lx$) is

$$z = 2l^{-1}(\tan^{-1} e^{lx} - \tan^{-1} e^{lx_0}) \cosh lx_0 \tag{3n-69b}$$

All the results previously obtained for plane waves (approximate theory) may now be applied to nonplanar one-dimensional waves simply by replacing u and x by w and z, as given by Eqs. (3n-67). For example, for sinusoidal excitation at $x = x_0$ the shock formation distance is found by putting $\bar{z} = \pm(\beta\epsilon k)^{-1}$ and then making use of Eq. (3n-67*b*).

Parametric Array. An application of particular interest is the so-called parametric, end-fired array, conceived by Westervelt (ref. 53). A source such as a baffled piston emits radiation consisting of two high-frequency carrier waves into an open medium. The carriers, whose frequencies are ω_1 and ω_2, interact nonlinearly to produce a difference-frequency wave (frequency $\omega_d = \omega_2 - \omega_1$). Also produced, of course, but not of interest here, are the harmonics of the two carriers as well as the sum-frequency and other intermodulation components (ref. 54). In Westervelt's original treatment the two carrier waves were assumed to be collinear beams of collimated plane waves. More recently, Muir (ref. 55) has taken the directivity and spherical spreading of the carriers into account. In any case, however, the interaction to produce the difference-frequency wave amounts to setting into operation a line of virtual sources of frequency ω_d, all phased so as to constitute an end-fired array. The result is that the difference-frequency wave has a very high directivity. In other words, a low-frequency beam is produced that is much more highly directive than would have been the case had the source emitted the difference-frequency signal directly. Typically, too, there are no minor lobes. Absorption by the medium may be relied upon to filter out the two carrier waves and the sum-frequency component, eventually leaving the difference-frequency wave as the most prominent signal. Experiments have confirmed the remarkable properties of the parametric array (refs. 55, 56), and many further studies of it have been done (ref. 57).

WEAK-SHOCK THEORY

3n-8. General Discussion. The appearance of shocks in a flow poses a serious challenge to the theory of simple waves as developed thus far. In the first place, the waveform gradient at a shock is so high that the dissipation terms in Eqs. (3n-2) and (3n-3), heretofore deemed negligible, are in fact very large. A second problem is that since the shock is (at least approximately) a discontinuity in the medium, it can cause partial reflection of signals that catch up with it. The presence of reflected waves invalidates the simple-wave assumption. Strictly speaking, therefore, the flow cannot be simple wave, once shocks form (ref. 9).

The situation is not quite so bad as it seems, however, provided we restrict ourselves to relatively weak waves, i.e., $u_0/c_0 < 0.1$, approximately. Under this condition the signals that are reflected from a shock in the waveform are so feeble as to be negligible. The simple-wave model may therefore be retained as a good approximation. Next, triple-valued waveforms of the kind shown in Fig. 3n-1 must be avoided. This requires that provision be made for dissipation. There are two approaches. First, one can take explicit account of the dissipation terms. This leads to Burgers' equation, or variations thereof; the method is described in Sec. 3n-12. Alternatively, one can postulate mathematical discontinuities—shocks—at places where the waveform would otherwise be triple valued. The Rankine-Hugoniot relations are invoked to relate conditions on either side of each shock. In this way dissipation is accounted for indirectly. A tacit assumption, it will be noted, is that all the dissipation takes place at the shocks.

The mathematical method is more fully appreciated if the physical aspects of the process are first understood. The history of a typical waveform is depicted in Fig. 3n-2 (taken from ref. 27). Figure 3n-2*a* shows the initial waveform. Numbered dots indicate initial phase points (values of ϕ) on the wave. In the beginning, distortion takes place as described in Sec. 3n-4 (Fig. 3n-2*b* and *c*). After the shock is born (Fig. 3n-2*c*), it travels supersonically. In consequence of Eq. (3n-72), however, phase points just behind, such as number 5, travel faster. As they catch up with the shock, it grows because the top of the discontinuity is always determined by the amplitude of the phase point that just caught up with it. (Conversely, the bottom of the discontinuity always coincides with the phase point just overtaken by

the shock.) The top reaches a maximum when phase point 5 catches up. After that, the top decays (Fig. 3n-2*e*). In Fig. 3n-2*f* the decay has progressed to the extent that all phase points of the original waveform between 4 and 6 have disappeared. Eventually all that remains (Fig. 3n-2*g*) is the shock and a linear section connecting it with the zero, phase point 7. This is the asymptotic shape toward which many waveforms or waveform sections tend (ref. 26).

3n-9. Mathematical Formulation of Weak-shock Theory. For the continuous sections of the waveform the most general solution from the approximate theory of simple waves is adopted, namely, Eqs. (3n-68), where w and z are given by Eqs. (3n-67). Plane, cylindrical, and spherical waves, which are not really "quasi-plane," are nevertheless included formally within the framework of this solution by taking $A = 1$, x, and x^2, respectively.

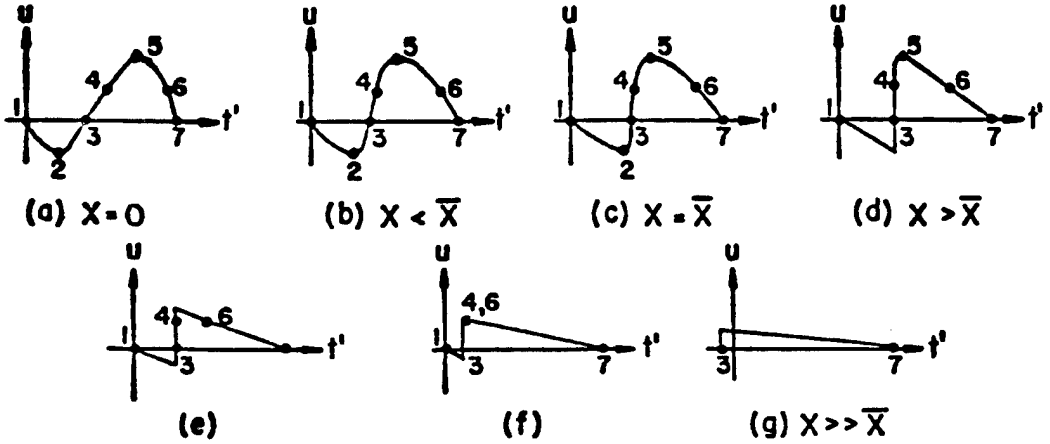


FIG. 3n-2. Development and decay of a finite-amplitude wave. Numbered points refer to initial phases (values of ϕ) of the wave. (*From ref.* 27.)

Suppose now that a shock begins to form at time $\bar{t}$ and distance $\bar{x}$. It will arrive at a subsequent point x at time t_s given by

$$t_s = \bar{t} + \int_{\bar{x}}^{x} v^{-1}\, d\mu \tag{3n-71}$$

where v is the shock's propagation speed. The Rankine-Hugoniot relations can be combined to give v in terms of u_a and u_b, the particle velocities just ahead of and just behind the shock, respectively. An approximation of the required relation is

$$v = \pm c_0 + \tfrac{1}{2}\beta(u_a + u_b) \tag{3n-72}$$

or, to the same order,

$$v^{-1} = \pm c_0^{-1} - \tfrac{1}{2}\beta c_0^{-2}(u_a + u_b) \tag{3n-73}$$

Substitution of this value in Eq. (3n-71) leads to

$$t_s' = \bar{t}' - \tfrac{1}{2}\beta c_0^{-2} \int_{\bar{x}}^{x} (u_a + u_b)\, d\mu \tag{3n-74}$$

where overbars continue to indicate values at the instant of shock formation, and primes denote retarded (or advanced) time. In terms of the generalized dependent and independent variables w and z, Eq. (3n-74) becomes

$$t_s' = \bar{t}' - \tfrac{1}{2}\beta c_0^{-2} \int_{\bar{z}}^{z} (w_a + w_b)\, d\mu \tag{3n-75}$$

An equivalent relation is

$$\frac{dt'_s}{dz} = -\tfrac{1}{2}\beta c_0^{-2}(w_a + w_b) \tag{3n-76}$$

Once the particle velocity u has been determined, the linear impedance relation, Eq. (3n-29), is used to find the pressure signal (ref. 23).

This completes the formal solution, except for some interpretation. The waveform in the continuous sections between shocks is prescribed by Eqs. (3n-68). For each shock the path and amplitude are determined by Eq. (3n-75) or Eq. (3n-76) together with Eqs. (3n-68), which are to be evaluated just ahead of the shock ($u = u_a$, $\phi = \phi_a$, $t' = t'_s$) and just behind it ($u = u_b$, $\phi = \phi_b$, $t' = t'_s$). In principle, Eqs. (3n-68) can be combined to eliminate the parameter ϕ as follows:

$$t' = g^{-1}(w) - \beta c_0^{-2} z w \tag{3n-77}$$

Hence just ahead of the shock

$$t'_s = g^{-1}(w_a) - \beta c_0^{-2} z w_a \tag{3n-78a}$$

and just behind

$$t'_s = g^{-1}(w_b) - \beta c_0^{-2} z w_b \tag{3n-78b}$$

Equations (3n-78a), (3n-78b), and (3n-75) or (3n-76) are to be solved simultaneously for w_a, w_b, and t'_s.

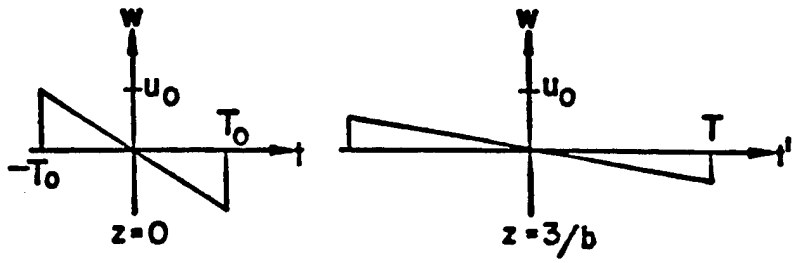


FIG. 3n-3. N wave.

3n-10. Applications of Weak-shock Theory. *N Wave.* Perhaps the most famous application is to the wave shaped like the letter N. The sonic boom is a cylindrical N wave in the far field. For the present consider outgoing waves only. Refer to Fig. 3n-3 for notation. At $t = 0$, $u = -u_0 t/T_0$ for $-T_0 < t < T_0$. Thus $g(\phi) = -u_0\phi/T_0$, and Eq. (3n-68b) yields $\phi = t'/(1 + bz)$, where $b = \beta u_0/c_0^2 T_0$. The solution is given by Eq. (3n-68a) as

$$w = -\frac{t'}{T_0}\frac{u_0}{1 + bz} \qquad -T < t' < T$$

To determine T, make use of Eq. (3n-76) for the head shock: that is,

$$\frac{dt'_s}{dz} = -\tfrac{1}{2}\beta c_0^{-2} w_b = \frac{\frac{1}{2}bt'_s}{1 + bz}$$

Integration gives

$$-t'_s = T = T_0(1 + bz)^{\frac{1}{2}}$$

The amplitude of the wave is therefore given by

$$u_b = \left(\frac{A_0}{A}\right)^{\frac{1}{2}} \frac{u_0}{(1 + bz)^{\frac{1}{2}}}$$

Next consider incoming waves. The major difference in the results is that z is replaced by $-z$. But z itself also changes sign [see the discussion following Eqs. (3n-67)]. The following formulas cover both incoming and outgoing waves:

$$w = \mp \frac{u_0}{1 + b|z|} \frac{t'}{T_0} \qquad -T < t' < T \tag{3n-79}$$

$$T = T_0(1 + b|z|)^{\frac{1}{2}} \tag{3n-80}$$

$$|u_b| = \left(\frac{A_0}{A}\right)^{\frac{1}{2}} \frac{u_0}{(1 + b|z|)^{\frac{1}{2}}} \tag{3n-81}$$

The growth of a converging wave ($A < A_0$) and the diminution of a diverging wave ($A > A_0$) are not comparable because the factor $(1 + b|z|)^{-\frac{1}{2}}$ acts to diminish both types of waves. Both waves spread at the same rate, however. From Eq. (3n-81) one obtains the classical results that outgoing plane, cylindrical, and spherical waves decay at great distances as $x^{-\frac{1}{2}}$, $r^{-\frac{1}{2}}$, and $r^{-1}(\ln r)^{-\frac{1}{2}}$, respectively.

Sawtooth Wave. Assume that the wave shown in Fig. 3n-3*a* is repetitive. The magnitude of the jump at the shock is now $2u_0$ to begin with. Because of the symmetry, we have $u_a = -u_b$, which means that, by Eq. (3n-72), the shocks all travel at sonic speed. Unlike the N wave, therefore, the sawtooth does not stretch out as it travels. The decay is more rapid however. Proceeding as before, we find the wave amplitude to be given by

$$|w_b| = \frac{\pi u_0}{\pi + \beta\epsilon k|z|} \tag{3n-82}$$

where k is the fundamental wave number of the wave. See ref. 28 for a discussion of power loss and related topics for sawtooth waves in an exponential horn.

Originally Sinusoidal Wave. It will be recalled that a sinusoidally vibrating piston gives rise to periodic waves whose mathematical description, for outgoing waves, is given by Eq. (3n-40), the Fubini solution. Weak-shock theory makes it possible to obtain a solution of this problem for distances beyond the point of shock formation. It turns out that after forming at $x = \bar{x} = (\beta\epsilon k)^{-1}$, the shocks reach a maximum amplitude at $x = \pi\bar{x}/2$ and thereafter decay. For distance greater than $3\bar{x}$ the wave is effectively a sawtooth of amplitude

$$u_b = \frac{\pi u_0}{1 + \sigma} \tag{3n-83}$$

where (see Sec. 3n-4) $\sigma = \beta\epsilon kx = x/\bar{x}$. This problem is treated in full in ref. 27, as is the similar one of an isolated sine-wave cycle. To generalize Eq. (3n-83) to other one-dimensional outgoing waves it is merely necessary to replace u_b by w_b and σ by $\beta\epsilon kz$.

3n-11. Limitations of Weak-shock Theory. The primary advantage of weak-shock theory over the method based on Burgers' equation (see below) is that results are obtained quickly and easily. Details of the actual profile of the wave in the neighborhood of each shock are suppressed simply by approximating the shock as a mathematical discontinuity. The method's strength is also its weakness, however. At great distances the shocks may become so weak that they become dispersed and are no longer approximate discontinuities.

As a test we may compare the shock rise time (ref. 29) τ with a characteristic period or time duration T of the wave. Thus consider the ratio

$$\frac{\tau}{T} = \frac{12\delta}{c_0|u_b|T} = \frac{12\delta}{c_0|w_b|T}\left(\frac{A}{A_0}\right)^{\frac{1}{2}} \tag{3n-84}$$

where δ is proportional to the viscosity and heat conduction coefficients of the fluid [see Eq. (3n-86)]. For an N wave $|w_b|T$ is a constant ($= u_0T_0$) so that τ/T is simply

proportional to $(A/A_0)^{\frac{1}{2}}$. Therefore, if the N wave is plane, τ/T is constant, which means that the validity of the weak-shock computation does not change with distance. The wave simply spreads out as rapidly as the shock. For all other outgoing N waves, however, the shock disperses more rapidly, and eventually $\tau \sim T$, beyond which point weak-shock theory should not be trusted. Let r_{max} designate the distance at which $\tau/T = 1$. For spherical N waves we obtain,

$$\frac{r_{max}}{r_0} = \frac{\beta u_0 c_0 T_0}{12\delta} \tag{3n-85a}$$

The comparable result for cylindrical N waves is

$$\frac{r_{max}}{r_0} = \left(\frac{\beta u_0 c_0 T_0}{12\delta}\right)^2 \tag{3n-85b}$$

For an outgoing sawtooth wave τ/T is proportional to $(1 + \beta\epsilon k|z|)(A/A_0)^{\frac{1}{2}}$, which means that weak-shock theory is always limited, even when the wave is plane. Even for converging waves τ may approach T in certain instances (refs. 17, 18). Care must therefore be exercized in using asymptotic formulas based on Eq. (3n-82). Calculations of r_{max} for sawtooth waves based on taking $\tau = T$ are in agreement with estimates obtained by other methods (ref. 27).

The importance of the limitation on weak-shock theory varies a great deal in practice. For sonic booms the limitation is apparently not significant. Typically at ground level τ is of the order of milliseconds, whereas T is measured in tenths of a second. For long-range propagation of pulses from underwater explosions (ref. 30), however, the limitation can be crucial.

In conclusion we remark that "weak-shock theory" is in some respects a misnomer. The theory is valid for weak shocks but not, in general, for very weak ones.

BURGERS' EQUATION AND OTHER MODELS

We now consider explicitly the effects that viscosity, heat conduction, and relaxation have on the propagation of finite-amplitude waves. The full-fledged equations—(3n-1), (3n-2), (3n-3), and (3n-6) or other equation of state—must be dealt with. Successful attacks on these equations have been mainly directed at specific problems, such as the profile of a steady shock wave (ref. 29). General exact results analogous to those for lossless waves are not known. The only general approach presently available, that based on Burgers' equation, is limited to relatively weak waves. For our purposes, however, this method is a fitting companion for weak-shock theory and its predecessor, the approximate theory of lossless simple waves.

3n-12. Thermoviscous Fluids. Burgers' Equation. *Plane Waves.* By employing an approximation procedure similar to that used to change Eq. (3n-7) into (3n-8), Lighthill (ref. 29) reduced the equations of motion for outgoing plane waves in a thermoviscous perfect gas to Burgers' equation,

$$u_t + \beta u u_{x'} = \delta u_{x'x'} \tag{3n-86a}$$

Here $x' = x - c_0 t$, $\delta = \frac{1}{2}\nu[\mho + (\gamma - 1)/\mathrm{Pr}]$, $\nu = \eta/\rho_0$ is the kinematic viscosity, $\mho = (\eta' + 2\eta)/\eta$ is the viscosity number, and $\mathrm{Pr} = \eta C_p/\kappa$ is the Prandtl number. The equation applies as well to fluids of the arbitrary equation of state (refs. 31, 32); simply let β be given by Eq. (3n-60*b*). In certain cases it applies also to solids (ref. 33).

Equation (3n-86*a*) is convenient for initial-value problems because the moving coordinate x' reduces to $x' = x$ at $t = 0$. For boundary-value problems a more convenient, yet equally valid, form is (refs. 31, 3, 34)

$$c_0^2 u_x - \beta u u_{t'} = \pm \delta c_0^{-1} u_{t't'} \tag{3n-86b}$$

where $t' = t \mp x/c_0$. [To make Eq. (3n-86a) apply to incoming as well as outgoing waves, redefine x' as $x \mp c_0 t$.]

Burgers' equation has a known exact solution. The introduction of the logarithmic potential ζ by

$$u = \pm \frac{2\delta}{\beta c_0} (\ln \zeta)_t = \pm \frac{2\delta}{\beta c_0} \frac{\zeta_{t'}}{\zeta} \tag{3n-87}$$

causes Eq. (3n-86b) to be reduced to

$$\pm c_0^3 \zeta_x - \delta \zeta_{t't'} = 0 \tag{3n-88}$$

which is a diffusion equation with the usual roles of space and "time" reversed. To avoid confusion we drop the multiple-sign notation at this point and focus attention on outgoing waves. It is clear that an incoming wave can be considered simply by replacing δ with $-\delta$. The solution of Eq. (3n-88) [with the (+) sign] is

$$\zeta = \sqrt{\frac{K}{\pi}} \int_{-\infty}^{\infty} \zeta_0(\lambda) \exp\left[-K(\lambda - t')^2\right] d\lambda \tag{3n-89}$$

where $K = c_0^3/4\delta x$. The quantity $\zeta_0(t') = \zeta(0,t')$ represents the transformed boundary condition. If the original boundary condition is given by Eq. (3n-48), then, by Eq. (3n-87),

$$\zeta_0(t') = \exp\left[\int_{-\infty}^{t'} \frac{\beta c_0}{2\delta} g(\mu)\, d\mu\right] \tag{3n-90}$$

Normally one takes $g(t) = 0$ for $t < 0$, in which case $\zeta_0 = 1$ for $t' < 0$, and the integral's lower limit is zero. The solution of Burgers' equation has been applied to a number of specific problems (refs. 29, 32).

The only solution reviewed here is the one for which the piston motion is sinusoidal (refs. 31, 34, 35): $u(0,t) = u_0 \sin \omega t H(t)$. Equation (3n-90) gives $\zeta_0 = \exp[\frac{1}{2}\Gamma(1 - \cos \omega t')]$ for $t' > 0$ ($\zeta_0 = 1$ otherwise), where

$$\Gamma = \frac{\beta c_0 u_0}{\delta \omega} = \frac{\beta \epsilon}{\alpha \lambda} \tag{3n-91}$$

and $\alpha\lambda = \alpha/k$ is the dimensionless small-signal attenuation coefficient ($\alpha\lambda = \omega\delta/c_0^3$). The dimensionless parameter Γ characterizes the importance of nonlinear effects relative to dissipation. The value $\Gamma = 1$ roughly marks the dividing line between the importance and unimportance of nonlinearity in a periodic wave (ref. 36). When the value of ζ_0 is substituted in Eq. (3n-89), the potential ζ can be separated into transient and steady-state parts. The steady-state part, to which we restrict ourselves, may be expressed as an infinite series,

$$\zeta = I_0(\tfrac{1}{2}\Gamma) + 2 \sum_{n=1}^{\infty} (-1)^n I_n(\tfrac{1}{2}\Gamma) e^{-n^2 \alpha x} \cos n\omega t' \tag{3n-92}$$

where I_n is the Bessel function of imaginary argument.

The most interesting case is that of strong waves, i.e., $\Gamma \gg 1$. In this circumstance ζ reduces to a theta function, and the logarithmic differentiation required by Eq. (3n-87) is easy to carry out. The result is (ref. 35)

$$\frac{u}{u_0} = \frac{2}{\Gamma} \sum \frac{\sin n\omega t'}{\sinh n(1 + \sigma)/\Gamma} \tag{3n-93}$$

which is Fay's solution (ref. 37) with Fay's const. α_0 taken to be Γ^{-1}. If σ is not large, the hyperbolic sine function may be approximated by its argument, giving

$$u = \frac{2u_0}{1 + \sigma} \Sigma n^{-1} \sin n\omega t' \tag{3n-94}$$

which represents a sawtooth wave of amplitude

$$u_b = \frac{\pi u_0}{1+\sigma}$$

This is exactly the same result found by means of weak-shock theory; see Eq. (3n-83).

For strong waves at great distances, i.e., $\sigma \gg \Gamma \gg 1$, the waveform is found, either by the Fay solution or directly by Eqs. (3n-92) and (3n-87), to be

$$u \cong 4\alpha\lambda c_0 \beta^{-1} e^{-\alpha x} \sin \omega t' \tag{3n-95}$$

The simple exponential decay is expected because the wave has now become quite weak. What is remarkable is the absence of the original amplitude factor u_0. The wave amplitude at great distances is independent of the source strength. In other words saturation is reached. This result is obviously of great importance. Saturation has been observed experimentally (refs. 15, 55, 58). Note from Eq. (3n-83) that the asymptotic amplitude given by weak-shock theory is (ref. 26)

$$u_b \cong \frac{\pi c_0^2}{\beta \omega x} \tag{3n-96}$$

but this result is accurate only in the sawtooth region, which is defined roughly by $3\bar{z} < x < \alpha^{-1}$ (ref. 35).

Nonplanar Waves. For other one-dimensional waves the analog of Eq. (3n-86*b*) is

$$c_0^2 (u_x + uA_x/2A) - \beta c_0 u u_{t'} = \delta u_{t't'} \tag{3n-97}$$

(again, for incoming waves replace δ by $-\delta$). It is necessary to make the far-field assumption in deriving this equation. The transformations that have proved so helpful in previous cases, namely, Eqs. (3n-67), lead to

$$c_0^2 w_x - \beta c_0 w w_{t'} = \delta \left(\frac{A}{A_0}\right)^{\frac{1}{2}} w_{t't'} \tag{3n-98}$$

which is similar to Burgers' equation, but has one variable coefficient. No exact solutions are known.

For periodic spherical and cylindrical waves, solutions of Eq. (3n-98) have been obtained that are valid in the shock-free region ($z < \bar{z}$) and in the sawtooth region (refs. 17, 18). These solutions correspond, respectively, to the Fubini solution for spherical and cylindrical waves and to the related weak-shock solutions (ref. 27). The latter are improved upon, however, because the detailed configuration of the waveform in the vicinity of the shocks is obtained. The behavior of the shock thickness is strongly dependent upon whether the wave is a diverging or a converging one. This can be seen from the form of Eq. (3n-98). A diverging wave ($A > A_0$) is equivalent to a plane wave in a medium in which the dissipation increases with distance. Conversely, for a converging wave ($A < A_0$) the dissipation seems to decrease with distance (refs. 17, 18).

3n-13. Equations for Other Forms of Dissipation. If dissipation is due to an agency other than the thermoviscous effects discussed in the last section, it may still be possible to derive an approximate unidirectional-wave equation similar to Burgers'.

Relaxing Fluids. An elementary example of a relaxing fluid is one that radiates heat in accordance with Eq. (3n-5)(ref. 38). For simplicity take the fluid to be a perfect gas, and let it be inviscid and thermally nonconducting. At very low frequencies infinitesimal waves travel at the isothermal speed of sound, given by $b_0^2 = p_0/\rho_0$. At very high frequencies the speed is the adiabatic value, given by $b_\infty^2 =$

$\gamma p_0/\rho_0$ (the notation b_∞ is used here in place of c_0 to emphasize the role played by frequency). The dispersion m, defined by

$$m \equiv \frac{b_\infty^2 - b_0^2}{b_0^2} \tag{3n-99}$$

is equal to $\gamma - 1$ for the radiating gas. If the dispersion is very small, i.e., $m \ll 1$ (which in this case implies $\gamma \doteq 1$), the following approximate equation for plane waves can be derived:

$$\left(q + \frac{\partial}{\partial t'}\right) u_x - b_0^{-2}\left(\beta_i q + \beta_a \frac{\partial}{\partial t'}\right) u u_t' = \pm \frac{m}{2b_0} u_{t't'} \tag{3n-100}$$

where $t' = t \mp x/b_0$. It is seen that the radiation coefficient q [see Eq. (3n-5)] is the reciprocal of a relaxation time. Subscripts a and i used with β indicate adiabatic and isothermal values, respectively; that is, $\beta_a = (\gamma + 1)/2$ and $\beta_i = (1 + 1)/2 = 1$. The two values are essentially the same, since it has been assumed that $\gamma \doteq 1$. At either very low frequencies ($\omega q^{-1} \ll 1$) or very high frequencies ($\omega q^{-1} \gg 1$) the left-hand side of the equation takes on the same form as Eq. (3n-47). If the equation is linearized, a dispersion relation can be found that gives the expected behavior for a relaxation process (the actual formulas for the attenuation and phase velocity agree with the exact ones for a radiating gas only for $m \ll 1$).

Polyakova, Soluyan, and Khokhlov considered a relaxation process directly and obtained a pair of equations that can be merged to form a single equation exactly like Eq. (3n-100) except that β_i and β_a are equal (ref. 39). Some solutions (refs. 39, 40) have been found. One represents a steady shock wave. The shock profile is single-valued for very weak shocks. But when the shock is strong enough that its propagation speed [see Eq. (3n-72)] exceeds b_∞, the solution breaks down (a triple-valued waveform is predicted). This illustrates an important fact about the role of relaxation in nonlinear propagation: Relaxation absorption can stand off weak nonlinear effects, but not strong ones. In frequency terms, relaxation offers high attenuation to a broad mid-range of frequencies. If the wave is quite weak, the distortion components are easily absorbed because their frequencies fall in the range of high attenuation. But if the wave is strong, many more very high frequency components are produced, and these are not attenuated efficiently by the relaxation process. To keep the waveform from becoming triple valued, it is necessary to include a viscosity term in the approximate wave equation. In ref. 40 the problem of an originally sinusoidal wave is treated. Quantitative approximate solutions are obtained for cases in which the source frequency is either very low or very high, and a qualitative discussion is given for source frequencies in between.

Marsh, Mellen, and Konrad (ref. 30) postulated a "Burgers-like" equation for spherical waves. It is similar to Eq. (3n-100) but is corrected to take account of spherical divergence. A viscosity term is added, and β_i and β_a are the same. At either very low or very high frequencies the equation takes on the form of Eq. (3n-98) [for spherical waves $(A/A_0)^{\frac{1}{2}} = r/r_0 = e^{s/r_0}$], and some initial attempts at solving this equation were described.

Boundary-layer Effects. Consider the propagation of a plane wave in a thermoviscous fluid contained in a tube. The wave can never be truly plane because the phase fronts curve a great deal as they pass through the viscous and thermal boundary layers at the wall of the tube. If the boundary-layer thicknesses are small compared with the tube radius, however, the curvature of the phase fronts is restricted to very narrow regions, and the wave may be considered quasi-plane. The boundary layers still affect the wave, causing an attenuation that is proportional to $\sqrt{\omega}$ and a comparable dispersion. If the frequency is low, the attenuation from this source is much

more important than that due to thermoviscous effects in the mainstream (central core of the fluid), and so it makes sense to find a Burgers-like equation for this case.

A one-dimensional model of time-harmonic wave propagation in ducts with boundary-layer effects treated as a body force has been given by Lamb (ref. 41). Chester (ref. 42) has generalized this model and applied it to compound flow in a closed tube. His method can be used to obtain the following equation for simple-wave flow:

$$u_x - \frac{\beta}{c_0^2} u u_{t'} = \mp \frac{1 + (\gamma - 1)/\sqrt{\mathrm{Pr}}}{c_0 D/2} \left(\frac{\nu}{\pi}\right)^{\frac{1}{2}} \int_0^\infty u_{t'}(x, t' - \mu) \frac{d\mu}{\sqrt{\mu}} \quad \text{(3n-101)}$$

where D is the hydraulic diameter of the duct (four times the cross-sectional area divided by the circumference). No solutions are presently available. But the equation does have proper limiting forms. If the effect of the boundary layers (right-hand side) is neglected, the result is Eq. (3n-47). If the nonlinear term is dropped, the time-harmonic solution can be found, and this solution yields the correct attenuation and dispersion. Because of the relative weakness of boundary-layer attenuation (the dimensionless attenuation $\alpha\lambda$ varies as $1/\sqrt{\omega}$), the higher spectral components generated as a manifestation of steepening of the waveform are not efficiently absorbed. Thus discontinuous solutions, modified somewhat by the attenuation and dispersion, are to be expected.

REFLECTION, STANDING WAVES, AND REFRACTION

3n-14. Reflection and Standing Waves. For plane interacting waves in lossless fluids we return to Eqs. (3n-24) to (3n-26). For perfect gases the Riemann invariants are given by

$$\mathfrak{r} = \frac{c}{\gamma - 1} + \frac{u}{2} \quad \text{(3n-102}a\text{)}$$

$$\mathfrak{s} = \frac{c}{\gamma - 1} + \frac{u}{2} \quad \text{(3n-102}b\text{)}$$

Equations (3n-26) tell us that the quantity $\mathfrak{r}$ is forwarded unchanged with speed $u + c = \frac{1}{2}(\gamma + 1)\mathfrak{r} - \frac{1}{2}(3 - \gamma)\mathfrak{s}$. Similarly, the speed for the invariant $\mathfrak{s}$ is $u - c = \frac{1}{2}(3 - \gamma)\mathfrak{r} - \frac{1}{2}(\gamma + 1)\mathfrak{s}$. The roles of independent and dependent variables can be reversed to give the following differential equation for the flow:

$$t_{\mathfrak{r}\mathfrak{s}} + N(\mathfrak{r} + \mathfrak{s})^{-1}(t_{\mathfrak{r}} + t_{\mathfrak{s}}) = 0 \quad \text{(3n-103)}$$

where $N = \frac{1}{2}(\gamma + 1)/(\gamma - 1)$. For monatomic and diatomic gases $N = 2$ and $N = 3$, respectively. An exact solution of this equation in terms of arbitrary functions $f(\mathfrak{r})$ and $g(\mathfrak{s})$ is known, but it is usually difficult to determine f and g from the initial conditions (ref. 4).

Reflection. Certain valuable information about reflection can be obtained without solving for the entire flow field. Consider the problem of reflection from a rigid wall. For the moment we need not be specific about the equation of state. Let the incident wave be an outgoing simple wave. The Riemann invariant $\mathfrak{r}$ for a particular signal in this wave is, by Eqs. (3n-21) and (3n-24),

$$2\mathfrak{r} = \lambda_i + u_i = 2\lambda_i$$

But $\mathfrak{r}$ can also be evaluated at the wall during the interaction of the incident and reflected waves: i.e.,

$$2\mathfrak{r} = \lambda_{\text{wall}} + u_{\text{wall}} = \lambda_{\text{wall}}$$

Elimination of $\mathfrak{r}$ between these two expressions gives

$$\lambda_{\text{wall}} = 2\lambda_i$$

This is an exact statement of the law of reflection for continuous finite-amplitude waves at a rigid wall: The quantity λ doubles, not the acoustic pressure.

To see what happens to the pressure, we must specify an equation of state. Take the case of a perfect gas, for which $\lambda = 2(c - c_0)/(\gamma - 1)$ (thus $c - c_0$ doubles at a rigid wall). Using Eq. (3n-11), we obtain

$$\left(\frac{p}{p_0}\right)_{\text{wall}} = \left[2\left(\frac{p_i}{p_0}\right)^{1/\mu} - 1\right]^{\mu} \tag{3n-105}$$

where $\mu = 2\gamma/(\gamma - 1)$. Now define a wall amplification factor $\mathcal{A}$ by

$$\mathcal{A} = \frac{p_{\text{wall}} - p_0}{p_i - p_0}$$

Substitution from Eq. (3n-105) gives

$$\mathcal{A} = \frac{[2(p_i/p_0)^{1/\mu} - 1]^{\mu} - 1}{p_i/p_0 - 1} \tag{3n-106}$$

An analogous result in terms of the source that generated the incident simple wave is given in ref. 43; Eq. (3n-106) was first obtained by Pfriem (ref. 44). For weak waves ($p_i - p_0 \ll p_0$) $\mathcal{A} = 2$, in agreement with linear theory. The limiting value for very strong waves is $\mathcal{A} = 2^{\mu}$ ($= 2^7$ for air), a quite startling result. It is only of passing interest, however, because a wave this strong would already have deformed into a shock by the time it reached the wall [for shocks the expression for $\mathcal{A}$ is entirely different; the limiting value for strong shocks is $\mathcal{A} = 2 + (\gamma + 1)/(\gamma - 1) = 8$ for air (ref. 4)]. In fact, the deviation from pressure doubling is small even for fairly strong waves. For an originally sinusoidal wave of sound pressure level 174 dB, the maximum deviation is about 6 percent (ref. 43).

For a pressure release surface the law of reflection for finite-amplitude waves is the same as for infinitesimal waves. To see this, evaluate τ as before, first in the incident wave ($2\tau = \lambda_i + u_i = 2u_i$) and then at the pressure-release surface ($2\tau = \lambda_{\text{surface}} + u_{\text{surface}} = u_{\text{surface}}$, since $\lambda = 0$ when $p = p_0$, $\rho = \rho_0$). The result is

$$u_{\text{surface}} = 2u_i$$

that is, the particle velocity doubles at the surface. The reflection has an interesting effect on the wave, however. Consider a finite wave train so that after interaction the reflected signal is a simple wave. To a good approximation, the acoustic pressure wave suffers phase inversion as a result of the reflection. A wave that distorts as it travels toward the surface therefore tends to "undistort" after reflection. This effect has been observed experimentally (ref. 45).

Reflection from and transmission through other types of surfaces, such as gaseous interfaces, are considered in ref. 43.

Oblique reflection of continuous waves from a plane surface has not been solved in any general way; see ref. 46 for a perturbation treatment.

Standing Waves. First consider finite-amplitude wave motion in a tube closed at one end and containing a vibrating piston in the other end. This problem is one of the few in which much experimental evidence is available (refs. 47, 48, 50). At resonance, if the piston amplitude is sufficiently high, shocks occur traveling to and fro between the piston and the closed end. Slightly off resonance, again for high enough amplitude, the waveform exhibits cusps. Below resonance the cusps occur at the troughs of the waveform, above resonance at the peaks. It would seem that such rich phenomena would have stimulated intensive theoretical treatments of the problem.

In fact, the theoretical problem has proved a difficult nut to crack. The Riemann solution [of Eq. (3n-103)] is of no avail because of the presence of shocks. There is no well-developed weak-shock theory for compound waves as there is for simple

waves. For weak waves perturbation treatments have been used (ref. 48). For strong waves one approach has been to assume the existence of shocks at the outset. The Rankine-Hugoniot relations are used to provide boundary conditions for the continuous-wave flow in between shocks (refs. 47, 49).

A more fundamental approach has been taken by Chester (ref. 42). His treatment is of general interest because of the way the effect of the boundary layer is assimilated in the one-dimensional model [see Eq. (3n-101) for an adaptation to simple waves]. An "inviscid solution" is first obtained; it contains discontinuities at and near resonance, and cusps at one point on either side of resonance. General agreement with experimental observation is thus good (ref. 50). Improved solutions are then considered in which thermoviscous effects, first in the mainstream and then in the boundary layers, are taken into account.

3n-15. Refraction. Treatments of oblique reflection and refraction at interfaces have mainly been confined to shock waves in which the flow behind the shock is basically steady. Slow, continuous refraction, such as that caused by gradual changes in the medium or by gradual variations along the phase fronts of the wave, has been treated, however (refs. 26, 51, 52). The basis of the method is ordinary ray acoustics. The propagation speed along each ray tube and the cross-sectional area of the tube are modified to take account of nonlinear effects. The approach is similar to that given in Sec. 3n-7 except that the cross-sectional area of the horn varies in a manner that depends on the wave motion.

Acknowledgment. Support for the preparation of this review came from the Aeromechanics Division, Air Force Office of Scientific Research.

References

1. Vincenti, W. G., and B. S. Baldwin, Jr.: *J. Fluid Mech.* **12**, 449–477 (1962).
2. Stokes, G. G.: *Phil. Mag.*, ser. 4, **1**, 305–317 (1851).
3. Blackstock, D. T.: Approximate Equations Governing Finite-amplitude Sound in Thermoviscous Fluids, *Suppl. Tech. Rept.* AFOSR-5223 (AD 415 442), May, 1963.
4. Courant, R., and K. O. Friedrichs: "Supersonic Flow and Shock Waves" Interscience Publishers, Inc., New York, 1948.
5. Earnshaw, S.: *Trans. Roy. Soc.* (*London*) **150**, 133–148 (1860).
6. Riemann, B.: *Abhandl. Ges. Wiss. Göttingen, Math.-Physik. Kl.* **8**, 43 (1860), or "Gesammelte Mathematische Werke," 2d ed., pp. 156–175, H. Weber, ed., Dover Publications, Inc., New York, 1953.
7. Poisson, S. D.: *J. École Polytech.* (*Paris*) **7**, 364–370 (1808). However, Poisson's solution is for the special case of a constant-temperature gas, which in our notation corresponds to $\beta = 1$.
8. Blackstock, D. T.: *J. Acoust. Soc. Am.* **34**, 9–30 (1962).
9. Stokes, G. G.: *Phil. Mag.*, ser. 3, **33**, 349–356 (1848).
10. Fubini, E.: *Alta Frequenza* **4**, 530–581 (1935). Fubini was the first to render the Fourier coefficients in terms of Bessel functions. He used Lagrangian coordinates, not Eulerian as in the derivation here, and attempted to calculate some of the higher-order terms. The mathematical similarity of this problem to Kepler's problem in astronomy is discussed in ref. 8.
11. Gol'dberg, Z. A.: *Akust. Zh.* **6**, 307–310 (1960); English translation: *Soviet Phys.—Acoust.* **6**, 306–310 (1961).
12. Thurston, R. N., and M. J. Shapiro: *J. Acoust. Soc. Am.* **41**, 1112–1125 (1967).
13. Breazeale, M. A., and Joseph Ford: *J. Appl. Phys.* **36**, 3486–3490 (1965).
14. Compare Eq. (3n-51) with Eq. (1), p. 481 in H. Lamb, "Hydrodynamics" 6th ed., Dover Publications, Inc., New York, 1945.
15. Laird, D. T., E. Ackerman, J. B. Randels, and H. L. Oestreicher: Spherical Waves of Finite Amplitude, *WADC Tech. Rept.* 57–463 (AD 130 949), July, 1957.
16. Blackstock, D. T.: *J. Acoust. Soc. Am.* **36**, 217–219 (1964).
17. Naugol'nykh, K. A., S. I. Soluyan, and R. V. Khokhlov: *Vestn. Mosk. Univ. Fiz. Astron.* **4**, 65–71 (1962) (in Russian).
18. Naugol'nykh, K. A., S. I. Soluyan, and R. V. Khokhlov: *Akust. Zh.* **9**, 54–60 (1963); English translation: *Soviet Phys.—Acoust.* **9**, 42–46 (1963).
19. Akulichev, V. A., Yu. Ya. Boguslavskii, A. I. Ioffe, and K. A. Naugol'nykh: *Akust. Zh.* **13**, 321–328 (1967); English translation: *Soviet Phys.—Acoust.* **13**, 281–285 (1968)

20. Cole, R. H.: "Underwater Explosions," Dover Publications, Inc., New York, 1965.
21. Taylor, G. I.: *Proc. Roy. Soc. (London)*, ser. A, **186**, 273–292 (1946).
22. Naugol'nykh, K. A.: *Akust. Zh.* **11**, 351–358 (1965) English translation: *Soviet Phys.—Acoust.* **11**, 296–301 (1966).
23. This solution has been derived by G. B. Whitham, *J. Fluid Mech.* **1**, 290–318, (1956), on a somewhat different basis.
24. Landau, L. D.: *J. Phys. U.S.S.R.* **9**, 496–500 (1945).
25. Friedrichs, K. O.: *Commun. Pure Appl. Math.* **1**, 211–245 (1948).
26. Whitham, G. B.: *Commun. Pure Appl. Math.* **5**, 301–348 (1952).
27. Blackstock, D. T.: *J. Acoust. Soc. Am.* **39**, 1019–1026 (1966).
28. Rudnick, I.: *J. Acoust. Soc. Am.* **30**, 339–342 (1958).
29. Lighthill, M. J.: In "Surveys in Mechanics," pp. 250–351, edited by G. K. Batchelor and R. M. Davies, eds., Cambridge University Press, Cambridge, England, 1956.
30. See, for example, H. W. Marsh, R. H. Mellen, and W. L. Konrad, *J. Acoust. Soc. Am.* **38**, 326–338 (1965).
31. Mendousse, J. S.: *J. Acoust. Soc. Am.* **25**, 51–54 (1953).
32. Hayes, W. D.: "Fundamentals of Gas Dynamics," chap. D, H. W. Emmons, ed., Princeton University Press, Princeton, N.J., 1958.
33. Pospelov, L. A.: *Akust. Zh.* **11**, 359–362 (1965); English translation: *Soviet Phys.—Acoust.* **11**, 302–304 (1966).
34. Soluyan, S. I., and R. V. Khokhlov: *Vestn. Mosk. Univ. Fiz. Astron.* **3**, 52–61 (1961) (in Russian).
35. Blackstock, D. T.: *J. Acoust. Soc. Am.* **36**, 534–542 (1964).
36. Gol'dberg, Z. A.: *Akust. Zh.* **2**, 325–328 (1956); **3**, 322–328 (1957); English translation: *Soviet Phys.—Acoust.* **2**, 346–350 (1956); **3**, 340–347 (1957).
37. Fay, R. D.: *J. Acoust. Soc. Am.* **3**, 222–241 (1931). Fay was concerned with a viscous gas.
38. Truesdell, C. A.: *J. Math. Mech.* **2**, 643–741 (1953).
39. Polykova, A. L., S. I. Soluyan, and R. V. Khokhlov: *Akust. Zh.* **8**, 107–112 (1962); English translation: *Soviet Phys.—Acoust.* **8**, 78–82 (1962).
40. Soluyan, S. I., and R. V. Khokhlov: *Akust. Zh.* **8**, 220–227 (1962); English translation *Soviet Phys.—Acoust.* **8**, 170–175 (1962).
41. Ref. 14, art. 360*b*.
42. Chester, W.: *J. Fluid Mech.* **18**, 44–64 (1964).
43. Blackstock, D. T.: Propagation and Reflection of Plane Sound Waves of Finite Amplitude in Gases, *Harvard Univ. Acoust. Res. Lab. Tech. Mem.* 43 (AD 242 729), June, 1960.
44. Pfriem, H.: *Forsch. Gebeite Ingenieurw.* **B12**, 244–256 (1941).
45. See, for example, R. H. Mellen and D. G. Browning: *J. Acoust. Soc. Am.* **44**, 646–647 (1968).
46. Shao-sung, F.: *Akust. Zh.* **6**, 491–493 (1960): English translation: *Soviet Phys.—Acoust.* **6**, 488–490 (1961).
47. Saenger, R. A., and G. E. Hudson: *J. Acoust. Soc. Am.* **32**, 961–970 (1960).
48. Coppens, A. B., and J. V. Sanders: *J. Acoust. Soc. Am.* **43**, 516–529 (1968).
49. Betchov, R.: *Phys. Fluids* **1**, 205–212 (1958).
50. Cruikshank, D. B.: An Experimental Investigation of Finite-amplitude Oscillations in a Closed Tube at Resonance, *Univ. Rochester Acoust. Phys. Lab. Tech. Rept.* AFOSR 69-1869 (AD 693 635), July 31, 1969.
51. Whitham, G. B.: *J. Fluid Mech.* **2**, 145–171 (1957).
52. Friedman, M. P., E. J. Kane, and A. Sigalla: *AIAA Journal* **1**, 1327–1335 (1963).
53. Westervelt, P. J.: *J. Acoust. Soc. Am.* **35**, 535–537 (1963).
54. Thuras, A. L., R. T. Jenkins, and H. T. O'Neil: *J. Acoust. Soc. Am.* **6**, 173–180 (1935).
55. Muir, T. G.: "An analysis of the parametric acoustic array for spherical wave fields," Ph.D. dissertation, University of Texas at Austin, Texas (1971).
56. Bellin, J. L. S. and R. T. Beyer: *J. Acoust. Soc. Am.* **34**, 1051–1054 (1962).
57. See, for example, Berktay, H. O.: *J. Sound Vib.* **5**, 155–163 (1967).
58. Lester, W. W.: *J. Acoust. Soc. Am.* **40**, 847–851 (1966).

ATMOSPHERIC ACOUSTICS

HENRY E. BASS, *Physical Acoustics Research Group, University of Mississippi, University, Mississippi, U.S.A.*

	Introduction	145
1.	**Sound Sources**	147
1.1	Aircraft Noise	147
1.2	Traffic Noise	149
1.3	Explosions	150
1.4	Supersonic Projectiles	151
1.5	Sounds in Nature	151
2.	**Propagation**	154
2.1	Absorption	154
2.2	Effects of the Ground Surface	156
2.3	Refraction and Diffraction	160
2.4	Effects of Topography	162
2.5	Effects of Turbulence	164
2.6	Nonlinear Effects	169
2.7	Infrasound Propagation	169
2.8	Numerical Techniques	170
3.	**Applications**	172
3.1	Acoustic Sensing of the Atmosphere	173
3.2	Perception of Sound	176
	Glossary	178
	Works Cited	179
	Further Reading	179

INTRODUCTION

The science of acoustics encompasses the production, transmission, and reception of sound. When applied to the atmosphere each of these fundamental aspects of acoustics is affected to some extent by the properties of the atmosphere itself. Atmospheric properties vary in both space and time giving rise to a close interaction between atmospheric acoustics and the broader field of atmospheric physics. The interaction of sound with this complex atmosphere as well as the surface of the earth requires that a wide spectrum of physical phenomena be understood to describe completely a sound field in the atmosphere.

The need to communicate audibly in the atmosphere has traditionally motivated studies of atmospheric acoustics. The physical phenomena which affect atmospheric acoustics were generally identified in the 1800s although many of the details and mathematical techniques for dealing with these phenomena have evolved more recently. An excellent example of early experimental work in atmospheric acoustics was that of Tyndall using the large horn shown in Fig. 1.

The large size of the horn results from the relatively low density [1.29 kg/m^3 at standard temperature and pressure (STP): 273 K, pressure of 1 atm with 0% relative humidity] and speed of sound in air (331.6 m/sec at STP with 0% relative humidity). More recently, studies of atmospheric acoustics have been motivated by concerns for environmental noise. Most activities associated with modern society result in some noise but perhaps none is more pervasive than that associated with transportation. Noise in and around airports and along highways has led to attempts to increase transmission loss as well as reduce source generation. A detailed treatment of many of the topics addressed in this article can be found in Pierce (1981).

A variety of natural processes give rise to atmospheric sound. Some of the more common sources are thunder associated with lightning, noise generated by wind, and sounds produced by animals of all types.

Sound is often classified in terms of amplitude and frequency. Amplitudes of acoustic waves are generally given in dB ref 20 micropascal (μPa). A very quiet anechoic chamber might have a sound level of 0 dB. Outdoors on a quiet day, the sound level is typically 40 dB. Above 90 dB, audible sound begins to result

Encyclopedia of Applied Physics, Vol. 2

© 1991 VCH Publishers, Inc.
1-56081-061-0/91/5.00+.50

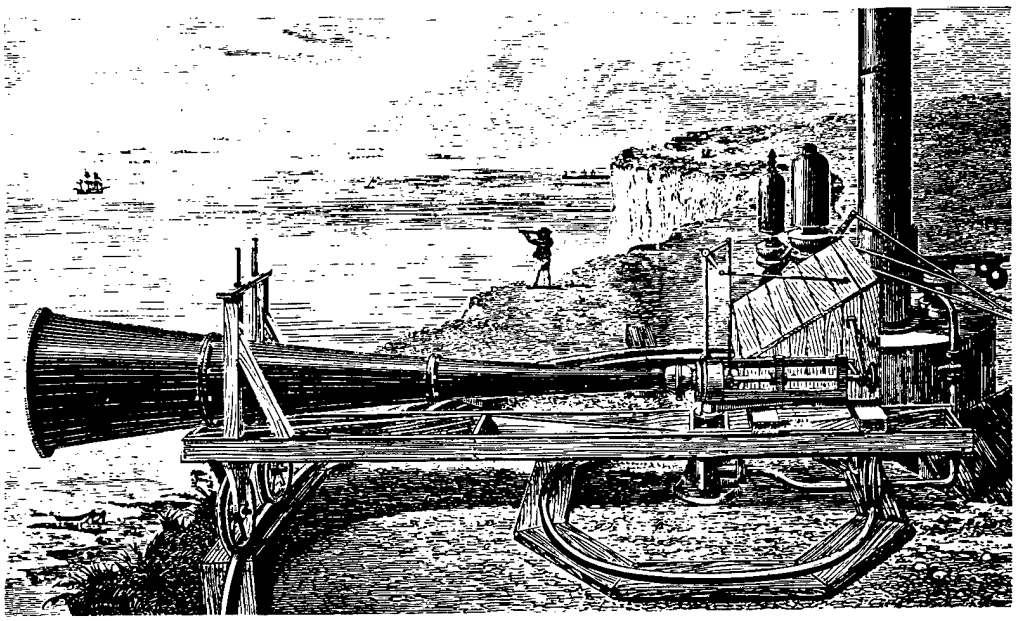


FIG. 1. Tyndall's apparatus for studying sound propagation in the atmosphere (from Tyndall, 1874).

in hearing loss in the case of very long exposures. A jet aircraft taking off might produce 120 dB. Above 130 dB, acoustic waves begin to exhibit nonlinear behavior. Above 140 dB, immediate and permanent hearing loss occurs.

Another way to categorize sound is according to frequency. Acoustic waves with a frequency above 20 Hz (units of sec^{-1}) but below 20 kHz are referred to as audible sound. Lower frequencies are classified as infrasound; higher frequencies as ultrasound. At low frequencies acoustic waves couple into the buoyant restoring force of gravity in the atmosphere and two modes result (Brown, 1987). At very low frequencies ($\sim 10^{-2}$ Hz) the propagation is dominated by the buoyant force (internal gravity waves). At the opposite end of the frequency spectrum, ultrasonic waves are rapidly attenuated in the atmosphere with an attenuation coefficient approximately proportional to frequency squared. As a result, ultrasound does not travel over great distances although over short distances it can be important (echolocation used by bats and echoranging by cameras, for example).

Once sound is generated, there are a variety of physical processes which affect transmission. The most obvious of these is spreading of the wave. The amplitude of a spherically spreading wave decreases as $1/r$ (6 dB per doubling of distance). Even highly directional sources (such as a bat) begin to show this type of spreading some distance from the source. For this reason, 6 dB/doubling of distance is generally taken as a standard loss and all other factors which affect the received amplitude are referred to in terms of excess attenuation (attenuation in addition to 6 dB/doubling of distance).

Physical processes which contribute to excess attenuation are molecular absorption, interaction with the ground, refraction due to temperature and wind gradients, diffraction due to those gradients or topography, and scattering. In all cases, except for the ground interaction (and even there to some degree), the effects depend strongly on specific local atmospheric conditions. This strong dependence of propagation on meteorological condition can be used as a basis to sense the atmosphere remotely using echo sounders or radio sounders.

Application of atmospheric acoustics usually defines the receiver. The detection and location of aircraft might involve an array of microphones sensitive to phase changes between individual elements. Noise control applications typically consider a human ear as

the receiver. These applications plus the atmospheric sounders already mentioned continue to provide motivation for studies of atmospheric acoustics.

1. SOUND SOURCES

A number of sources of atmospheric sound were mentioned in the Introduction. In this section, specific sources will be considered in more detail to illustrate the different phenomena which give rise to sound and the atmospheric properties which affect the sound generated. First, man-made sources will be considered, then natural sources.

1.1 Aircraft Noise

Jet aircraft represent a pervasive noise source. Aircraft noise is generated by both propulsive and nonpropulsive sources (Smith and Schien, 1989). Propulsion sources normally result in the dominant noise. Propulsion sources include turbomachinery noise generated internal to the engine and jet exhaust mixing and shock noise attributed to processes external to the engine. Nonpropulsive noise is generated in flight by the air flowing past the vehicle structure and is called airframe noise.

Turbomachinery noise sources include fan noise, compressor noise, combustion noise, and turbine noise.

Fan noise is caused by the unsteady component of the flow which interacts with rotor or stator blade rows. When a perturbation in flow pressure or velocity encounters a blade or strut surface, an acoustic pressure pulse is generated. Primary fan noise generation mechanisms are

1. inlet flow inhomogeneities interacting with the rotor blades, and
2. rotor wakes and pressure fields interacting with the stator blades.

Sound is generated primarily at the blade passing frequency and its harmonics, although there is a broadband component as well.

Key parameters affecting fan noise generation are the number of rotor and stator blades, rotor-stator axial spacing, blade loading, fan rpm, stage pressure ratio, and rotor tip speeds. Upstream stators (inlet guide vanes) increase acoustic levels. At high engine power settings, supersonic tip speeds give rise to tones at harmonics of the shaft rotation speed, which are known as multiple pure tones.

Compressor noise is similar in nature to fan noise, and has the same key parameters as the fan, applied to the compressor stages. Compressor noise is usually of higher frequency than fan noise and is attenuated as it passes through adjacent blade rows. Similar in nature to fan and compressor noise, turbine noise is generated downstream of the combustor and radiates from the core exhaust duct.

Instabilities in combustors give rise to sound waves. These instabilities are usually a result of flow turbulence, so that combustion noise is broadband in nature, normally peaking at 400–500 Hz. Key engine operating parameters affecting combustion noise generation are engine mass flow rate, compressor pressure and temperature ratio, and combustor temperature rise.

Jet exhaust noise is generated by two mechanisms in the jet plumes aft of the exhaust nozzles: turbulent mixing and shock cell perturbations.

Jet mixing noise is generated by the turbulent mixing process between the exhaust flow and the atmosphere. Noise generation is strongest in the turbulent shear layer between the laminar core and the ambient air. This source correlates most strongly with the velocity of the jet at the plane of the exhaust nozzle, varying as approximately the eighth power of the velocity. Secondary parameters are the jet flow density and exhaust nozzle area. Forward motion of the aircraft has a strong attenuation effect on jet mixing noise, because it reduces the magnitude of the shear layer velocity gradient.

Jet mixing noise is spectrally broadband in nature, usually peaking at frequencies below 1000 Hz, depending on the nozzle diameter and jet velocity. It is most intense in the downstream direction and peaks at a radiation angle of approximately $\theta = 140°$, measured from the aircraft nose (i.e., tail-to-nose vector is $\theta = 0°$).

The jet mixing noise generated by air vehicles and high-performance military aircraft is a significant source of noise. This source is difficult to suppress, since it is generated en-

tirely outside the aircraft. Most successful reductions in jet noise have been due to reductions in jet velocity, which generally requires an increase in fan bypass ratio to maintain thrust. Modern commercial aircraft (referred to as Stage 3 aircraft) have large bypass ratio engines. Such engines are visually greater in diameter.

When an air vehicle's convergent exhaust nozzles are operated supersonically, a quasi-periodic shock cell structure will form in the jet. Shock cell noise is produced by the unsteady interaction between the flow turbulence and this structure. The intensity of shock noise is a function of the Mach number (ratio of speed to total speed of sound) of the fully expanded exhaust flow.

Shock-associated noise possesses characteristics distinct from those of turbulent mixing noise. In contrast to the mixing noise radiation pattern, broadband shock cell noise will be most prominent in the forward arc ($\theta < 90°$) directions for a static engine, due to the high jet mixing noise levels. However, for high flight speed, shock noise may dominate the signal in the forward arc and at all radiation angles.

Airframe or nonpropulsive noise is generated by the airflow over the wings, fuselage, vertical and horizontal stabilizers, slats, gears, or any element of an aircraft. The interaction of aircraft elements with large scale turbulence in the atmosphere results in large fluctuations in lift and drag. These forces and the subsequent radiation occur with any lifting element of the aircraft. The sound power radiated from this source of noise increases with the size of the lifting surface, the turbulence intensity, and the fourth power of the aircraft speed. This source of noise has been considered as insignificant for low flight speeds but could be a significant contributor to the aircraft signature when the aircraft speed and the turbulence intensity are high. For this case the radiated power would be high and the frequencies would be low resulting in little excess attenuation due to propagation through the atmosphere.

There are four components of noise associated with radiation from the turbulent boundary layer. These include the following:

1. a surface distribution of dipoles associated with the fluctuating pressures within the flow causing fluctuating forces on a rigid surface;
2. a volume distribution of quadrupoles generated by the fluctuating forces in the boundary layer;
3. reflected quadrupole noise due to the presence of the surface;
4. noise radiation resulting from the motion of the surface beneath the turbulent layer.

Current thinking, based on commercial aircraft experience at low Mach number, considers only the surface distribution of dipoles. For such a distribution, the sound pressure is taken to vary as AU^6 (Reed, 1977) where U is the flow velocity and A is the radiating area. At low Mach numbers this source is less than that predicted for the trailing edge.

Although the surface distribution may be a valid model throughout the subsonic Mach number range there is no evidence to show that dipole-type radiation will be dominant at high subsonic Mach numbers. That is, quadrupole-type sources may become the dominant source at high subsonic Mach numbers. This has not been satisfactorily answered at this time.

The acoustic radiation from structures excited by turbulent boundary layers has been investigated primarily in terms of the noise levels in aircraft cabins and noise radiation underwater. There has been little emphasis on this source of sound with respect to airframe noise. There is something of a conflict as to its importance at low Mach numbers.

The noise radiated directly from wakes and vortices is quadrupole in nature and, as such, is not a significant source. However, when wakes or vortices impinge on a structure or rotating blade, significant levels of noise can be generated. These sources are similar to whole body noise, but at higher frequencies because of the smaller scales.

Trailing edge noise is generated by turbulence being swept over the trailing edge with the creation of half-baffled dipoles. The source of the turbulence is the turbulent boundary layer. There is a large data base at low Mach number ($M < 0.4$) which appears to confirm that trailing edge noise is proportional to the aircraft velocity to the fifth or sixth power.

Rotary wing and propeller-driven aircraft generate sound quite different from that of jet aircraft even though many of the noise generation mechanisms are the same. Machinery

noise, for example, is common to both. The biggest difference is the role that the helicopter blade or aircraft propeller plays. In many cases, for these types of aircraft, this blade noise is dominant and characterizes the audible sound.

Four mechanisms are important for noise generation provided the blade tip speed does not exceed the speed of sound. One is the discrete loading noise which dominates the lower harmonics of the blade passage frequency. Next is the blade-vortex interaction which is a discrete source of impulsive character. At higher frequencies, broadband noise due to blade interactions with turbulence in and around the rotor wake begin to dominate. Broadband self-noise due to blade interaction with boundary-layer and near wake turbulence controls the high-frequency part of the spectrum.

The relative importance of each of the above mechanisms depends upon aircraft operating conditions. In most cases, however, the aircraft is characterized by sound associated with the blade passage frequency. This frequency is variable for a propeller-driven aircraft depending upon flight operations. Rotary wing aircraft tend to operate at a lower, fixed, blade passage frequency. A military UH-1D helicopter main rotor rotates at 324 rpm and has twin blades, which gives a fundamental frequency of 10.8 Hz. The tail rotor rotates more rapidly producing a fundamental of 60.5 Hz at a level much lower than for the main blade. Since the human ear is not sensitive to 10.8 Hz, audible sound is due to harmonics of the fundamental blade passage frequency.

1.2 Traffic Noise

Another pervasive source of man-made noise is that due to traffic (Dept. of the Army, 1978). Traffic noise results primarily from the engine and exhaust system and interaction of tires with the road surface. The maximum noise emitted by an automobile increases approximately with the third power of vehicle speed. This is due primarily to tire noise created by the tire-roadway interaction. (Figure 2 illustrates automobile noise spectra at different speeds.)

The noise output of trucks is a more complicated phenomenon. First, trucks should be considered in three distinct classes according

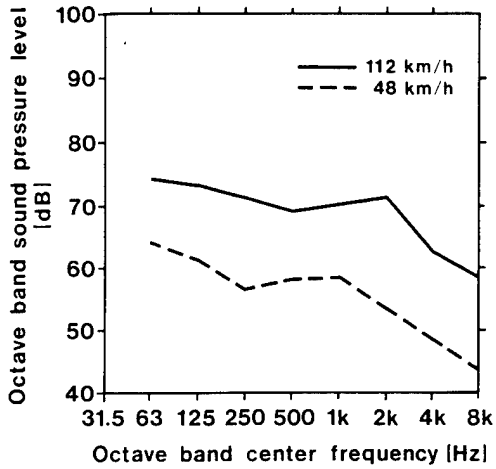


FIG. 2. Typical automobile spectra for two average speeds.

to their noise emission characteristics: light, medium, and heavy. *Light trucks* are two-axle, four-wheel vehicles such as panel and pickup trucks; their noise characteristics are similar to those of automobiles. *Medium trucks* are typically gasoline-powered two-axle, six-wheel vehicles, such as city trucks without a vertical exhaust muffler. The noise generation characteristics of these vehicles are also similar to those of automobiles. However, medium trucks are usually 10 dB noisier than automobiles for the same flow and speed.

Heavy trucks are a more complex noise source. These diesel-powered, three- or more-axle vehicles have a multitude of noise mechanisms, i.e., tire noise, exhaust noise, intake noise, engine noise, and gear noise. Shown in Fig. 3 is a typical truck noise spectrum for the three major component sources: tires, engine, and exhaust. Tire-roadway interaction, the major noise source for automobiles and light and medium trucks, occurs at ground level. For heavy trucks an additional noise source, the exhaust stack opening, is nominally located eight feet above the ground. Heavy truck noise does not exhibit great variability. While tire noise varies with speed, the engine noise sources generally show little dependence upon road speed. Furthermore, drivers tend to maintain relatively constant engine speed for all road speeds.

As previously discussed, roadway noise exposure is a function of the traffic flow and pa-

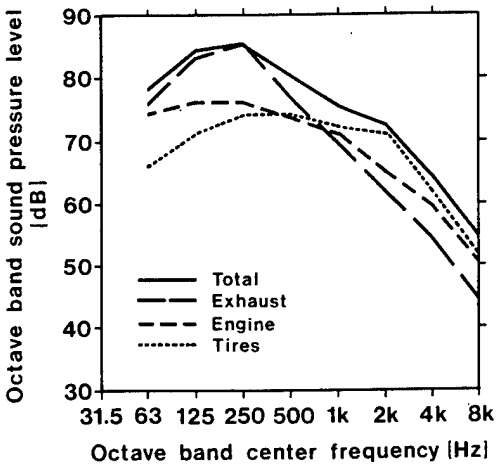


FIG. 3. Typical spectra for diesel truck and component sources.

rameters of the classes of vehicles using the roadway. Where vehicles are uniformly distributed along a single-lane roadway that is straight, infinitely long, flat, level terrain, the noise exposure is a function of the volume flow and average speed of each group of vehicles.

In practice, roadway factors will affect noise levels. Noise exposure is *increased* by uphill grades (for heavy trucks and transport vehicles). The noise exposure is decreased by buildings, land forms, or other barriers located between the roadway and the observer.

The drop-off of noise levels with distance from a roadway will typically range from 4 to 5 dB per doubling of distance. This drop-off rate is affected substantially by ground cover. Beyond two to three thousand feet, the drop-off can increase to about 6 dB per doubling of distance, due to the additional effect of atmospheric attenuation. However, the noise levels from roadway traffic will rarely be high enough to be of concern at these larger distances.

1.3 Explosions

Less common, though more dramatic, sources of sound are impulsive. Man-made explosions ranging in amplitude from nuclear blasts to the blast from a hunter's rifle can provide very large peak sound pressure levels. The variation of pressure with time from an explosive source can be approximated by the wave form shown in Fig. 4.

Near the explosion, the wave form is not so fully developed. In the far field, the most noticeable features of the impulse are the rise time, peak overpressure, and wave duration. The overpressure is given approximately as (Reed, 1977)

$$\Delta p = 0.0469 W^{b/3} R^b, \tag{1}$$

where the value 0.0469 of the coefficient for yield W is given in kg of high explosives, the distance R in km, and Δp in kPa. Values for b are determined empirically to be 1.1–1.2. The wave duration determines the fundamental frequency $\nu_f = 1/\tau$ where τ is the duration in sec. The duration for a 0.57-kg charge of C-4 is typically 0.025 sec (at 100 m). Assuming spherical spreading, ν_f should be proportional to $W^{-1/3}$. Application of these scaling laws to atmospheric conditions other than

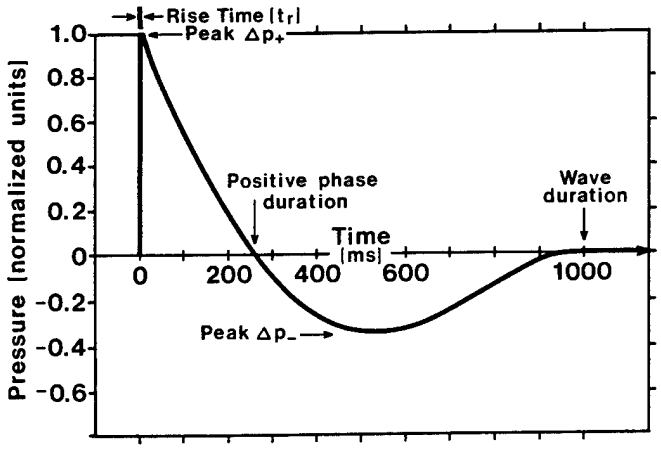


FIG. 4. Explosion-wave pressure/time signature [Reed (1977)].

STP can be accomplished, assuming the charge does PdV work against an ideal gas atmosphere.

1.4 Supersonic Projectiles

When an object is traveling at a speed v greater than the local speed of sound c, a wake forms behind the object. Since the object travels a distance vt while pressure waves travel a distance ct, a Mach cone develops ahead of which there is no sound and behind which the pressure rises very rapidly. The angle between the edge of the Mach cone and the projectile flight path is given as θ_M where $\cos\theta_M = c/v$. As the Mach cone passes over an observer on the ground, a sonic boom is heard.

The duration of the shock wave generated by supersonic projectiles depends upon the size of the projectile and the distance to the observation point. Near a projectile with simple geometry the duration $\tau \cong -l/v$ where l is the projectile length. As described later, propagation effects will cause the shock wave to increase in duration as it propagates.

1.5 Sounds in Nature

Natural processes provide a wide variety of atmospheric sounds (Brown, 1987). Referring to Fig. 5, we note that in the infrasound regime, earthquakes, large explosions (volcanoes, for example), and weather fronts give rise to very low-frequency atmospheric waves. Since these low-frequency waves are attenuated very little, they propagate great distances. It is not unusual to detect infrasound from these events after full circuits around the earth.

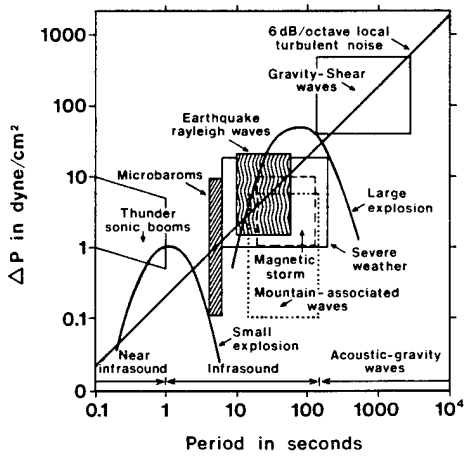


FIG. 5. Regions for infrasound sources. (Courtesy of A. J. Bedard.)

Earthquakes and other strong infrasound sources have the strange properties that the initial part of the received wave can arrive as much as a half hour before the main feature. The explanation for this effect is that high-speed Rayleigh waves traveling through the surface of the earth at a speed much greater than the speed of sound in air constantly leak energy into the atmosphere which can be detected long before the infrasound wave propagation through the atmosphere. It is frequently reported that birds and animals are startled by such arrivals, giving a warning of the upcoming infrasound.

Wind generates noise as it blows across leaves, wires, etc. The primary mechanism for the sound generation is turbulent vortex shedding. In addition to acoustic noise generated at one region and propagating to a receiver, a receiver such as a microphone or the ear responds to pressure fluctuations carried by wind. These fluctuations are not propagating but, since they result in variations in pressure as they flow past the receiver, the receiver cannot distinguish them from propagating waves. A receiver placed in the flow can also give rise to turbulent eddies providing additional noise.

Since wind noise results from the flow of turbulence, one would expect the amplitude to increase as U^2 where U is the mean flow velocity. In addition, wind noise decreases in amplitude as the frequency increases as $\nu^{-5/3}$ for moderate wind speeds. At very large wind speeds, the distribution of turbulence (discussed in more detail later) dictates fewer turbules available to flow by the wind screen resulting in a leveling off in amplitude with frequency. This effect is most pronounced at low frequencies.

An extreme case of wind noise is the sound of tornadoes (Arnold *et al.*, 1976). A model of a severe-traveling-storm cell chosen for discussion purposes is shown in Fig. 6.

Rotation of the air masses is chosen to be cyclonic in the model. The sources of sound S_i located at the coordinate (X_i, Y_i, Z_i) considered are presented in Fig. 6 and may be described as

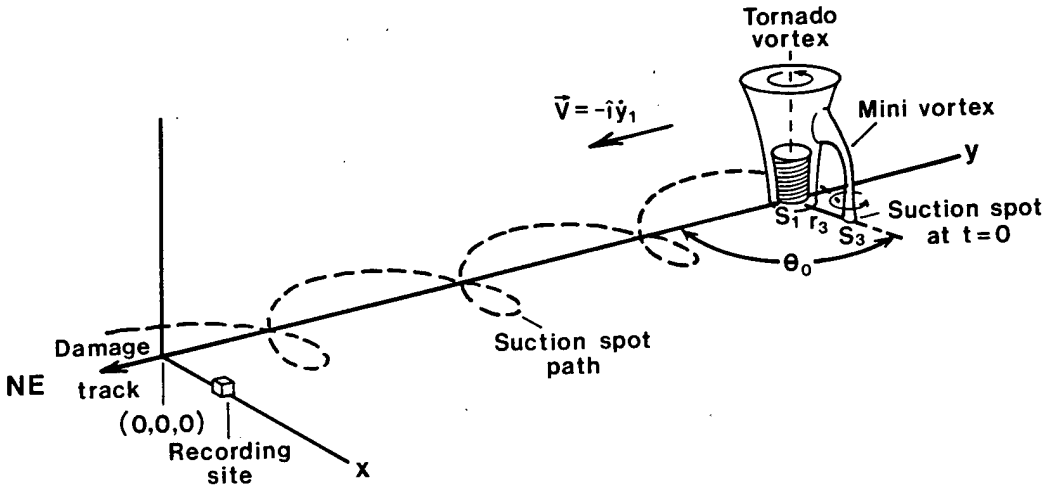


FIG. 6. Geometry assumed for the track of the Guin, Alabama, tornado and a possible minivortex. The terrain has been assumed flat although there were variations in elevation of approximately +15–20 m from the elevation at the recording site.

1. the wind-surface noise S_1 generated by interaction of the high-speed tornado winds with the ground,
2. atmospheric noise associated with the tornado vortex,
3. wind-surface noise S_3 generated by the suction spot of a minivortex,
4. atmospheric noise associated with a minivortex,
5. atmospheric noise associated with the parent storm, and
6. general nonlocalized atmospheric noises associated with the parent storm(s).

We define I_i as the acoustic power from the ith source and I_0 to be the total acoustic power (expressed in dB) when the noise (integrated over frequency) is a maximum. Maximum power probably occurs when the sound source S_1 (wind-surface noise generated by the tornado vortex) is located at or near the origin (0,0,0), and therefore the source S_1 is chosen to be located at the origin when the maximum (integrated over frequency) power is recorded at the recording site (located at position r_0,0,0).

For discussion purposes, only two principal sound sources S_1 and S_3 (which represent wind-surface interactions for the tornado vortex and suction spot, respectively) need to be considered. When the storm is most distant, the sources can be treated as spherical or point sources of sound.

One tornado studied in detail was the one which destroyed Guin, Alabama in 1974. There were no good photographic records for the Guin storm, so the distance of the suction spot from the center of the tornado vortex was assumed to be 0.6 km, and the small vortex was assumed to be moving in a circular path around the rotation axis of the tornado with a period of 60 sec. This means that the rotational (linear) speed of the suction spot was 240 km/h (150 miles/h). Combining this with the translational speed [assumed to be 96 km/h (60 miles/h)] indicates that the suction spot would have maximum wind speeds towards the recording site of 340 km/h and minimum speeds away from the receiver of 140 km/h. The high speeds involved here immediately suggest that Doppler effects should be included. For the present, however, assume that the tornado was a white-noise source. In this case, the distance from the suction spot to the receiver is given by

$$R_3 = \{[6.4 - 26t - 640\cos(\pi t/30)]^2 + [160 - 640\sin(\pi t/30)]^2\}^{1/2}. \quad (2)$$

Now consider contributions at the receiver from both sources S_1 and S_3. It should be possible to combine noise from both sources and achieve a reasonable account for the qualitative and quantitative variations in the acoustic power level as the tornado approached and receded. For convenience we have cho-

sen the level generated by each to be equal. Figure 7 presents a comparison of the relative acoustic power data and the relative power at the recording site if noise from the spherical (or point) sources S_1 and S_2 are combined. The agreement between the data and the calculated relative power is better than considering the sources separately.

Audibly, the tornado sounds like low-frequency noise (not unlike a freight train or jet aircraft landing) modulated by the rotation of the suction spot.

Thunder represents an excellent example of natural impulsive sound. As lightning discharges, the gas along the discharge channel rapidly heats and expands. This expansion propagates outward, away from the channel. To a good approximation, the frequency of maximum emission for a lightning discharge is given by (Bass, 1980)

$$\nu_n = c/[2.6(E_l/P_0\pi)^{1/2}], \quad (3)$$

where E_l is the energy per unit length dissipated in the discharge (typically 3 kJ/m) and P_0 is the ambient pressure.

Near the discharge channel, the pressure wave is a cylindrically spreading, highly nonlinear wave. For a truly cylindrical wave, all observers would hear a single clap associated with the point along the discharge path nearest to them. Real lightning, however, is very tortuous. This tortuosity gives rise to a "string of pearls" effect where thunder can be modeled as a long line of spherically spreading sources. As the thunder propagates, nonlinearity and absorption shift the spectrum to low frequencies. These combined effects give rise to the extended low-frequency rumbles associated with distant lightning and explain the sharp crack heard when lightning strikes nearby.

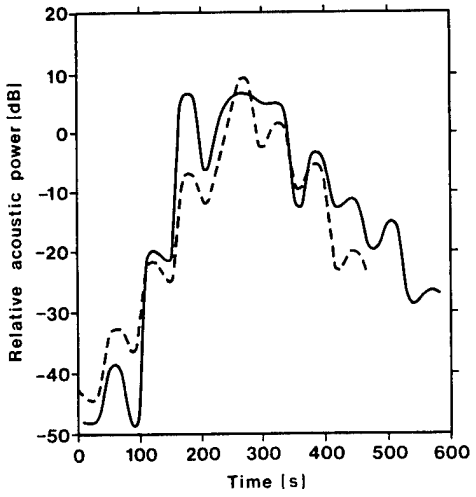


FIG. 7. Temporal variations in the relative acoustic power at 350 Hz for the Guin, Alabama, tornado compared to the variation expected from point sources S_1 and S_3. Solid line represents experimental data corrected for automatic volume control and the dashed line is predicted.

The animal world produces diverse sounds for communications (Busnel, 1977). Many animals use these sounds for warning and mating signals. The range of emission organs is quite varied and are found on all parts of the body. Invertebrates may use a toothed file to stimulate a vibrating element, vibration of nonspecialized wings (mosquitoes), semiresonant plates activated by muscle contractions, or reedlike organs activated by aspiration and respiration of air. In higher vertebrates specialized organs generate sound using forced or aspirated air. The frequency of the sound generated is controlled by varying the geometry or tension on the organs. At this point in time, sound generation mechanisms are well understood but there has been very little research which carefully characterizes total sound energy emitted. The recent discovery of communications between elephants via infrasound further emphasizes our general lack of knowledge.

Whether an animal equipped with sound emission organs will or will not activate them depends upon a number of factors such as temperature, humidity, light, and hormonal state. When sound is emitted, it follows a code involving rhythm, pulse, repetition rate, amplitude variation, and frequency. In lower animals, many of these quantities are controlled by the physical properties of the sound-producing mechanisms. In higher-order species, the code is more controlled by the central nervous system.

Messages broadcast by animals contain two specific elements of information. The first indicates presence and in a general sense, location. Hierarchical status is included for gregarious species. The second element of the message is motivation. It might also contain information giving location of an object, a territory, or a predator. Vocabularies in the animal kingdom vary widely in their complexity.

Perhaps the most interesting application of acoustics by animals is the sonar used by bats to capture flying prey. The sonar operates in two modes, search and home. In the search mode, the 35-kHz signal propagates reasonably well and has a beam width sufficiently wide to cover a large volume. Once an echo indicates a potential meal, the bat switches to a high frequency chirp (40–90 kHz). The higher frequencies provide better spatial resolution, and as the bat approaches the target, the added atmospheric absorption characteristic of high frequencies is less important. Spatial resolution automatically improves as the range to target decreases.

2. PROPAGATION

2.1 Absorption

Atmospheric absorption occurs by two basic mechanisms (Bass *et al.*, 1984):

1. classical losses associated with the transfer of acoustic energy (or the organized systematic motion associated with the kinetic energy of the molecules) into equivalent heat energy (uncoordinated random motion associated with thermal agitation) by a redistribution of translational energy among the molecules; and
2. relaxation losses associated with the redistribution (relaxation) of translational or internal energy of the molecules.

A complete calculation of sound absorption would necessarily include not only the contribution of each mechanism singly but also their interdependence; e.g., the effect of molecular relaxation processes on the classical loss mechanisms. Fortunately, it has been demonstrated by successive approximations to a solution to the Boltzmann equation and measurements at low pressures that, for frequencies below 10 MHz, absorption due to classical losses and molecular relaxation are additive. The effects of absorption will be represented by an absorption coefficient $\alpha(m^{-1})$ such that the acoustic pressure $P = P_0 e^{-\alpha r}$. Often, the absorption coefficient is given with units of neper/m. The neper is not a unit in the normal sense but, rather, specifies that it is associated with the Naperian base e. The component of absorption due to viscous effects is

$$\alpha = (\omega^2/2\rho_0 c^3)(4\mu/3), \tag{4}$$

where ω is the angular frequency ($2\pi f$) in s^{-1}, c is the speed of sound in m/s, ρ_0 is the density in kg m^{-3}, and μ is the coefficient of viscosity in kg m^{-1} s^{-1}.

Energy loss from the sound wave can be separated into two distinct processes. First, since the sound wave causes a sinusoidal variation in temperature in space, heat will be conducted from the regions of high temperature to those of lower temperature. This heat conduction process converts the energy of the sound wave into random thermal motion of the gas molecules. Heat conduction increases the absorption to

$$\alpha = (\omega^2/2\rho_0 c^3)[4\mu/3 + (\gamma - 1)\kappa/(\gamma c_v)], \tag{5}$$

in m^{-1}, where κ is the coefficient of thermal conductivity in J (kg mol)$^{-1}$ K^{-1} kg m^{-1} s^{-1}, and c_v is the specific heat at constant volume (J/K).

In addition to energy loss due to heat conduction, energy can also be lost due to relaxation of internal degrees of freedom. As the local temperature increases, the energy of internal modes also increases. If the temperature changes very slowly, the internal energy changes by an amount equal to $c_{int}\Delta T$, where c_{int} is the specific heat of the internal energy modes (e.g., due to rotational and vibrational processes), and ΔT is the local difference between the ambient temperature and the temperature in the presence of the sound wave. For low frequencies, any energy which goes into the internal modes while the local temperature is at a maximum will be diminished as soon as the local temperature begins to decrease. However, this transfer of energy from the translational energy of the gas to internal energy modes and back can only take place during collisions between the gas molecules. During any one collision, there is the probability of energy being transferred which depends on the particular internal mode of concern. So, as the frequency is increased, the process of transferring energy back and forth between internal energy modes and translation introduces a time lag between the time needed for the local temperature to decrease

and the time needed for the internal energy modes to give energy back to the translational modes. This time delay results in a decrease in the energy of the sound—an attenuation. For a particular energy mode, at very low frequencies, no absorption will result. As the frequency is increased to a value referred to as the relaxation frequency f_r of the mode (or alternatively as the frequency of maximum absorption), the absorption increases. Then, as the variations in the local temperature become so rapid that the internal mode is always behind in transferring energy, the relaxation absorption becomes constant. For any particular relaxation process, the absorption can be expressed by a relation having the form

$$\alpha=\frac{\pi s_v}{c}\frac{f^2/f_r}{1+(f/f_r)^2} \tag{6}$$

in Np m^{-1}, where s_v is the relaxation strength in nepers, which depends on the specific heat of the relaxing mode.

For air, three relaxation processes are important; those associated with O_2 vibration, N_2 vibration, and rotation. Rotational relaxation occurs at very high frequencies so that f/f_r in the denominator of Eq. (11) can be ignored below 10 MHz at 1 atm of pressure (limit decreases in proportion to the local pressure). In this case, rotational relaxation has the same form as classical absorption and the two can be combined into a single term. Using atmospheric values for density, viscosity, and the speed of sound, the combined term is

$$\alpha_{cr}=1.83\times10^{-11}\frac{(T/T_0)^{1/2}f^2}{(P/P_0)}, \tag{7}$$

where $T_0=293.15$ K and $P_0=1$ atm.

Oxygen and nitrogen relaxation depend strongly on the mole fraction of water vapor. In terms of the relative humidity h_r, in percent, the mole fraction, h_f in percent, can be written as

$$h_f=h_r(P_{sat}/P_0). \tag{8}$$

The ratio of saturated vapor pressure to atmospheric pressure (P_{sat}/P_0) adopted by the World Meterological Organization is

$$\begin{aligned}\log_{10}(P_{sat}/P_0)=&10.795\ 86[1-(T_{01}/T)]-5.028\ 08\log_{10}(T/T_{01})\\&+1.504\ 74\times10^{-4}\{1-10^{-8.296\ 92[(T/T_{01})-1]}\}\\&+0.428\ 73\times10^{-3}\{10^{4.769\ 55[1-(T_{01}/T)]}-1\}-2.219\ 598\ 3,\end{aligned} \tag{9}$$

where T is the air temperature in kelvins, and T_{01} is the triple-point isotherm temperature with the exact, internationally agreed value of 273.16 K.

The values for the relaxation frequencies of oxygen ($f_{r,O}$) and nitrogen ($f_{r,N}$) can be extracted from the experimental data. The best available values for these quantities are

$$(f_{r,N})=(P/P_0)(9+200h_f) \tag{10}$$

and

$$(f_{r,O})=(P/P_0)\{24+4.41\times10^4h_f[(0.05+h_f)/(0.391+h_f)]\}. \tag{11}$$

With these relaxation frequencies and assuming that dry air is composed of 20.9% O_2 and 78.1% N_2, the total absorption becomes

$$\alpha=f^2\left[1.83\times10^{-11}(P_0/P)(T/T_0)^{1/2}+(T_0/T)^{5/2}\left(1.278\times10^{-2}\frac{e^{-2239.1/T}}{f_{r,O}+(f^2/f_{r,O})}\right)\times\left(1.069\times10^{-1}\frac{e^{-3352/T}}{f_{r,N}+(f^2/f_{r,N})}\right)\right] \tag{12}$$

in Np m^{-1}, with $f_{r,O}$ given by Eq. (11) and $f_{r,N}$ by Eq. (10) and with the percent mole fraction h_f related to relative humidity h_r by Eqs. (8) and (9). In Eq. (12), air temperature T is in kelvins; $T_0=293.15$ K; atmospheric pressure P is in N m^{-2}; and $P_0=1.013\ 250\times10^5$ N m^{-2} (1 atm). The combined effects of these different loss terms are shown in Fig. 8.

Fogs and aerosols can introduce additional sound absorption mechanisms (Wei *et*

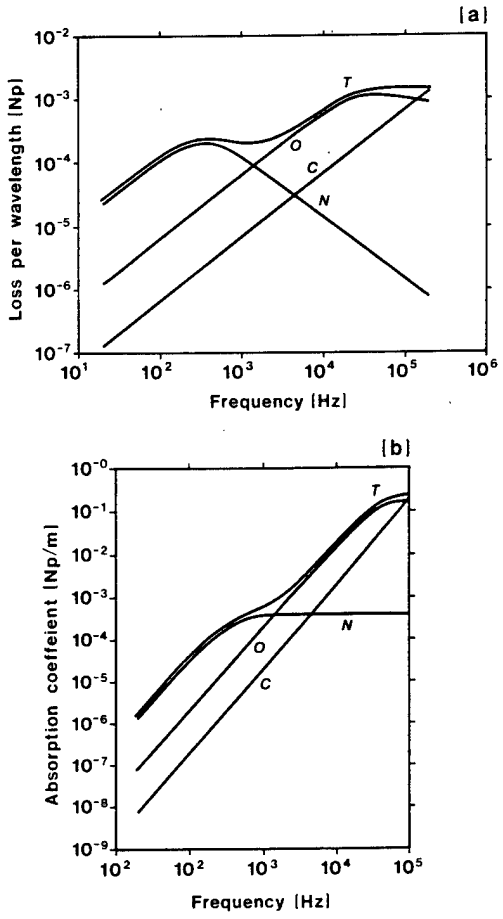


FIG. 8. Components and general behavior of total air absorption in air in terms of (a) loss per wavelength and (b) loss per unit distance. The contributions to the total absorption T, are C, classical and rotational relaxation; O, oxygen relaxation, and N, nitrogen relaxation.

al., 1987). These include evaporation and condensation around the droplets, partial entrainment of droplets by velocity fluctuation of the sound wave, heat transfer between droplets and surrounding air, the excitation of shape oscillations in the droplets, and gradients in the temperature and velocity along with heat fluxes and viscous effects. There have been little data taken in real fogs so our understanding of the mechanisms relies on theory and laboratory data.

Real fogs are composed of droplets with a wide range in size and number per unit volume. An average droplet radius is probably 8 μm and a typical ratio of liquid water content to air mass per unit volume might be 1.5×10^{-3}. A typical curve of fog absorption per unit length is given in Fig. 9 for different size droplets.

For comparison, molecular absorption at 100% relative humidity is also included. It can be seen, by comparison of these, that the two are comparable at low frequencies but that standard molecular absorption begins to dominate at high frequencies.

2.2 Effects of the Ground Surface

Research results of the past decade have greatly advanced understanding of sound propagation near the surface of the earth. Two major contributions have led to this rapid increase in understanding. The first was development of an accurate mathematical description of the sound field from a point source near a complex impedance boundary (Attenborough *et al.*, 1980). The second was the development of a microscopic description of the interaction of airborne sound with the earth (Sabatier *et al.*, 1986). Due to these advances, the physical understanding of sound propagation above plane homogeneous earth rivals that of atmospheric absorption. Referring to Fig. 10, the total acoustic field at the receiver can be written as

$$\phi_{\mathrm{TOT}}=\frac{e^{ik_1R_1}}{R_1}+\frac{e^{ik_1R_2}}{R_2}\{R(\theta_i)+B[1-R(\theta_i)]F(w)\}. \tag{13}$$

In this expression for ϕ_{TOT}, k_1 is the propagation constant for air (m^{-1}), R_1 is the direct distance from the source to receiver, and R_2 is the path length from source to receiver that

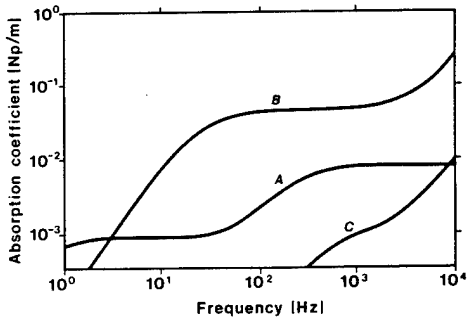


FIG. 9. Sound attenuation coefficient α_f, in units of Np/m, vs frequency. Curve A, $r=8$ μm, curve B, $r=1$ μm, curve C, molecular absorption [see Bass (1980)].

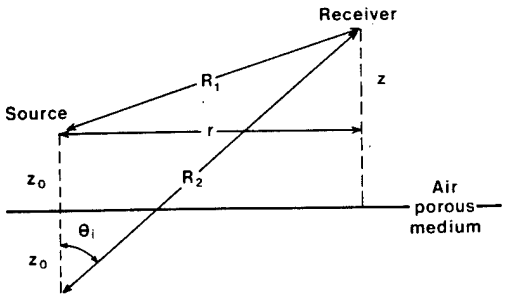


FIG. 10. Geometry of point-to-point propagation over a porous half-space.

includes one ground reflection. The quantity in curly braces is a reflection coefficient which has been modified for spherical waves reflecting from a complex-impedance flat boundary. Inside the curly braces is the usual plane-wave reflection coefficient $R(\theta_i)$, plus a term that depends on B and $F(w)$. The term B takes into account phase changes upon reflection from a complex impedance surface. For a locally reacting surface; (typical outdoors) $B \cong 1$. $F(w)$ arises from the need to match the curved wave front to a plane boundary. The quantity w is referred to as the numerical distance. It is useful to consider the numerical distance as representing the propagation distance scaled by the impedance. As w increases, wave fronts become more planar and $F(w)$ decreases.

The plane-wave reflection coefficient is given by

$$R(\theta_i)=\frac{\cos\theta_i - m(n^2-\sin^2\theta_i)^{1/2}}{\cos\theta_i + m(n^2-\sin^2\theta_i)^{1/2}}, \tag{14}$$

where the angle θ_i is the angle of incidence of the incoming wave with the ground, measured from the vertical. The symbol n is the wave-number ratio

$$n=k_2/k_1, \tag{15}$$

where k_2 is the propagation constant in the ground. The value for m may be deduced from the relation

$$\beta_c = mn = 1/Z_c, \tag{16}$$

where β_c and Z_c are known as the normalized characteristic admittance and the normalized characteristic impedance, respectively. The characteristic impedance can be written in terms of real and imaginary parts as $Z_c = R + iX$.

The equations for B and $F(w)$ are presented in their extended reaction form, both for added generality and a slight increase in accuracy. For the extended reaction case,

$$B=\left(\frac{[\cos\theta_i+\beta_c(1-\sin^2\theta_i/n^2)^{1/2}](1-n^{-2})^{1/2}}{[\cos\theta_i+\beta_c(1-n^{-2})^{1/2}/(1-m^2)^{1/2}](1-\sin^2\theta_i/n^2)^{1/2}}\right) \times\left(\frac{[(1-m^2)^{1/2}+\beta_c(1-n^{-2})^{1/2}\cos\theta_i+\sin\theta_i(1-\beta_c^2)^{1/2}]^{1/2}}{(1-m^2)^{3/2}(2\sin\theta_i)^{1/2}(1-\beta_c^2)^{1/4}}\right). \tag{17}$$

For local reaction, which is often a good approximation, $B \cong 1$. The equation for the "sphericity" factor, $F(w)$, is

$$F(w)=1+i\pi^{1/2}we^{-w^2}\,\mathrm{erfc}(-iw), \tag{18}$$

where erfc denotes the complementary error function, w is the so-called "numerical distance" defined by

$$w^2=ik_1R_2\{1+[\beta_c\cos\theta_i(1-n^{-2})^{1/2} - \sin\theta_i(1-\beta_c^2)^{1/2}]/(1-m^2)^{1/2}\}, \tag{19}$$

and all square roots are taken so that their real parts are non-negative.

The first and second terms in Eq. (13) are readily identified as the direct and reflected waves. The last term ($[1 - R(\theta_i)]F(w)$ recalling that $B \cong 1$) arises mathematically from the need to match the boundary conditions, in particular the variation in wave-front curvature with distance (Piercy *et al.*, 1977). The function $F(w)$ in Eq. (18) is plotted for various values of the phase angle [$\phi = \tan^{-1}(X/R)$] in Fig. 11.

In order to understand the third term in Eq. (13) consider the case where the source and receiver are both on the ground ($\theta_i = 90°$). For this case, $R(\theta_i)$ in Eq. (13) becomes -1 and, since $R_1 = R_2$, the first two

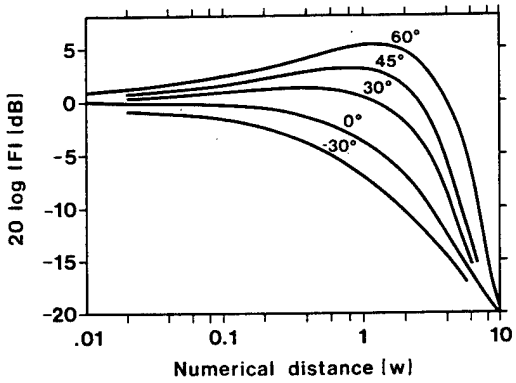


FIG. 11. Amplitude factor $F(w)$ of the ground and surface waves vs numerical distance w for various values of the phase angle ϕ of the surface impedance.

terms in Eq. (13) completely cancel to form a shadow zone. This shadow zone is penetrated by the third term. For $\phi=0$ (purely resistive surface), the third term is referred to as a ground wave in analogy to electromagnetic propagation. The curve for $\phi=0$ in Fig. 11 indicates that for short distances, the ground wave suffers no attenuation in excess of spherical spreading but for longer distances falls off with an additional 6 dB per doubling of distance.

It will be shown later that for real ground surfaces, ϕ varies between 30° and 60° so F in Fig. 11 differs from that for $\phi=0$ by the pressure of a substantial increase for $w \geqslant 1$. This increase is due to the contribution of a surface wave in the air which propagates with an amplitude that decreases exponentially with height above the boundary. For $w<1$, the contribution of the surface wave is smaller than that of the ground wave because the surface-wave amplitude decreases at only 3 dB per doubling of distance. For $w>1$, the amplitude of the surface wave again becomes smaller because of attenuation by viscous losses in the pores of the boundary.

The scale of the distance and heights of the ground and surface waves is given in Table 1 for a grassy surface. Figure 12 gives relative contributions of the various waves.

Values of w, ϕ, and $R(\theta_i)$ all depend upon the complex normalized characteristic impedance of the sound. The four-parameter model of Attenborough (1985) can be used to calculate ground impedance. For ground surfaces commonly encountered, the complex

Table 1. The propagation distance d_w for numerical distance=1, and height z_{sw} for the amplitude of the surface wave to decrease by $1/e$ from its value at the boundary, calculated from the impedance for mown grass.

f (Hz)	d_w (m)	z_{sw} (m)
50	4000	58
100	1500	23
200	270	7.3
500	24	0.9
1000	3.6	0.5
2000	1	0.15

normalized characteristic impedance of the ground can be approximated by

$$Z_c \cong \left(\frac{4q^2}{3\Omega}+\frac{iS_f^2\sigma}{\omega\rho_0}\right)\Big/ k_b, \tag{20}$$

where $q^2=\Omega^{-n'}$ and n' is a constant which relates tortuosity (q) to porosity (Ω), S_f is the pore shape factor ratio, Ω is the porosity of the ground, σ is the flow resistivity of the ground (kg s^{-1} m^{-3} or g s^{-1} m^{-3}), ω is the angular frequency (s^{-1}), ρ_0 is the density of air (kg m^{-3}), and k_b is the normalized wave number (m^{-1}).

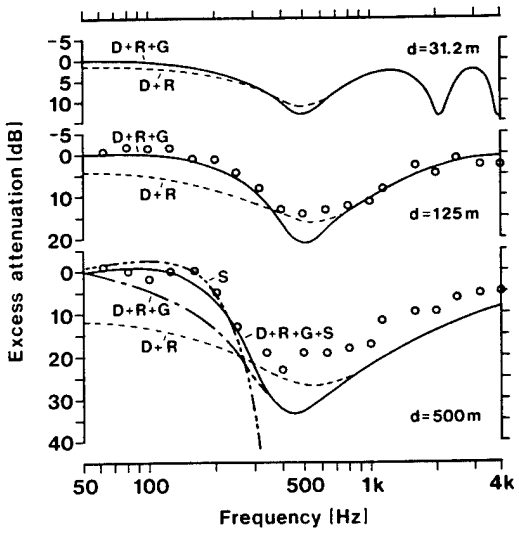


FIG. 12. Excess attenuation for propagation from a point source over mown grass, source height=1.8 m, receiver height=1.5 m. The calculated curves show the contributions from the various waves—direct D, reflected R, ground G, and surface S. The points are measurements of jet noise at comparable distances. The excess attenuation is relative to that for a point source placed on a perfectly hard surface.

The normalized wave number k_b is computed as

$$k_b \cong (\gamma\Omega)^{1/2}\left[\left(\frac{4}{3}-\frac{\gamma-1}{\gamma}N_{pr}\right)\frac{q^2}{\Omega}+\frac{iS_f^2\sigma}{\omega\rho_0}\right]^{1/2}, \tag{21}$$

where γ is the ratio of specific heats (1.4), and N_{Pr} is the Prandtl number (0.72).

The variables S_f, Ω, and n' have been varied individually until agreement between theoretical and measured values for impedance has been achieved for a variety of surfaces. As a result of such comparisons, n' is always set equal to 0.75 and S_f is always set equal to 0.875. Recommended values for porosity and flow resistivity are given in Table 2.

At intermediate to high frequencies, the semiempirical expression by Delany and Bazley (1977) and Chessel (1977),

$$Z_c=(R+iX)/[1-(k/k_\alpha)]^{1/2}, \tag{22}$$

where

$$R=[1+9.08(f/\sigma_e)^{-0.75}],$$

$$x=11.9(f/\sigma_e)-0.73, \tag{23}$$

$$k_\alpha=(\omega/c_0)[1+10.8(f/\sigma_e)^{-0.70} -i10.3(f/\sigma_e)^{-0.59}],$$

has been found to give adequate agreement between predictions and experiment. In these equations, f is in Hz and σ_e is in cgs rayls. The effective flow resistance $\sigma_e \cong S_f^2\sigma/\Omega$. For large σ_e, Attenborough's four-parameter model reduces to

$$Z_c \cong 0.218(1+i)(\sigma_e/f)^{1/2}. \tag{24}$$

Some ground surfaces have a layered structure which results from the gradual deposition of material over a soil base. The effective impedance $Z(d)$ for a layer of depth d and impedance Z_1 above a semi-infinite layer of impedance Z_2 is given by

$$Z(d)=\left(\frac{Z_2-iZ_1\tan(k_bd)}{Z_1-iZ_2\tan(k_bd)}\right)Z_1, \tag{25}$$

where k_b is the bulk propagation constant in the upper layer. The quantities Z_1, Z_2, k_b are computed from Eqs. (21) and (22) or (23).

The effect of ground reflection can be seen in Fig. 13. The excess attenuation is given relative to what it would be in the free field (positive excess attenuation means lower level). The computed curve is for a source height of 15 m, receiver height of 0.55 m, range of 500 Hz, and specific impedance Z of $9.34-8.75i$.

The first dip is the so-called ground effect dip caused by interference between direct and reflected sound paths. At frequencies above the ground effect dip, constructive interference leads to an enhanced signal level followed by additional interference minima. At very large ranges, the direct and reflected paths will be nearly equal leading to a surface-induced shadow region.

Table 2. Parameters required for the four-parameter model; n' is always set equal to 0.750 and S_f (shape factor ratio) is always set equal to 0.875.

Description of surface	Ω	σ (cgs)
Dry snow, new fallen 0.1 m over about 0.40 m old snow	0.850	23
Sugar snow	0.825	48
In forest, pine or hemlock	0.825	60
Grass: rough pasture, airport, public buildings, etc.	0.675	330
Roadside dirt, ill-defined, small rocks up to 0.1 m mesh	0.575	960
Sandy silt, hard packed by vehicles	0.475	3 470
"Clean" limestone chips, thick layer (0.01–0.025 m mesh)	0.425	6 470
Old dirt roadway, fine stones (0.05 m mesh) interstices filled	0.400	7 500
Earth, exposed and rain-packed	0.350	17 100
Quarry dust, fine, very hard-packed	0.300	41 700
Asphalt, sealed by dust and light use	0.250	120 000

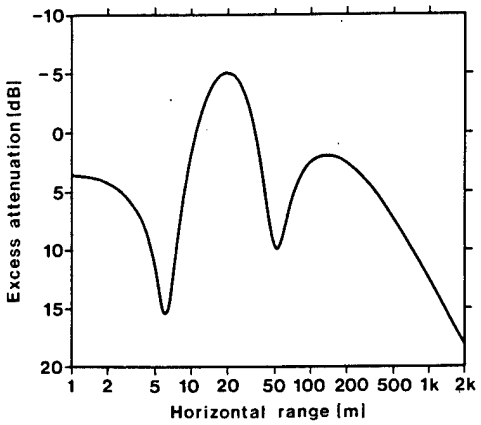


FIG. 13. Excess sound attenuation due to surface.

2.3 Refraction and Diffraction

The atmosphere is never quiet, but in some cases it is convenient to describe it in terms of temperature or, in the case of a steady wind, speed of sound as a function of altitude (Brown and Hall, 1978). Throughout the troposphere and stratosphere, the atmosphere has a relatively constant (with some variation with humidity) molecular weight of 28.97. Above infrasonic frequencies, the speed of sound c will be given by

$$c=343.3(T/293.15)^{1/2}\ \text{m s}^{-1} \tag{26}$$

and the density will be given by the ideal gas law, $\rho = P_0/RT$ where P_0 is the static pressure and R is the universal gas constant. Figure 14, from Lindsey (1974), shows the speed of sound as a function of altitude for a model atmosphere. A model atmosphere, however, seldom if ever exists (Brown and Hall, 1978). The first few kilometers above the ground undergo significant diurnal variations. Inversions caused by radiative heating at night give rise to a sound speed which in creases with altitude near the surface. Following sunrise, rapid surface cooling causes surface air to rise pushing this inversion higher before destroying it. Wind causes the speed of sound to be dependent upon direction and can eliminate inversions by mixing air from different levels.

When the speed of sound as a function of altitude varies linearly, high-frequency sound follows a curved path represented by the arc of a circle shown in Fig. 15.

Wind causes the local speed of sound to become dependent upon direction. Sound propagating upwind will be refracted upward; downwind, waves are diffracted downward. When computing the local speed of sound, the important quantity for near horizontal propagation is

$$c=c+w\cos\theta, \tag{27}$$

where θ is the angle between the direction of propagation (propagation speed c) and the wind (wind speed w). Gradients in this quantity give rise to enhanced refraction and, at low frequencies, diffraction.

For horizontally stratified media, the ray paths are given by Snell's law,

$$\frac{\cos\theta(z)}{c_z}=s=\frac{\cos\theta_s}{c_s}, \tag{28}$$

where c_z is the sound speed at altitude z, c_s is the sound speed at the source, and s is a constant. The initial angle measured from horizontal is θ_s with upward-sloping rays having positive launch angles. When the sound speed gradient ($g_s = dc/dz$) is linear,

$$\cos\theta_f=s(c_s+g_s h), \tag{29}$$

where θ_f is the angle made with the adjoining interface and h is the separation between interfaces. Solving for h yields

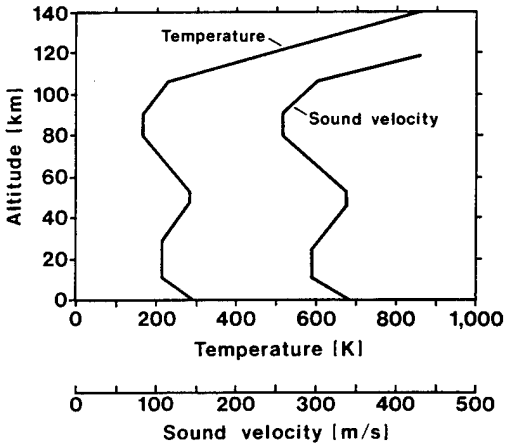


FIG. 14. Temperature and sound velocity in the model atmosphere as developed by the Air Research and Development Command in 1959. Details of the real atmosphere vary with location on the earth's surface and with the seasons. Wind speed also influences sound velocity (Lindsey, 1974).

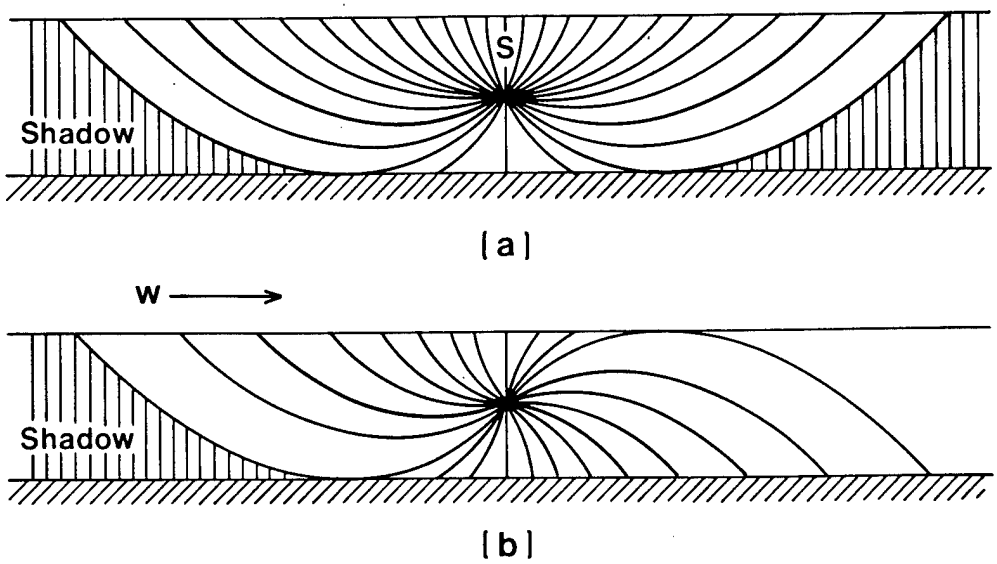


FIG. 15. Acoustic rays and shadow zones for (a) decreasing temperature and (b) increasing horizontal wind.

$$h=\frac{1}{sg_s}\cos\theta_f+\text{const.} \tag{30}$$

We note that the ray path is circular with radius $1/sg_s$. A single linear variation in c seldom applies in the first hundred meters or so. In addition, when significant (greater than 1%) variations in the speed of sound occur in a distance comparable to the acoustic wavelength, diffraction results. Ray tracing is not valid in this case and one must resort to numerical or approximate solutions to the wave equation.

The effects of diffraction are most pronounced near a shadow boundary. For an upward-refracting atmosphere, rays emanating from a source will bend upward creating an ever enlarging zone on the surface where rays cannot penetrate. In this shadow boundary, diffracted and scattered waves dominate.

A residue series has been obtained to treat the case of a stratified atmosphere where the sound speed varies linearly with height above a flat surface (Pierce, 1981; Berengier and Daigle, 1988). When $r > h_s + h$, where $z=h$ is the height of the receiver (see Fig. 16), the solution can be written

$$p(r,z)=\frac{\pi e^{i\pi/6}}{l}\sum_n H_0^{(1)}(k_n r)\,\frac{\mathrm{Ai}[b_n-(h_s/l)e^{2i\pi/3}]\mathrm{Ai}[b_n-(z/l)e^{2i\pi/3}]}{[\mathrm{Ai}'(b_n)]^2-b_n[\mathrm{Ai}(b_n)]^2}, \tag{31}$$

where

$$b_n=\tau\exp(2i\pi/3)=(k_n^2-k_0^2)l^2\exp(2i\pi/3) \tag{32}$$

are the zeros of the expression

$$\mathrm{Ai}'(b_n)+q[\exp(i\pi/3)]\mathrm{Ai}(b_n)=0. \tag{33}$$

The notation $H_0^{(1)}$ signifies the zeroth-order Hankel function of the first kind and Ai signifies an Airy function with primes denoting derivatives.

The abbreviations in Eqs. (31)–(33) are

$$q=ik_0\rho_0 lc/Z_c,$$
$$l=(R/2k_0^2)^{1/3},$$
$$\tau=(k^2-k_0^2)l^2,$$

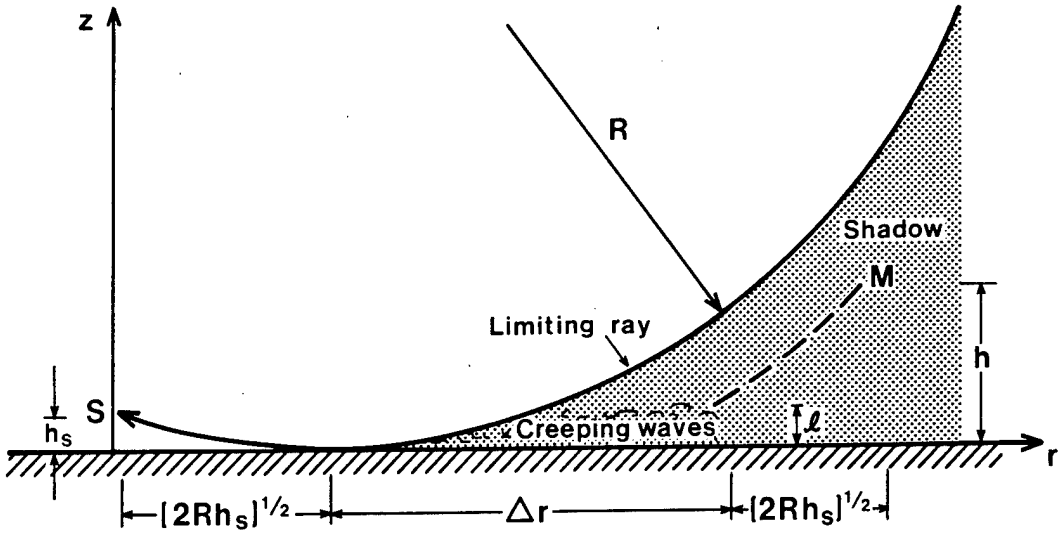


FIG. 16. Sketch showing the geometrical definitions for propagation in a stratified atmosphere above flat ground.

where $k_0 = \omega/c(0)$ and Z_c is the specific impedance of the surface.

The solutions of Eq. (33) for the b_n are simple only in the limiting cases of a rigid surface ($Z_c \approx \infty$) and that of a pressure release surface ($Z_c \approx 0$). In the intermediate cases of a surface of complex finite impedance Z_c, Eq. (33) must be solved numerically for its zeros for each value of n required in the series. The expression in Eq. (33) is a complex function of the complex variable b. Finding the zeros involves solving two nonlinear equations with two real unknowns for which a number of numerical methods exist.

The theory [Eq. (31)] is valid for propagation above a curved surface if z is interpreted as the height h transverse to the curved surface, r is interpreted as the arc length along the surface, and the radius of curvature R becomes, for the case studied here, the radius of the curved cylindrical surface.

When the receiver is deep within the shadow, but not close to the ground, the expression in Eq. (31) can be approximated by the first term in the series and, further, the Hankel and Airy functions can be approximated by their asymptotic behavior. For the purposes of this article it is sufficient to restrict the analysis to the magnitude of the pressure and these approximations yield the following expression:

$$|p(r,z)| = \left(\frac{1}{2k_0 r l^2}\right)^{1/2} \frac{1}{K(Z_c)} \left(\frac{1}{h_s}\right)^{1/4} \times \left(\frac{1}{z}\right)^{1/4} \exp[-\alpha(Z)\Delta r], \quad (34)$$

where k_0 and l are defined above. The function $K(Z)$ essentially describes the contribution of the denominator in Eq. (31) and has limiting values of $K(\infty) = 1.04$ in the case of a rigid surface and $K(0) = 1.74$ in the case of a pressure release surface. Similarly, the function $\alpha(Z)$ describes the attenuation of the creeping wave along the portion of the path Δr and takes on the following limiting values:

$$\alpha(\infty) = 0.7(k_0/R^2)^{1/3}, \quad (35)$$

in the case of a rigid surface, and

$$\alpha(0) = 1.61(k_0/R^2)^{1/3}, \quad (36)$$

in case of a pressure release surface.

As the frequency gets lower, the creeping wave approximation requires more and more terms in the expansion favoring a numerical solution to the wave equation.

2.4 Effects of Topography

Up to this point, the surface of the Earth has been assumed flat. The presence of hills, trees, buildings, and constructed sound barriers will modify, to some extent, the amplitude

of the received sound field. The sound field behind a semi-infinite, thin, rigid plate with no ground surface can be predicted using the integral equations of Fresnel. In this simple treatment, the top edge of the barrier acts as a line source diffracting energy into the region behind the barrier. In practice, the problem is much more difficult.

If one considers a long barrier (Isei *et al.*, 1980), the sound paths shown in Fig. 17 will all contribute to some extent to the field at the receiver R. The total field Φ at receiver R depends on the fields from the real source S and the image source in the ground T, and the values of these fields at the image of the receiver position X, namely

$$\Phi=\Phi_{SR}+\Phi_{TR}+\Phi_{SX}+\Phi_{TX}, \tag{37}$$

where Φ_{SR} is the direct diffracted field due to source S at receiver R (shortest path $SBR = d_{SB} + d_{BR}$), Φ_{TR} is the diffracted field due to image source T in the ground at receiver R (shortest path $TBR = d_{TB} + d_{BR}$), Φ_{SX} is the diffracted field due to image source T at image receiver X, both in the ground (shortest path $TBX = d_{TB} + d_{BX}$). More rigorously, the contribution from image source U in the barrier should not be overlooked. Then, instead of Eq. (37), we have

$$\Phi=\Phi_{SR}+\Phi_{TR}+\Phi_{SX}+\Phi_{TX}+\Phi_{UR}+\Phi_{UX}, \tag{38}$$

where Φ_{UR} is the diffracted field due to image source U in the barrier at receiver R (shortest path $UBR=SBR$) and Φ_{UX} is the diffracted

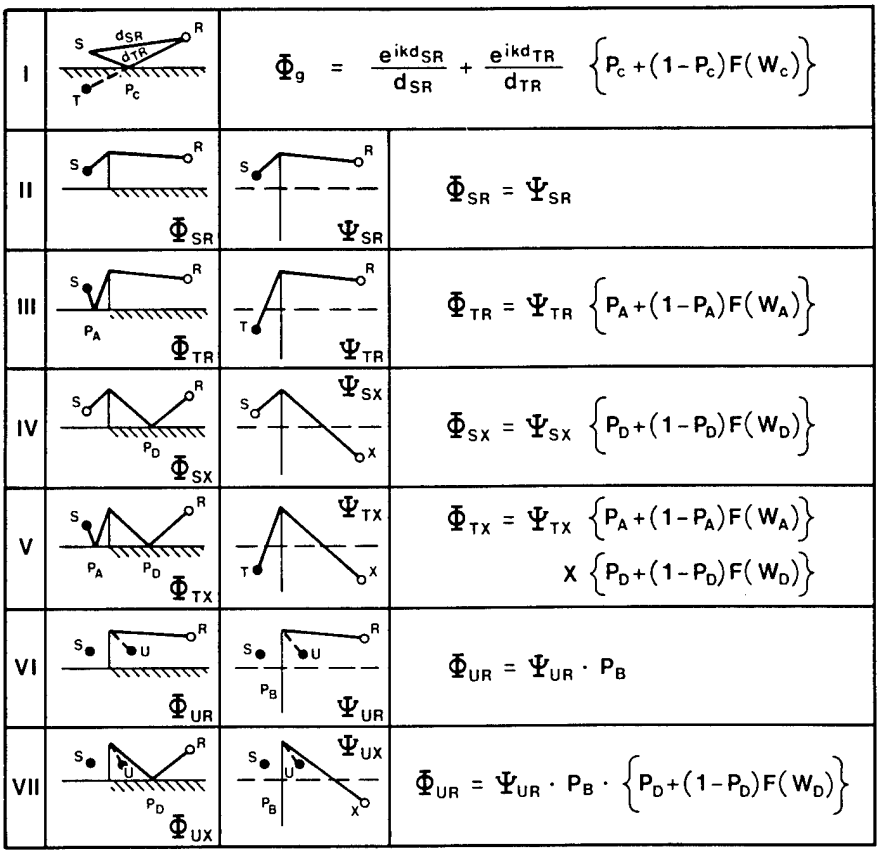


FIG. 17. Ray path diagrams and equations used for calculating the six components (lines II–VII) of the diffracted sound field at R. Line I shows how to calculate, without the barrier, the reference sound field needed for describing barrier insertion loss. Column 1 shows the actual ray path, including reflections, that relates to the field component Φ. Column 2 shows the equivalent ray path used for calculations of the diffracted field corrected for reflection and possible presence of ground waves, using the equation of column 3, in order to obtain the desired field component Φ.

field due to image source U at the image receiver X (shortest path $UBX=SBX$). Furthermore, if the receiver R is directly illuminated by the source S, or its image T, then another two fields, due to geometrical spreading from each source, should be added to the field Φ.

The evaluation of each of these terms gives rise to predictions such as that shown in Fig. 18. As expected, at low frequencies, barriers provide little shielding.

Natural barriers such as hills can be represented by wedge shaped barriers or a wedge with a horizontal top (wide barrier). Again, the largest insertion loss is observed at high frequencies.

In practical applications, the insertion loss shown in Fig. 18 near 5 kHz is seldom achieved. Refracted and scattered sound often fill in interference dips providing less attenuation. Typical results are shown in Fig. 19 (from Daigle, 1982) with highway noise as a source and a receiver on the opposite side of a barrier 1 m above an asphalt surface. Note that at high frequencies where diffraction theory predicts the greatest insertion loss, the experimental measurements liewell above theoretical predictions. Scattering degrades the effect of the barrier.

For noise control, it is often desirable to make use of plants (trees, shrubs, etc.) (Price *et al.*, 1988). At low frequencies (below~800 Hz), the primary effect of plant growth along the propagation path is a modified surface impedance due to deposited organic matter. The

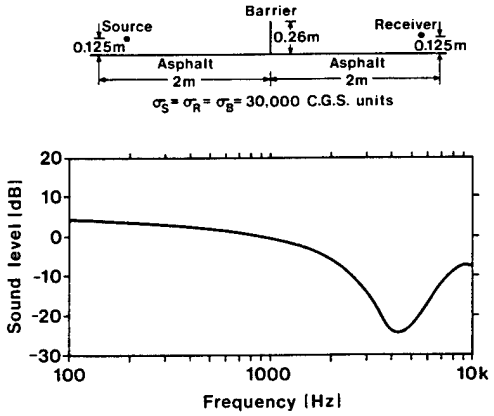


FIG. 18. Attenuation spectrum predicted by diffraction theory for the particular symmetrical configuration of source, receiver, barrier, and hard ground sketched at the top of the figure.

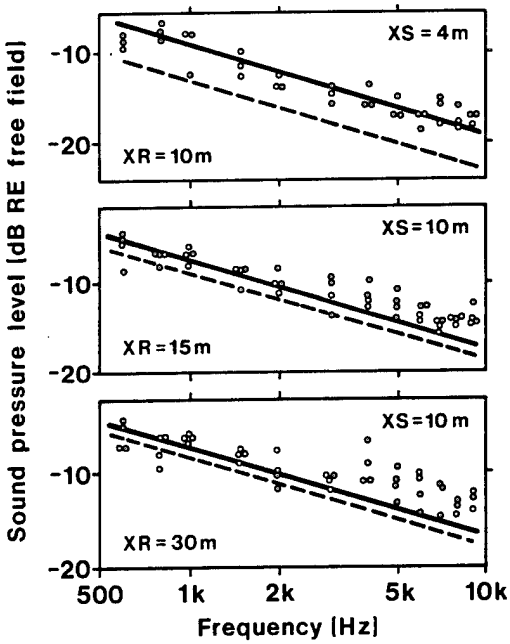


FIG. 19. The points are the measured sound pressure levels from a point source diffracted at the edge of a barrier 2.44 m high. XS and XR are, respectively, the horizontal distances from source to barrier and receiver to barrier (on hard ground—asphalt). The dashed curve is the prediction of diffraction theory.

surface flow resistance is typically decreased in such regions to below 100 cgs rayls. Further, the organic layer requires the use of a two-layer model for the surface impedance. At higher frequencies, scattering by trees can become significant for dense growths.

Even at high frequencies, however, insertion loss due to trees is typically small. Figure 20 gives results for a very dense coniferous site where measurable sound attenuation is achieved. The trees can be modeled as scattering cylinders which are assumed to be identical, vertical, and infinitely long. Multiple scattering is allowed. The use of two arrays is intended to represent scattering both by tree trunks and by foliage.

2.5 Effects of Turbulence

The sound level at a receiver is seldom the value expected assuming the mechanisms discussed to this point. The real atmosphere is far from stationary with changes in acoustic index of refraction occurring on a time scale of milliseconds to seconds due to atmo-

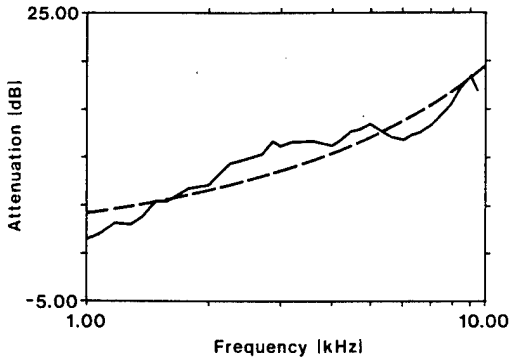


FIG. 20. Mean measured attenuation in dB per 24 m at site 3 (mixed coniferous) compared with a prediction (broken line) based upon the sum of two scattering arrays of cylinders: (a) rigid, radius 0.066 m, density 0.0724 m^{-2} and (b) nonrigid, radius 0.001 m, density 100 m^{-2}; σ_e (of surface) 860 g^{-1} cm^{-3}.

spheric turbulence. There are two sources for the turbulence. The first is that due to heating and cooling of the surface which gives rise to thermal plumes (thermally driven turbulence). The second is due to the presence of the ground surface which causes a gradient in wind leading to shear and turbulent eddy formation. In most cases, this wind-driven turbulence dominates. Acoustically, this turbulence serves to scatter sound into shadow zones and causes sound beams to spread.

First, consider a simple model of turbulence effects on the acoustic field. A simplified representation is shown in Fig. 21. As an acoustic ray travels toward the receiver, its direction and phase will be modified slightly when it encounters a turbule. The received signal will be some vector sum of these individual path segments plus any random noise at the same frequency. The next (either later or adjacent) acoustic ray will follow a slightly different path arriving with different amplitude and phase. The net result is that the received amplitude and phase will form a pattern in phase space rather than a point.

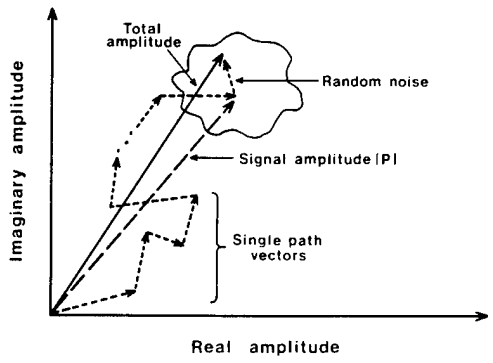


FIG. 21. Representation of acoustic ray scattered by turbulence.

If we now look along the amplitude vector, the probability (P) of observing the received sound with a given amplitude, p, can be represented by a Ricean distribution (Rayleigh at high frequencies, Gaussian at low frequencies) as shown in Fig. 22 and given by

$$P(p)=\frac{p}{\delta^2}\,(e^{-(p^2+p_0^2)/\delta^2})I_0(pp_0/\delta^2), \tag{39}$$

where p_0 is the average sound pressure, δ is a measure of the width of the distribution, and I_0 is the modified Bessel function.

It is sometimes convenient to represent phase and log-amplitude fluctuations in terms of structure functions D_s and D_x, respectively, in a plane perpendicular to the direction of propagation. For two microphones separated by distance ρ at a distance r from the source,

$$D_x(r,\rho)=\langle[x(\mathbf{r}+\boldsymbol{\rho})-x(\mathbf{r})]^2\rangle \tag{40}$$

and

$$D_s(r,\rho)=\langle[\phi(\mathbf{r}+\boldsymbol{\rho})-\phi(\mathbf{r})]^2\rangle, \tag{41}$$

where $x(\mathbf{r})$ is the log-amplitude measured at one microphone and $x(\mathbf{r}+\boldsymbol{\rho})$ is the log-amplitude measured at the second.

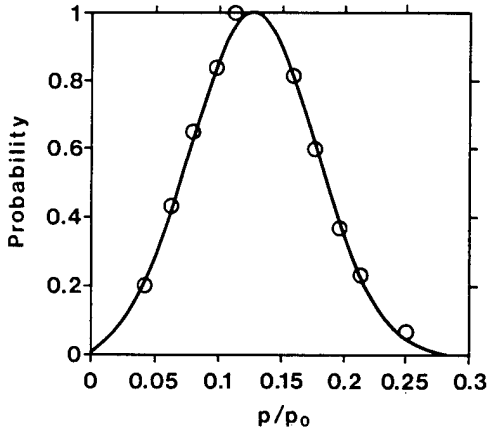


FIG. 22. Probability of observing an acoustic amplitude p, as a function of that amplitude divided by the free-field amplitude p_0. Open circles are experimental points. Solid line computed for Eq. (39) with $p_0=0.117$, $\delta=0.052$.

For $(\lambda r)^{1/2} \gg L$ where L is the scale of turbulence, simple single-scattering theory predicts that

$$D_x(r,\rho) = 2[\langle x^2 \rangle - B_x(\rho)], \tag{42}$$

$$D_s(r,\rho) = 2[\langle s^2 \rangle - B_s(\rho)], \tag{43}$$

where

$$\langle x^2 \rangle = \langle s^2 \rangle = \left(\frac{\sqrt{\pi}}{2}\right) \langle \mu^2 \rangle k^2 r L, \tag{44}$$

μ is the rms fluctuation in the acoustic index of refraction, and k is the propagation constant. $B_x(\rho)$ and $B_s(\rho)$ are, respectively, the covariances of the log-amplitude and phase fluctuations,

$$\frac{B_x(\rho)}{\langle x^2 \rangle} = \frac{B_s(\rho)}{\langle s^2 \rangle} = \frac{\Phi(\rho/L)}{\rho/L}, \tag{45}$$

where

$$\Phi(\rho/L) = \int_0^{\rho/L} e^{-u^2} du. \tag{46}$$

In practice Eq. (44) agrees with experimental results for $\langle s^2 \rangle$. However, measurements show that at large distances $\langle x^2 \rangle \ll \langle s^2 \rangle$ and, in addition, the log-amplitude fluctuations quickly saturate.

The two parameters that are used to characterize fluctuating wind speed and temperature are the strength and the scale (Johnson *et al.*, 1982). The strength of the turbulence is represented statistically by the standard deviation of the fluctuations (denoted by σ_T or σ_v for the temperature or wind speed, respectively). The scale of the fluctuations (denoted by L) is determined from the spectral density function or the correlation function.

To model the variation in scale and magnitude of convective and mechanical turbulence with altitude under unstable conditions, we must introduce two types of scaling: Monin–Obukhov scaling and mixed-layer scaling.

Monin–Obukhov's similarity theory allows the development of relationships between the turbulence variables and the surface roughness, wind speed, and convection to be expressed in terms of the single scale L_{MO} and the altitude z. The Monin–Obukhov length is a function of wind speed, heat flux, and surface roughness.

In practice, L_{MO} can be estimated from the Turner classes, which were developed for use in air pollution studies. The Turner class is determined from the wind velocity and an estimate of solar radiation (see Table 3). Since L_{MO} is also a function of surface roughness, this parameter must also be included in estimating L_{MO} from the Turner class. In Table 4 we present the range of Monin–Obukhov lengths for a given Turner class for a site which has a roughness length of about 0.05 m. For all conditions investigated, L_{MO} is negative.

The planetary boundary layer (PBL) represents the farthest extent of surface heating and cooling. This layer varies in thickness from a few hundred meters at night to a few kilometers on a sunny day. The region in the planetary boundary layer above $-L_{MO}$ is known as the "mixed" layer. The scaling parameter in this layer is the mixing depth h, the vertical extent of the planetary boundary layer. This type of scaling is called mixed-layer scaling. In the mixed layer, certain turbulence statistics can be written as universal functions of z/h. Some statistics, in particular those associated with vertical velocity fluctuations, still obey Monin–Obukhov scaling in the mixed layer.

With these two scaling factors in mind, we can now examine the theoretical variation of

Table 3. Estimation of Turner classes (daytime only).

Surface wind speed (at 10 m) (m/s)	Incoming solar radiation		
	Strong	Moderate	Light
<2	1	1	2
2–3	1–2	2	3
3–5	2	2–3	3
5–6	3	3–4	4
>6	3	4	4

Class 4 should be assumed for overcast conditions.

Table 4. Turner class versus L_{MO} for a flat site in Illinois.

Turner class	$-L_{MO}$ (m)
1	8–12
2	12–20
3	20–60
4	60

convective and mechanical turbulence with altitude for the case of an unstable to near neutral daytime atmosphere.

The dominant term in the vertical velocity fluctuation has scales which are a monotonically increasing function of height. As $-L_{MO}$ increases, this function becomes approximately linear. The magnitude of the scale is independent of stability for $z > -0.7L_{MO}$.

Horizontal velocity fluctuations obey mixed-layer scaling. The scale of the horizontal fluctuations is determined by the height of the mixed layer and is nearly constant, as a function of altitude, in the planetary boundary layer. These scales will be larger than the vertical velocity scales for unstable measurement conditions.

The temperature fluctuations are closely related to the vertical velocity fluctuations and are expected to obey Monin–Obukhov scaling. Since the temperature fluctuations are caused by vertical velocity fluctuations acting on the temperature gradient, the temperature fluctuation scale is expected to increase monotonically with altitude under unstable conditions.

The variance of the wind speed will also be determined by fluctuations in both the horizontal and vertical. Examining the theoretical variance in the surface layer, the variance in the vertical wind component is expected to increase slowly with height for unstable conditions as a function of z/L_{MO}:

$$\sigma_w \propto [1-3(z/L_{MO})]^{1/3}. \tag{47}$$

The larger L_{MO}, the slower the variation. The variances of the horizontal components should be constant in the unstable surface layer. Above the surface layer, all wind variances should decrease as the altitude approaches the top of the planetary boundary layer.

The temperature variance in unstable air is a more rapidly decreasing function of height in the surface layer than the variance of the vertical wind velocity. Johnson *et al.* (1982) use the form

$$\sigma_T \propto [1-16(z/L_{MO})]^{-1/2}. \tag{48}$$

[Note that Eq. (48) corrects an error in Eq. (2) of Johnson *et al.* (1982).] This implies, for example, that at $-L_{MO}$ the variance should be one-fourth of that close to the ground. Above L_{MO}, the variance in temperature should decrease as $(z/L_{MO})^{-1/3}$.

A summary of the average results of measurements made at Bondville Field Station in Illinois serves as indicators of typical results. This summary represents typical measured values of turbulence strength and scale at a location that has good fetch. A rougher terrain with poor fetch would have values of L and $\langle\mu^2\rangle$ that would be slightly larger near the ground. The variation of these variables with height for poor fetch would probably be similar to these results at heights away from the ground, but there are no data to confirm this.

(a) Variation of scale.

For overcast, light wind conditions (implying a shallow PBL),

$L \approx 1.0$ m @ 1 m to $L \approx 7.0$ m @ 33 m.

For sunny, windy conditions (implying a deep PBL),

$L \approx 1.0$ m @ 1 m to $L \approx 12.0$ m @ 33 m.

The variation of L with height is approximately linear.

(b) Variation of strength.

Sunny, light wind (<2 m/sec) conditions: $\langle\mu^2\rangle \approx 5\times10^{-6}$.

Sunny, moderate wind (2–4 m/sec) conditions: $\langle\mu^2\rangle \approx (9\text{–}10)\times10^{-6}$.

Sunny, strong wind (>4 m/sec): $\langle\mu^2\rangle \approx (15\text{–}25)\times10^{-6}$.

Overcast, light wind conditions: $\langle\mu^2\rangle \approx 3\times10^{-6}$.

Overcast, moderate wind conditions: $\langle\mu^2\rangle \approx (8\text{–}9)\times10^{-6}$.

Overcast, strong wind conditions: $\langle\mu^2\rangle \approx (15\text{–}25)\times10^{-6}$.

There is very little variation of $\langle\mu^2\rangle$ with height through 10 m. Above 10 m, $\langle\mu^2\rangle$ may be slightly increasing, as determined by the wind speed variance.

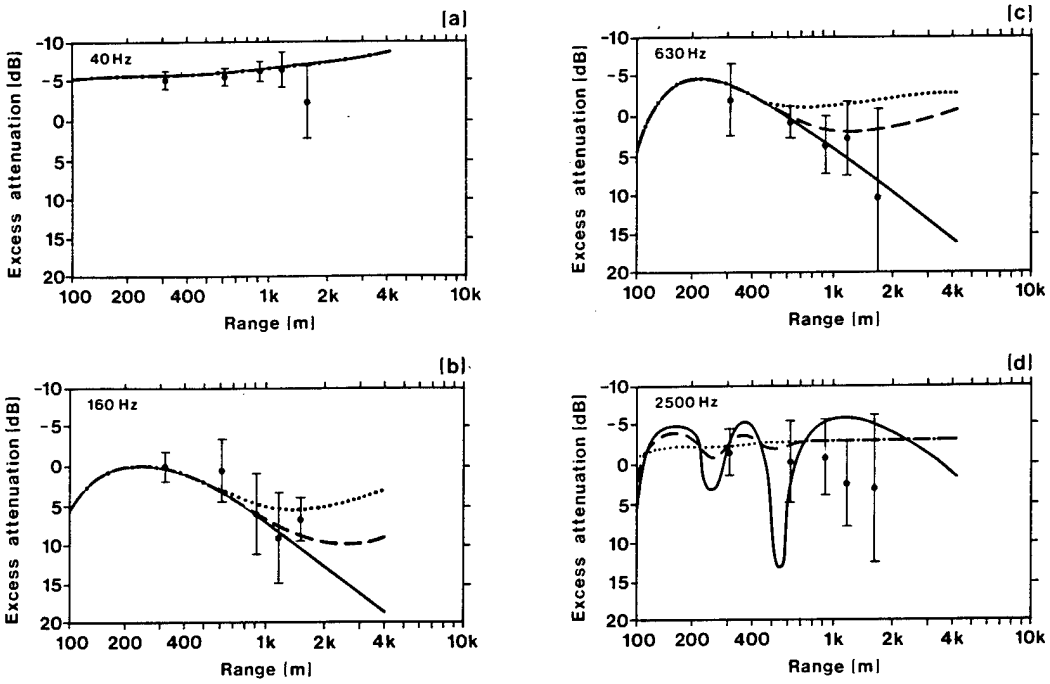


FIG. 23. Excess attenuation vs horizontal distance for 40, 160, 630, and 2500 Hz. The data are the points with error bars. The data at 1550 m are limited since wind noise reduced the sample size. The solid lines are the excess attentuation prediction for a nonturbulent atmosphere. The dashed lines are predictions for $L_0=1.0$ m, $\langle\mu^2\rangle=3\times10^{-6}$, and the dotted lines are for $L_0=1.0$ m, $\langle\mu^2\rangle=15\times10^{-6}$.

Figure 23 displays the variation in excess attenuation with distance for four frequencies measured at Bondville Field Station, Illinois. The source is a set of eight speakers on top of a tower at a height of 31 m. The receivers were 1.2 m high. The data displayed are the average of eleven 15-min measurements of excess attenuation selected as representing a nearly homogeneous atmosphere. These sets had a logarithmic sound speed gradient between -0.11 m/sec and 0.11 m/sec and the standard deviation of the gradient was less than 0.05 m/sec. The predicted quiet-atmosphere excess attenuation with distance for the average ground impedance condition is also displayed in Fig. 23 as the solid lines. These calculations display the expected asymptotic decay of excess attenuation at $1/r^2$ for large distances. The transition to $1/r^2$ depends on frequency, source height, and impedance.

The data for 160 and 630 Hz behave more coherently than was expected. The 40-Hz data and calculation display little variation from coherent behavior and will not be discussed further. The 2500-Hz data, on the other hand, display near incoherent behavior at all ranges.

Calculations based on the extremes of the variance of the index of refraction are displayed as the dashed lines in Fig. 23. These results are calculated for a scale length of 1 m, typical of the scale lengths measured at 1-m height. It is observed that the turbulence effects are less than would be observed for ground to ground propagation for the same scale length. Further, the calculation indicates a larger difference in the coherent and turbulent results than was measured.

For ground to ground propagation, in a quiet homogeneous atmosphere, there is cancellation between the direct wave and plane-wave reflected wave and only the lower-amplitude ground wave is measured. Turbulence reduces the coherence between the equal-amplitude waves, and therefore increases the amplitude of the resultant wave.

2.6 Nonlinear Effects

Large-amplitude waves and impulses will give rise to nonlinear propagation effects. If we consider an impulse as shown in Fig. 4, the point on the pressure wave form corresponding to maximum pressure will travel more rapidly than points of lower pressure according to the relation

$$\Delta v = c[(\gamma+1)/(\gamma-1)][(p/P_0)^{(\gamma-1)/2\gamma}-1], \tag{49}$$

where Δv gives the speed at which a point on the pressure wave form with pressure p travels in excess of the $p=P_0$ point; Δv may be positive or negative. The amount the positive phase of the wave form lengthens can be found from

$$\Delta L = \int_{t_0}^{t} \Delta v \, dt'. \tag{50}$$

The term Δv is a function of time, since from Eq. (49), Δv depends on p/P_0, and $\Delta p/P_0 = (p - P_0)/P_0$ decreases as r raised to a negative power depending upon the spreading law involved, which in turn depends on t. The relationship between r and t is not simple, since

$$r = vt + r_0 \tag{51}$$

and $v = c + \Delta v$, which is a function of (p/P_0) from Eq. (49).

As the wave propagates, the wave steepening predicted by Eq. (49) will tend to transfer energy from low frequency to high. Since atmospheric absorption increases approximately as the frequency squared, these high frequencies will be rapidly attenuated. The overall rate of decay of the wave form, then, will depend upon how rapidly low-frequency energy is transferred to high frequencies which in turn depends upon the amplitude of the acoustic pressure.

One indicator of this wave steepening versus high-frequency attenuation competition is the shock rise time discussed earlier. Referring to Fig. 24, we can see that the rise time decreases as the over-pressure increases. Since the rise time is a measure of the highest frequencies in the impulse, the small rise time (high frequency) associated with high amplitude shows that nonlinear steepening is replacing energy more rapidly for high amplitude waves.

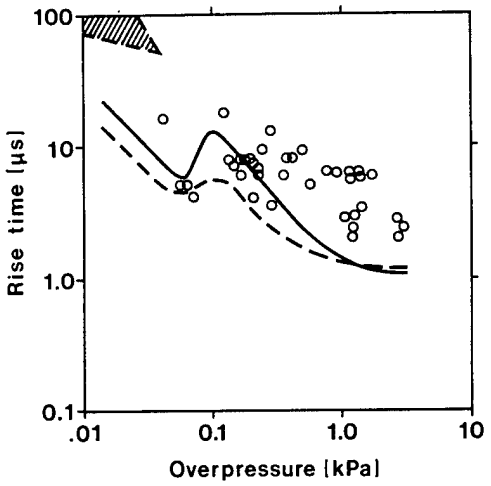


FIG. 24. Experimental and computed rise times. The hatched area and circles represent experimental data from Ref. 1. The solid line was computed for $T=305$ K, relative humidity 90%. The dashed line was computed for $T=293$ K, relative humidity 50%.

Harmonic waves are affected by nonlinearity in slightly different ways. Rather than pulse lengthening, harmonic waves exhibit subharmonic generation. As the subharmonics grow (they are attenuated less than the harmonics), they can produce harmonics and subharmonics until, provided the original amplitude was great enough, the harmonics and subharmonics will completely fill the frequency spectrum and appear as noise.

2.7 Infrasound Propagation

The primary mode of propagation for infrasound at long ranges is by multiple hops; each hop is typically 250 km. This process results from temperature inversions in the upper atmosphere (altitudes of 50–120 km) and reflections at the earth's surface. Referring to the section on absorption, we find that at low frequencies absorption is very small; hence long propagation paths are allowed. The surface impedance is very large at low frequencies so very little energy is lost to the surface on each hop. After a few hops, surface-generated infrasound is constrained to a channel and spreads cylindrically rather than spherically. Empirically it is found that (ANSI, 1983)

$$p \sim R^{-n} \text{ with } n \cong 1.1. \tag{52}$$

Infrasound propagation depends strongly on winds in the atmosphere which change seasonally. As a result, propagation is more effective toward the east in the winter when atmospheric winds tend to blow in that direction. The velocity of propagation for infrasound also varies seasonally with wind and temperature of the primary transmission channel. As shown in Fig. 14, the local sound velocity can rise dramatically at high altitudes but the fact that the waves bounce back and forth between the surface and the upper transmission channel increases the effective propagation distance. Therefore, infrasound velocities are typically a little less than the speed of sound near the earth's surface.

2.8 Numerical Techniques

The conditions for which ray tracing is applicable are

1. $\lambda n' \ll n^2$ where n' is the rate of change in the index of refraction n, and λ is the acoustic wavelength;
2. $\lambda(\ln A)' < 1$ where $(\ln A)'$ is the rate of change in the logarithm of the acoustic amplitude; and
3. $(\lambda/2\pi R)^{1/3} < \theta$ where R is the propagation range and θ is the grazing angle.

Note that each condition becomes increasingly difficult to meet as λ increases (low frequency). Two common sources of low-frequency sound are explosions and rotary wing aircraft.

Air pressure waves from very energetic explosions [the Mount St. Helen's eruption (Reed, 1987), for example] travel through the upper atmosphere and are refracted slowly so that n' is always very small. In these events, ray tracing routines accurately predict quiet zones relatively near the sources and very dramatic damage much further away due to refractive focusing. Low-flying helicopters do not produce sufficient energy to result in detectable sound levels after reflection by the upper atmosphere. The received sound depends, primarily, on index of refraction gradients near the ground surface. In this case, n' can be quite large. For such cases ray tracing can lead to poor predictions of long-range sound propagation and a numerical solution to the wave equation becomes necessary.

Full wave solutions take a variety of forms. The most accurate is a finite element solution. For propagation distances greater than a few meters, such algorithms take many hours of supercomputer time. Most computations involve modified ray tracing routines, the parabolic approximation, or the fast field program. At long ranges, corrected ray tracing routines are typically more time consuming than the other two approaches so ray tracing is generally limited to cases where extensive modifications are not required (sources nearly overhead or an almost neutral atmosphere). Of the remaining two approaches, the PE (parabolic equation) has the capability of including the largest number of effects but is least developed. At very long ranges, the FFP (fast field program) has an advantage in run time but can still consume several minutes on a minicomputer.

For purposes of this contribution, only the FFP will be described in detail (Raspet *et al.*, 1985). The problem considered is sketched in Fig. 25. In the space $0<z<z_{n-1}$, there exist $n-1$ liquid layers. A time-harmonic point source of constant pressure is located at the lower interface of layer S with cylindrical coordinates $(r=0, z=z_S)$. The problem at hand is to determine the pressure at the observation point located at the lower interface of layer D with coordinate $(r=r, z=z_D)$.

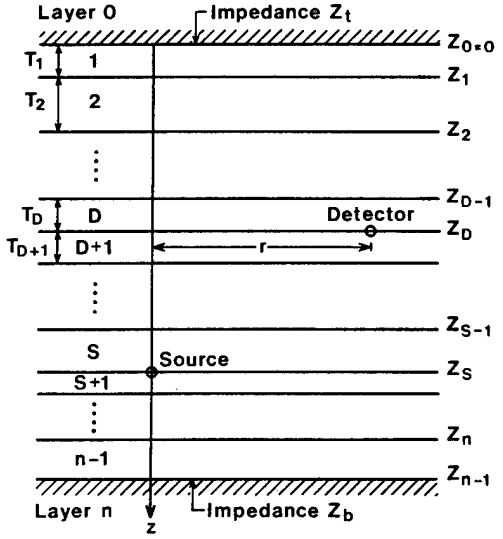


FIG. 25. Liquid layered medium with source and detector bounded by impedance surfaces.

In a typical layer m, the acoustic field is described by pressure $p_m(r, z)$ and a vertical particle velocity $\dot{w}_m(r, z)$. Both of them are expressed in terms of the familiar Fourier–Bessel transform

$$\begin{bmatrix} p_m(r, z) \\ \dot{w}_m(r, z) \end{bmatrix} = \int_0^\infty \begin{bmatrix} p_m(k, z) \\ \dot{w}_m(k, z) \end{bmatrix} J_0(kr)dk. \quad (53)$$

Here, the transform variable k can be interpreted as the propagation constant. Two surface-impedance boundary conditions are to be enforced at $z=0$ and at $z = z_{n-1}$.

At a surface $z=a$, the surface impedance in the transform domain is defined by

$$Z(k) = \text{sgn}\,\frac{p(k,z=a)}{\dot{w}(k,z=a)}, \quad (54)$$

where

$$\text{sgn} = \begin{cases} +1 & \text{if } \dot{w} \text{ moves into impedance surface,} \\ -1 & \text{if } \dot{w} \text{ moves away from the surface.} \end{cases}$$

For the configuration in Fig. 25, sgn$=-1$ if $a=0$ for top impedance Z_t, and sgn$=+1$ if $a=z_{n-1}$ for bottom impedance Z_b. It is assumed that the space above $z=0$ is a liquid half-space, i.e.,

$$Z_t = \omega\rho_0/(k_0-k^2)^{1/2}, \quad (55)$$

where, for $m=0,1,2,\ldots,n$,

ρ_m = density of layer m in kg/m^3,

c_m = acoustic wave speed in layer m in m/s,

$k_m = \omega/c_m$ = wave number in layer m in m^{-1}.

Use the exp$(+i\omega t)$ time convention; then (54) assumes the following values:

$$Z_t(k) = \begin{cases} +|Z_t(k)| \text{ (resistive)} & \text{if } k_0 > k, \\ +i|Z_t(k)| \text{ (inductive)} & \text{if } k_0 < k. \end{cases}$$

The surface impedance Z_b at the bottom surface has already been described.

An integral solution for the pressure $p(r)=p_D(r,z=z_D)$ at the detector's location can be written as

$$\frac{p(r)}{p_0} = (-i)\int_0^\infty F(k)J_0(kr)dk, \quad (56)$$

where

$$F(k) = \frac{2k}{\omega\rho_s}[(D_D)_{21} - Z_t(D_D)_{22}]\left(\frac{(D_s)_{22}(D_{21}+Z_bD_{11}) - (D_s)_{21}(D_{22}+Z_bD_{12})}{(D_{21}+Z_bD_{11}) - Z_t(D_{22}+Z_bD_{12})}\right); \quad (57)$$

p_0 = source pressure at a unit distance away in a homogeneous medium.

In Eq. (57), we have used propagation matrices

$$D_s = d_s d_{s-1}\cdots d_2 d_1,$$

$$D = d_{n-1}d_{n-2}\cdots d_2 d_1,$$

where d_m is a 2×2 matrix with elements given by

$$(d_m)_{11} = (d_m)_{22} = \cos P[\xi_m(z_m - z_{m-1})],$$

$$(d_m)_{12} = (i\xi_m/\omega\rho_m)\sin P[\xi_m(z_m - z_{m-1})],$$

$$(d_m)_{21} = (i\omega\rho_m/\xi_m)\sin P[\xi_m(z_m - z_{m-1})],$$

$$\xi_m = \sqrt{k_m^2 - k^2},$$

D_{12}, for example, is the (1,2) element of matrix D.

Now consider the evaluation of the integral involved in Eq. (56). Of primary interest is the far field solution, i.e., the solution when the observation point is far away from the source. To this end, $J_0(kr)$ can be rewritten as

$$J_0(kr) = \tfrac{1}{2}[H_0^{(1)}(kr) + H_0^{(2)}(kr)]. \quad (58)$$

The component involving the incoming wave factor $H_0^{(1)}(kr)$ does not contribute significantly, and can be neglected. Thus, Eq. (56) is approximated by

$$\frac{p(r)}{p_0} \approx \frac{1}{2i}\int_0^\infty F(k)H_0^{(2)}(kr)dk. \quad (59)$$

Two further approximations are used in (59), namely,

1. for the far field, $H_0^{(2)}$ is replaced by its asymptotic expansion:

$$H_0^{(2)}(kr) \sim (2/kr)^{1/2} e^{-i(kr-\pi/4)}, \quad kr \to \infty;$$

2. the range of integration is truncated at $k = k_{max}$. The final expression suitable for numerical evaluation is

$$\frac{p(r)}{p_0} \approx \frac{1-i}{2\sqrt{\pi r}} \int_0^{k_{max}} \frac{F(k)}{\sqrt{k}} e^{-ikr} dk, \quad r \to \infty, \tag{60}$$

which is the normalized pressure at a distance r from the source.

Next consider the numerical evaluation of (60) by the fast Fourier transform. To this end, four parameters are required:

k_{max} = maximum horizontal wave number in integration range,

N = number of steps of integration,

k_0 = starting point of integration,

r_0 = minimum range.

The integral in Eq. (60) is approximated by a summation:

$$\frac{p(r_{n+1})}{p_0} = (\Delta k) \frac{1-i}{2(\pi r_{n+1})^{1/2}} \times \sum_{m=0}^{N-1} \left(\frac{F(k_{m+1})}{k_{m+1}^{1/2}} e^{-imr_0 \Delta k} \right) \times e^{-i2\pi mn/N}, \tag{61}$$

for $n = 0,1,2,\ldots,N-1$, where

$$\Delta k = (k_{max} - k_0)/(N-1), \tag{62}$$

$$\Delta r = 2\pi/N(\Delta k), \tag{63}$$

$$k_m = k_0 + m(\Delta k), \quad m = 0,1,2,\ldots,N-1,$$

$$r_n = r_0 + n(\Delta r), \quad n = 1,2,3,\ldots,N.$$

Consider the case that c decreases from 335 m/s at the ground to 327 m/s at 8-m height. This is a sample model of shadow zone propagating with a noninfinite negative gradient up to 8 m. The source and the detector heights are 1 and 4 m, respectively. The detector is located in the shadow zone when $r > 71$ m. Regions near the ground are approximated by 16 layers of 0.5-m thickness. For source frequency $f = 40$ Hz, the pressure versus range curve calculated by FFP is shown in Fig. 26 (solid line). The quantity

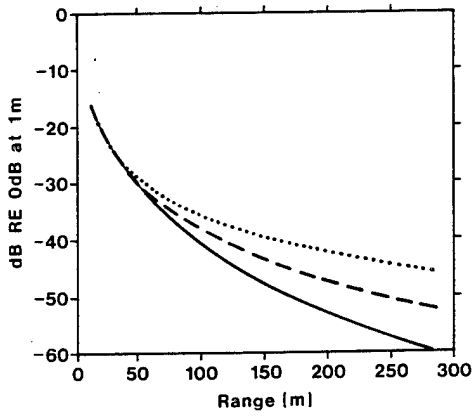


FIG. 26. Normalized pressure vs range for the inhomogeneous half-space sketched in Fig. 25 at a frequency of 40 Hz. In the diagram the dotted line is for the homogenous case, $s = 0$, the dashed line is for $s = 1.567 \times 10^{-3}$ m^{-1}, and the solid line is for $s = -2.985 \times 10^{-3}$ m^{-1}.

$$s = \Delta c / c \Delta h \tag{64}$$

is a measure of the speed of sound gradient. The curve for a homogeneous atmosphere ($c = 335$ m/s for all heights, $s = 0$) is shown by the dotted line. Much larger attenuation is predicted for the inhomogeneous case due to the upward refraction of the sound. In the region $r < 71$ m this reduction in level is principally caused by the reduction in grazing angle; beyond 71 m the detector is in the shadow zone and receives only diffracted sound energy. For ranges greater than 71 m, a simple ray trace treatment would predict infinite attenuation. In Fig. 26, a third curve is shown corresponding to an intermediate gradient $s = -1.567 \times 10^{-3}$ m^{-1}. As expected, this curve falls between the other two.

3. APPLICATIONS

Applications of atmospheric acoustics are quite varied. For purposes of this article only two will be considered. The first is the use of acoustics to determine the state of the atmosphere. Acoustic sensing of the atmosphere takes advantage of the fact that acoustic propagation is very sensitive to atmospheric conditions. The second, and most widespread, application is for prediction of noise levels. In addition to the nature of the source and propagation path discussed to this point, the re-

ceiver must be considered. When the receiver is a human listener, human perception of sound becomes an important factor. Based upon a knowledge of the source and transmission path, the sound level at the receiver can be accurately predicted. Human response to that sound, however, is much less predictable.

3.1 Acoustic Sensing of the Atmosphere (Brown, 1987)

The invention of the echosonde (also called the sodar), which McAllister patented in 1972, added a powerful tool for atmospheric research and for applications to forecasting especially in the atmospheric boundary layer. McAllister combined improved antenna design with the use of the facsimile recorder patented by Marti in 1919, which shows graphically the turbulent structure in the boundary layer. The use of the echosonde has become so widespread that results derived with it have become a ubiquitous part of the environmental impact studies required in many countries around the world. In its more complex form—the Doppler echosonde—the instrument can provide measurements of mean winds, sea breezes, and the flows in and out of valleys. An outgrowth of the instrument, called the echometer, shows promise of providing vertical profiles of mean temperatures as well. The monostatic echosonde relies on backscatter (scattering angle $\theta_s=\pi$) from a field of turbulence-driven temperature fluctuations. The bistatic echosonde ($\theta_s\neq\pi$) includes separated transmitter and receiver, with a signal that results from both temperature and velocity fluctuations. A usable return signal from a transmitted pulse in the monostatic geometry requires both a gradient in the temperature and a turbulent velocity field to shift air parcels to different positions, where the parcels will assume different actual temperatures T from their neighbors and produce a random field of refractive index fluctuations.

The equations of fluid mechanics, together with two assumptions valid in acoustic remote sensing, that the turbulence remains a nonzero-vorticity and incompressible field that produces no sound, while the acoustic wave remains a vorticity-free longitudinal compression field that produces no changes in the turbulence, lead to the governing acoustic remote-sensing equation. If the turbulence "eddy" sizes (here, an eddy represents a Fourier space component, not a physical swirl or blob) lie within the inertial range (tenths of meters to tens of meters), and Taylor's "frozen turbulence" hypothesis* holds, the "single scatter," or Born, approximation is valid. Then, the scattering cross section σ_s (in reciprocal meters) as a function of scattering angle θ_s becomes

$$\sigma_s(\theta_s)=1.52k_0^{1/3}\cos^2(\theta_s)\times[0.13C_n^2+\cos^2(\theta_s/2)C_v^2/(4c^2)]\times[2\sin(\theta_s/2)]^{-11/3}, \quad (65)$$

with k_0 the initial wave number, C_v^2 the structure parameter for velocity fluctuations ($m^{4/3}\,s^{-2}$), and C_n^2 the structure parameter for refractive index fluctuations ($m^{-2/3}$) given by

$$C_n^2=C_T^2/4T_0^2+2(0.307)C_{eT}/4P_0T_0+(0.307)^2C_e^2/4P_0^2, \quad (66)$$

where subscript e refers to water vapor, and T to temperature, fluctuations (C_T^2 and C_{eT} are in $K^2\,m^{-2/3}$ and $K\,N\,m^{-8/3}$, respectively). One should note that a single fluctuation would have negligible effect on the scattered wave. However, the nonzero correlation and cross-correlation distances for turbulence velocity, vapor pressure, and temperature fluctuations imply a quite different kind of interaction than that from a field of randomly spaced point scatterers. Therefore, wavelets scattered from nearby regions can interfere constructively and destructively to produce a partially coherent scattered wave resulting from the integration of many wavelets over a nonzero region of space. The same physics underlies the importance of the spatial Fourier transform decomposition of the random fluctuation field.

A typical monostatic echosonde employs a single antenna for both transmission and reception, with an electronic switch to alternate between the two. The antenna includes a high-power pressure transducer (similar to those used in airport public address systems)

*Taylor's "frozen turbulence hypothesis" maintains that temporal changes in local meteorological parameters caused by turbulence are related to spatial changes through the mean wind.

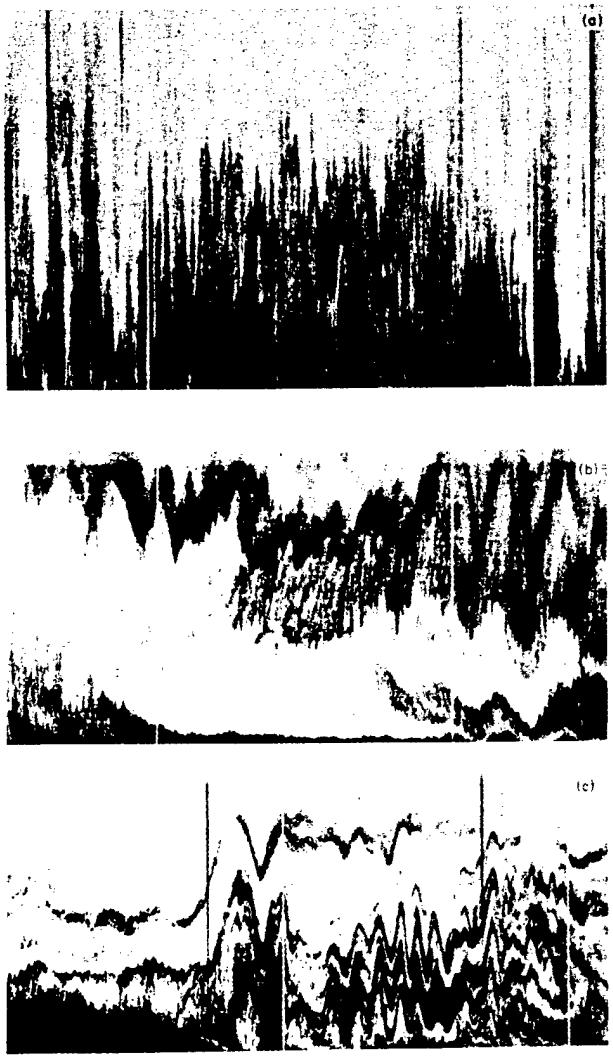


FIG. 27. Echosonde records of the boundary layer. Vertical scale: 0–500 m; horizontal scale: time, increasing to the right; vertical white streaks are hour markers. (Courtesy of C. W. King.) (a) Daytime thermal plumes; (b) decreasing ground layer (below) and descending inversion (above) in early evening, followed by triggering of internal waves; (c) fully developed internal wave activity after midnight.

together with an exponential feed horn (for impedance matching) mounted at the focus of a parabolic dish, with the entire assembly enclosed in a shield, or "cuff," made from plywood lined with an acoustic foam that includes a lead-sheet septum. The cuff provides substantial side-lobe suppression and, if the sides flare, an increase in the effective aperture of the antenna. Typical antennas achieve angular beamwidths of 8°–10°. The reason for using the same antenna for reception rather than a more sensitive microphone lies in this narrow beamwidth and the resulting much greater signal-to-noise ratio.

Figure 27 shows a sequence of boundary layer structures over a typical diurnal period. (Time in the figure moves from left to right.) The top of the figure corresponds to an altitude of 500 m; vertical thin white stripes represent hour markers. Figure 27(a) shows a typical unstable, daytime record of thermal plumes carried through a vertical echosonde beam by the mean wind. Such plumes result from the conductive random heating of air near the ground, itself heated by solar radiation. Randomness may result from variations in surface albedo and from the presence of clouds. Figure 27(b) shows the boundary layer in the late afternoon and early evening. The decrease and then disappearance of solar

radiation and the consequent cooling of the ground brings about the vanishing of the thermal plumes, a concurrent creation of a ground-based shear layer that on a finer time scale would show a "herringbone" structure, and the slow descent of an elevated inversion previously off the top of the record. Near its end, the figure shows the beginning of internal gravity wave activity. In Fig. 27(c) a series of large-scale internal gravity waves dominate the record. Especially noteworthy, the particle displacement double amplitude (peak to peak) reaches values of the order of 300 m. In the descending part of the cycle such large displacements require a horizontal outflow near the ground that would appear similar to wind gusts, except for the regularity of their return with the period of the waves.

Although the bistatic configuration has many other uses, its greatest value lies in vertical profiles of Doppler measurements of mean winds. Figure 28 outlines the geometry of one leg of such a system. (With two horizontal legs at right angles to each other and the addition of a vertical Doppler measurement, one can map the entire three-dimensional wind field at all heights within the echosonde range.) A number of different arrangements have shown various advantages and drawbacks. The following describes a particular design that seems one of the more flexible and dependable and has the advantage that the vertical narrow receiver "pencil beam" rejects much of the noise coming in on low-elevation paths from road traffic, distant aircraft, and residential areas. The typical transmitter antenna uses a "fan beam" geometry, one that sprays an initial pulse over a range of angles. Time-gating of the received signal then gives a Doppler value at each desired height. Other configurations use a set of three monostatic echosondes pointing in three different directions, allowing greater simplicity in the hardware but producing greater errors when the atmosphere lacks horizontal homogeneity because of the separations between the three different scattering volumes. The wind velocity will be denoted by a vector $\mathbf{w}$ or a Mach vector defined as $\mathbf{m}=\mathbf{w}/c$. For the monostatic Doppler echosonde, when $\mathbf{m}$ is directed along the beam, and with no refraction, the frequency shift Δf satisfies

$$\Delta f/f=\pm 2m, \tag{67}$$

where the plus sign holds when the wind blows toward, and the minus sign away, from the antenna, and $m=|\mathbf{m}|$. For the bistatic system, shown in Fig. 28 (together with some simplifying assumptions, such as horizontal stratification and negligible vertical wind or cross wind) the horizontal component of the Mach vector m_x, with $m_x = \pm|m_x|$, and with an obvious choice of sign, satisfies

$$\Delta f/f=-m_x \sin\theta+(m_x \sin\theta)^2. \tag{68}$$

Frequently, an average over, say, 20 min will cancel the effect of any vertical wind; that is, one can consider it zero. However, a complete horizontal wind measuring system may often face conditions where cross winds on one leg or the other become important, or some other simplifying assumption no longer holds.

Researchers prefer various methods for determining the frequency shift Δf from the raw data, especially to cope with the ubiquitous presence of noise. In the frequency domain, many use the FFT (fast Fourier transform), at the same time trying to measure or estimate the spectral contribution of the noise. However, the FFT requires a fairly powerful computer. In the time domain, the CXCV (complex covariance) method provides good data with a much less expensive microprocessor but requires careful choice of sampling frequency and filter bandpass; a choice of sampling frequency such that the Nyquist frequency remains near the filter half-power point minimizes aliasing.

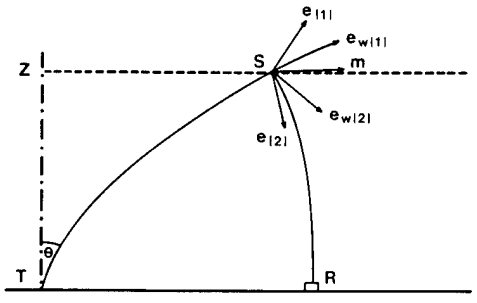


FIG. 28. Geometry for Doppler wind measurements. Transmitter *T* and receiver *R*. Scattering volume center *S* at height *z*. Other symbols defined in the text.

With the increasing amount of inexpensive computer memory becoming available, echosonde designers now have the choice of turning to completely coherent signal processing, for example; averaging the signal in both amplitude and phase, and, thus, reducing random noise to a negligible problem. The noise of rain hitting the receiving antenna makes echosonde observations impossible during a storm. In contrast, the temperature-measuring echometer, by virtue of the (here necessary) coherent processing, operates successfully through a rain squall, producing data in close agreement with those from a nearby *in situ* thermometer on a tower. No obvious reasons exist that would rule out similar coherent processing of echosonde signals to obtain a large increase in signal-to-noise ratios.

The boundary layer can have either an "open" structure for the unstable case, acting as a pathway for fluxes of momentum, heat, and water vapor to reach higher levels and perturb mesoscale processes, or a "closed" structure for the stable case, acting as an impenetrable container disconnected from large-scale events. In the first case, information on the structure may have an impact on future weather-prediction methods. In the second, boundary layer data should assist in predicting the strength and duration of air pollution events. Many of the parameters needed, however, derive from higher-order measurements that have a slightly shaky theoretical basis, and experimental values of unknown precision. Thus, they represent work still in progress, on the leading edge of the research in acoustic remote sensing.

Among other parameters, dependable measurements of mean temperature profiles $T(z)$ still elude researchers. One method, called RASS (radio acoustic sounding system), uses an acoustic pulse containing of the order of 50 phase fronts to reflect a radar signal coherently back to a receiver. With an acoustic wavelength equal to one-half the radar wavelength, strong returns arrive from altitudes throughout the boundary layer. The Doppler shift in the radar return derives—with a correction for mean vertical wind w_z—from the speed of the acoustic waves.

When an elevated inversion lies above an unstable atmosphere near the ground, the height z_i of the inversion becomes a useful parameter. The upper part of the boundary layer, the well-mixed layer, as its name implies, has very little fluctuation in temperature except in the thermal plumes and thus shows no echosonde returns elsewhere, even if (velocity) turbulence exists. Within the inversion the atmosphere remains stable, that is, without turbulence, even though the temperature T changes, and thus again the echosonde shows no returns. However, the rising air in the plumes has a tendency to overshoot and, before descending, entrain air with a different temperature from the bottom of the inversion in its turbulent motion. As a result echosonde facsimile records frequently show a dark elevated band corresponding to this entrainment layer above the region of thermal plumes. Comparisons with radar or FM-CW radar seem to support good agreement between their measurements of the base of an elevated inversion z_i and the echosonde measurement of the height of the bottom of the dark band produced by an entrainment layer.

3.2 Perception of Sound

Measurements and measures associated with atmospheric acoustics are typically related to the hearing response of the human ear. Such a treatment is especially applicable when dealing with acoustic noise and noise control. The loudness level is defined as the sound pressure level of a 1000-Hz tone that sounds as loud as the sound in question (see ACOUSTICS, PSYCHOLOGICAL). The unit of loudness is the phon. The famous Fletcher–Munson equal-loudness contours shown in Fig. 29 give an indication of the loudness level at different sound pressure levels. Note that a sound pressure level of 70 dB at 25 Hz has the same loudness as a 10-dB tone at 1000 Hz.

When measuring noise in the atmosphere, it is common to express the level in dBA, dBB, or dBC. A sound level meter with these ratings has a frequency response given in Fig. 30. The A-weighted curve approximates the hearing response shown in the previous figure for relatively low sound pressure levels ($\sim$40 dB). The B weighted curve is applicable at higher levels (80–90 dB). C-weighted sound levels show less frequency compensation and, therefore, are a better measure of sound pressure level.

Measures of noise are designed to give an indication of human response. Typically, the

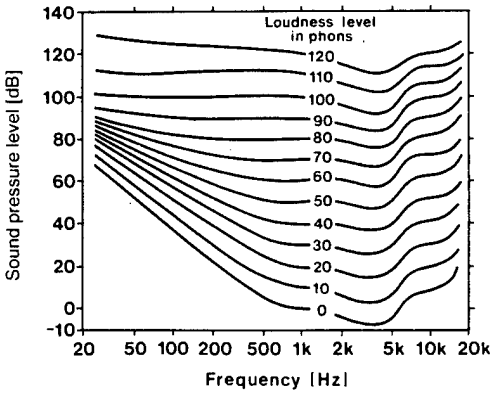


FIG. 29. Equal-loudness contours for pure tones by Fletcher and Munson. (American Standard for noise measurement, Z24.2-1942.)

goal is to reduce annoyance. Since different types of sound present different annoyances at the same C-weighted instantaneous level, a variety of noise measures have been advanced. Some of the more commonly used are perceived noise level (PNL), sound exposure level (SEL), and day-night sound level (L_{dn}).

In contrast to the A level, the perceived noise level weighting function is based upon subjective assessment of the noisiness of the various frequency components of the signal, rather than upon *loudness* considerations. There is more emphasis on the upper portion of the noise spectrum (2000–4000 Hz) in this weighting function than in the A-weighting function. The perceived noise level, in units of PNdB, has been used for many years in the U.S. as a measure of aircraft noise.

Subjective tests indicate that human response to noise is not only a function of the maximum level, but of the duration of the signal and its temporal variation. Time-related changes may range from a sound level constant over time, as produced by a continuously operating machine, to the typical haystack-shaped time history produced by an aircraft flyover, to the constantly varying noise level perceived near highways.

Historically, several methods have been used to introduce time characteristics. With recent advances in electronics and instrumentation technology, there are now instruments which can integrate, or sum, noise signals as a function of time. (Integrated noise levels are a measure of the physical energy in the noise signal.)

Significant evidence indicates that two signals with equal sound energy will produce the same subjective response. For example, a noise with a constant level of 85 dB occurring for 10 min would be judged equally as annoying as an 82-dB noise signal lasting for 20 min, i.e., one-half the energy lasting for twice the time period. This is known as the "equal energy" principle. In practice, the integration process is often replaced by a summation of levels occurring at one-half second intervals over the upper 10 dB of the noise signal.

Integration of the A-weighted noise level over the period of a single event (such as an aircraft flyover) gives the sound exposure level, in dB. The day-night average sound level (referred to as L_{dn}) is obtained by energy-averaging noise levels over a 24-hour period, with a 10-dB penalty to nighttime noise levels. The day-night average sound level can be applied to all sources of noise. For discrete event noise sources, such as aircraft operations, L_{dn} may be computed by decibel summation of noise levels occurring during daytime and nighttime periods (with the nighttime penalty). Areas subjected to L_{dn} in excess of 75 are useful only for selected commercial applications. Areas of a 70–75 L_{dn} may be used for a wider variety of commercial applications while areas impacted by 65–70 L_{dn} may be used for most applications including residential with proper sound treatment. Below 65 L_{dn} some treatment may be required for uses requiring a quiet background.

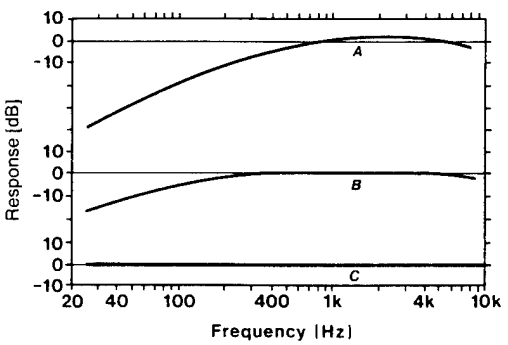


FIG. 30. Frequency-response characteristics taken from the American Standard for sound-level meters, Z24.3-1944.

Combining perception with propagation efforts described earlier provides practical guidelines for predicting community objections to different types of noise sources. In most cases involving community noise (Makarewicz, 1987), the source and receiver heights are much smaller than the source–receiver separation so $|z|\cos\theta_i \ll 1$. For this geometry, the A-weighted mean-square sound pressure at distance d can be given as

$$P_A^2 = \tilde{P}_A^2[S(\sigma_e,\mu,\beta d)/d^2], \tag{69}$$

where $\tilde{P}_A^2$ is the A-weighted mean-square sound pressure for the direct wave, σ_e is the effective flow resistivity of the surface, β is the absorption coefficient α divided by the frequency f, and μ is a parameter which describes the source spectrum $[P(f) = P(0)\exp(-\mu f)]$. Far away from a source ($d \gg 500$ m), $P_A^2 \sim d^{-7}/1$ m which suggests that the A-weighted sound level will decrease at a rate of 21 dB per doubling of distance. In practice, turbulence does not allow complete cancellation of direct and reflected waves so such a rapid falloff is not expected. Nonetheless, atmospheric absorption and surface effects combine to give very rapid attenuation of sources close to the horizon.

GLOSSARY

(For complete definition of commonly used acoustical terms see American National Standards Institute, ANSI S1.1-1960 [R 1976])

Absorption Loss: That part of the transmission loss due to the dissipation or conversion of sound energy into other forms of energy.

Admittance: Inverse of impedance.

atm: One standard atmosphere of pressure $= 1.01\times10^6$ dyn/cm$^2 = 1.01\times10^5$ Pa $= 1.01\times10^5$ N/m^2.

Attenuation Coefficient: Logarithmic rate of change in sound pressure (dB/unit length or neper/unit length).

Creeping Wave: Dominant term in an approximate solution (applicable at long ranges) for sound diffracted into a shadow zone.

Diffracted Wave: One whose front has been changed in direction by an obstacle or other inhomogeneity in a medium.

Echo Sounder (Acoustic Sounder): A device which measures atmospheric conditions by sending and receiving acoustic pulses.

Excess Attenuation: Attenuation in excess of 6 dB/doubling of distance (+dB means added loss).

FFT: Fast Fourier transform.

Impedance: The complex ratio of the sound pressure to the particle velocity. The specific impedance, the ratio of impedance to impedance of air, is generally used.

Infrasound: Sound at frequencies less than 20 Hz.

Insertion Loss: The increase in transmission loss caused by the addition of a barrier.

Mach Number: Ratio of speed to local speed of sound.

neper: Unit of level when the logarithm is on the Naperian base.

Nyquist Frequency: Highest frequency allowed in a digitally sampled signal.

Octave Band: Band of frequencies extending from $f_c/\sqrt{2}$ to $\sqrt{2}\,f_c$, where f_c is the center frequency in Hz.

Pascal: mks unit of measure for pressure (1 N/m^2).

Radio Sounder: Similar to echo sounder but the acoustic wave is monitored with a radar system.

rayl: Unit of measure for impedance; 1 mks rayl $= 1$ kg/m^2 s $= \frac{1}{10}$ cgs rayl.

Rayleigh Waves: Surface waves associated with the free boundary of a solid such that a surface particle describes an ellipse.

Refraction: The process by which the direction of sound propagation is changed due to spatial variation in the speed of sound.

Scattering: Acoustic scattering is the irregular reflection, refraction, or diffraction of sound in many directions.

S/N: Signal-to-noise ratio typically measured in dB.

Sound Pressure Level: 20 times the logarithm to the base 10 of the ratio of the pressure to a reference pressure (conventionally 20 μPa).

Standard Temperature and Pressure: (STP) 273 K, 1 atm. Note that in atmospheric physics the standard atmosphere at sea level (288 K, 1 atm) is commonly used.

Transmission Loss: Decrease in sound pressure level between two points measured in dB re 20 μPa.

Turbule: A single element of turbulence defined in a statistical sense.

Ultrasound: Frequencies above 20 kHz.

List of Works Cited

American National Standards Institute (ANSI) S2.20-1983, *Estimating Airblast Characteristics for Single Point Explosions in Air with a Guide to Evaluation of Atmospheric Propagation and Effects*, New York: American National Standards Institute.

Arnold, R. T., Bass, H. E., Bolen, L. N. (1976), *J. Acoust. Soc. Am.* **60**, 584–593.

Attenborough, K. (1985), *J. Sound Vib.* **99**, 521–544.

Attenborough, K., Hayek, S. I., Lawther, J. M. (1980), *J. Acoust. Soc. Am.* **68**, 1493–1501.

Bass H. E. (1980), *J. Acoust. Soc. Am.* **67**, 1959–1966.

Bass, H. E., Sutherland, L. C., Piercy, J., Evans, L. (1984), in: *Physical Acoustics XVII*, New York: Academic, pp. 145–232.

Berengier, M., Daigle, G. A. (1988), *J. Acoust. Soc. Am.* **84**, 1055.

Brown, E. H. (1987), *Encyclopedia of Physical Science and Technology 3*, New York: Academic, pp. 147–163.

Brown, E. H., Hall, F. F., Jr. (1978), *Rev. Geophys. Space Phys.* **16**, 47–110.

Busnel, R. G. (1977), in: T. A. Sebeok (Ed.), *How Animals Communicate*, Bloomington: Indiana University.

Chessel, C. I. (1977), *J. Acoust. Soc. Am.* **62**, 825–834.

Daigle, G. A. (1982), *J. Acoust. Soc. Am.* **71**, 847–854.

Delany, M. E., Bazley, E. N. (1977), *J. Sound Vib.* **16**, 315–322.

Dept. of the Army (15 June 1978), *Dept. of the Army Technical Manual 5-803-2.*

Isei, T., Embleton, T. F. W., Piercy, J. E. (1980), *J. Acoust. Soc. Am.* **67**, 46–58.

Johnson, M. A., Raspet, R., Bobak, M. T. (1982), *J. Acoust. Soc. Am.* **81**, 638–646.

Lindsey, R. B. (1974), *Physical Acoustics*, Benchmark Papers in Acoustics, 69. New York: American Institute of Physics.

Makarewicz, R. (1987), *J. Acoust. Soc. Am.* **82**, 1706–1711.

Pierce, A. D. (1981), *Acoustics: An Introduction to its Physical Principles and Applications*, New York: McGraw-Hill.

Piercy, J. E., Embleton, T. F. W., Sutherland, L. C. (1977), *J. Acoust. Soc. Am.* **61**, 1403–1418.

Price, M. A., Attenborough, K., Heap, N. W. (1988), *J. Acoust. Soc. Am.* **84**, 1836–1844.

Raspet, R., Lee, S. W., Kuester, E., Chang, D. C., Richards, W. F., Gilbert, R., Bong, N. (1985), *J. Acoust. Soc. Am.* **77**, 345–352.

Raspet, R., Bass, H. E., Ezell, J. (1983), *J. Acoust. Soc. Am.* **74**, 267–274.

Reed, J. W. (1977), *J. Acoust. Soc. Am.* **61**, 39–47.

Reed, J. W. (1987), *J. Geophys. Res.* **92**, 11 979–11 992.

Sabatier, J. M., Bass, H. E., Bolen, L. N., Attenborough, K., Sastry, V. V. S. S. (1986), *J. Acoust. Soc. Am.* **79**, 1345–1352.

Smith, D., Schien, D. (1989), Description of jet aircraft noise provided by employees of the Northrop Corporation.

Tyndall J., (1874), *Philos. Trans. R. Soc. London, Ser. A* **164**, 183–244.

Wei, R. J., Tian, Y. R., Lu, Q. J. (1987), *J. Acoust. Soc. Am.* **81**, 1350–1354.

Further Reading

Bass, H. E., Sutherland, L. C., Piercy, J., Evans, L. (1984), in: *Physical Acoustics XVII*, New York: Academic, pp. 145–232.

Brown, E. H., Hall, F. F., Jr. (1978), *Rev. Geophys. Space Phys.* **16**, 47–110.

Busnel, R. G. (1977), in: T. A. Sebeok (Ed.), *How Animals Communicate*, Bloomington: Indiana University.

Pierce, A. D. (1981), *Acoustics: An Introduction to its Physical Principles and Applications*, New York: McGraw-Hill.

Piercy, J. E., Embleton, T. F. W., Sutherland, L. C. (1977), *J. Acoust. Soc. Am.* **61**, 1403–1418.

Benchmark cases for outdoor sound propagation models

K. Attenborough and S. Taherzadeh
The Open University, Milton Keynes MK7 6AA, England, United Kingdom

H. E. Bass, X. Di,[a] and R. Raspet
The University of Mississippi, University, Mississippi 38677

G. R. Becker and A. Güdesen
Atlas Elektronik GmbH, D-2800 Bremen 44, Germany

A. Chrestman
USAE Waterways Experiment Station, Vicksburg, Mississippi 39180-6199

G. A. Daigle and A. L'Espérance[b]
National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada

Y. Gabillet
Centre Scientifique et Technique du Bâtiment, 38400 Saint-Martin-d'Hères, France

K. E. Gilbert
Pennsylvania State University, State College, Pennsylvania 16804

Y. L. Li and M. J. White
U.S. Army Construction Engineering Research Laboratory, Champaign, Illinois 61826-9005

P. Naz
French-German Institute of Research of Saint Louis, 68301 Saint Louis, France

J. M. Noble
U.S. Army Research Lab, White Sands Missile Range, New Mexico 88002-5501

H. A. J. M. van Hoof
Physics and Electronics Lab/TNO, 2509 J.G. The Hague, The Netherlands

(Received 15 June 1993; revised 12 May 1994; accepted 22 August 1994)

The computational tools available for prediction of sound propagation through the atmosphere have increased dramatically during the past decade. The numerical techniques include analytical solutions for selected index of refraction profiles, ray trace techniques which include interaction with a complex impedance boundary, a Gaussian beam ray trace algorithm, and more sophisticated approximate solutions to the full wave equation; the fast field program (FFP) and the parabolic equation (PE) solutions. This large array of computational approaches raises questions concerning under what conditions the various approaches are reliable and concerns about possible errors in specific implementations. This paper presents comparisons of predictions from the several models assuming a complex impedance ground and four atmospheric conditions. For the cases studied, it was found that the FFP and PE algorithms agree to within numerical accuracy over the full range of conditions and agree with the analytical solutions where available. Comparisons to ray solutions define regimes where ray approaches can be used. There is no attempt to compare calculated transmission losses to measurements.

PACS numbers: 43.28.Fp, 43.20.Bi

INTRODUCTION

Propagation of sound outdoors involves a number of physical phenomena including geometric spreading, molecular absorption, reflection from a complex impedance boundary, refraction, diffraction, and scattering.[1] Accurate predictions of sound transmission loss (TL) from a source to a receiver must somehow account for all of these phenomena simultaneously. Although this goal is beyond current capabilities, a great deal of progress has been made in the past decade. Specifically, numerical techniques originally developed for applications in underwater acoustics have been modified for atmospheric predictions and new analytical results provide solutions for more realistic conditions. Some progress has been made toward including turbulence and terrain in the propagation models, but research on these aspects is still in progress.

Ray tracing has been a standard approach extensively used to predict outdoor sound levels in the presence of speed of sound gradients even though the mathematical conditions

[a] Also with National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada.
[b] Also with Université de Sherbrooke, Sherbrooke, Quebec J1K 2R1, Canada.

 0001-4966/95/97(1)/173/19/$6.00

for ray solutions are seldom met outdoors. Recent analytical solutions include a refractive atmosphere and a complex impedance boundary[2,3] but are only applicable to cases where the speed of sound varies with altitude in a specific functional form. Numerical solutions to the wave equation include the diffraction effects which limit the applicability of ray solutions and can be applied to arbitrary speed of sound profiles.[4,5] For this reason, the development of numerical algorithms to predict transmission loss through the atmosphere is being pursued by a number of groups throughout the world.

Research groups in many countries have developed propagation models to predict transmission loss. The models can be generally classified as analytic (such as normal mode models), ray trace, a variation of the fast field program (FFP), or a form of the parabolic equation (PE). Though many of the groups involved share results at international meetings, the codes were generally developed independently. Members of NATO Panel 3 Research Study Group 11 (RSG-11) began an effort in 1990 to compare predictions of the various models and different implementations of the same model. The goals of this effort were to identify any differences and sources of those differences and to provide a set of results which future researchers can use to check new models and codes.

Thus the objective of this paper to look for consensus between models or, more accurately, significant differences, to investigate regimes of validity and to examine artifacts. We have chosen not to evaluate accuracy in any systematic way nor computational requirements since these change rapidly.

Assessing the quality of numerical computation schemes is an issue that has recently received considerable attention in underwater acoustics.[6] At technical meetings it became common to find numerical results presented in profusion that differed for the same problem. Researchers in underwater acoustics therefore explored the utility of benchmark solutions as standards against which numerical codes could be tested. The success of their efforts inspired the members of RSG-11 to also develop benchmarks.

We define a "benchmark" throughout this paper as a well-defined environment given in terms of a set of numerical parameters. This set of parameters represents the input data for model runs.

Comparison of different codes and models requires selection of suitable atmospheric conditions, ground surface, frequencies, and ranges. These were limited, in many cases, by the current state of development of the different approaches. Since most models do not include turbulence, quiet atmospheres were assumed. Similar reasons led the group to consider only flat terrain. Even with these limitations, however, much of the physics is included.

The absence of turbulence or real terrain in the calculations severely limits comparison to experimental data. In fact, no such comparisons are made here. There are limited comparisons of data to individual models in many of the references cited. In those cases, however, differences are often attributed to atmospheric inhomogeneities. That may well be the reason, but it involves an assumption not necessary to meet the goals of this study. If predictions based upon the same set of assumptions do not agree, the reason for the difference should provide guidance as to the applicability of the models.

In the case of the underwater acoustic benchmarks results of individual contributions were published in *The Journal of the Acoustical Society of America* following the format presented in Ref. 6. In the present case members of RSG-11 concluded that much of the material used in the atmospheric benchmarks could be found in existing archival papers and that it would be more expedient to publish the results as a single summary paper.

The next section explains in more detail the propagation conditions assumed for the benchmark calculations. The conditions selected represent cases where there is no refraction, downward refraction, upward refraction (including shadow zones), and ducted propagation. The exact profiles are not realistic but these conditions should provide good tests of predictive abilities of the different approaches. Further, the exact profiles were chosen because there are analytical solutions for the first three cases. Section II describes the different models. In some cases, differences in implementations are also described. Although the issue of CPU time is discussed, a comprehensive comparison is beyond the scope of this paper and is not considered in detail here. In the fourth section, results are given and interpreted for selected cases. A complete set of results is given only for the FFP/PE calculations (these two approaches are usually in agreement to within about 0.5 dB) since it is these two methods that appear most accurate and versatile.

I. DEFINITION OF BENCHMARK CASES

The aim of the benchmark cases outlined in this section is to create a set of well-defined test cases for which computer codes for airborne sound propagation, of any origin, may be compared with respect to computational effort and numerical accuracy. It is intended to check numerical performance of computer codes by comparing their outputs with test cases for which analytical solutions exist. Researchers of other laboratories are invited to use the standard test cases and to compare their results with examples discussed in Sec. III.

A. General considerations

Benchmark cases for numerical tests of propagation codes should be defined in such a way as to cover the range between "simple" cases for which analytical solutions might be available and relevant cases with boundary conditions close to real world situations. Due to the large variations in propagation conditions of the real world, cases must be idealized; e.g. simplified in such a way as to allow for complete analytical solutions or at least piecewise analytical solutions. This is also motivated by the need to study typical numerical limitations such as

(1) discretizing sound profiles,
(2) truncation of numerical approximations,
(3) capping of profiles at a certain height.

Aside from the "simple" case or cases which can serve as the basis for computing a numerical reference it is very

difficult to define other cases which are both realistic and few in number and still possibly rewarding to overcome the problems of complexity due to all possible variations of parameters. Another problem area is the types of outputs one needs to compare the outputs of various codes; e.g., whether one cares for the details in the interference nulls or just for the average level. Again a compromise has to be found in order to keep the number of graphs as low as possible without neglecting essential features of the codes. A description of the benchmark cases is given below.

B. Common parameters for benchmark cases

1. Basic parameters

All comparisons rely on a model geometry of two infinite half-spaces, the airspace and the groundspace. Unless otherwise noted the interface between both spaces is regarded as a smooth infinite plane. There is an acoustic point source at source height h_s above the interface and a receiving microphone at a horizontal range R and receiver height h_r. The source emits a constant tone of frequency f.

The following quantities were chosen for calculations:

$h_s = 5$ m (source height),

$h_r = 1$ m (receiver height),

$R = 200$ m; 10 000 m

(horizontal range to receiver),

$f = 10; 100; 1000$ Hz (frequency).

The transmission loss (TL) was considered to be the most useful quantity to be used for comparison purposes. The following definition is used:

$$\mathrm{TL} = -20 \log \frac{(\text{total acoustic pressure at a field point})}{(\text{acoustic pressure of direct sound field at 1 m from source})}. \quad (1)$$

2. Plot parameters

For ease of quantitative comparison the graphical outputs of the various propagation codes are standardized by using the following three types of output graphs:

1. TL versus horizontal range.
2. TL versus frequency at $R = 200$ m and $R = 10\,000$ m.
3. Contour plot: Height versus range with TL contours at 6-dB increments, maximum height being 1000 m.

C. Characterization of the ground

Ground impedance plays a major role in airborne sound propagation near the surface. Typically outdoor ground surfaces are uneven and their character will vary with range but here it is assumed

(a) that the ground is transversely uniform and its properties do not vary with range,

(b) that all interfaces are flat and horizontal.

The properties of the ground surface that have the most influence on outdoor propagation below 1000 Hz are its flow resistivity and any near surface layering. Typically the ground surface may be treated as an impedance boundary. At higher frequencies the tortuosity and porosity will be important. Moreover in predicting the response of a buried geophone to acoustic sources above the surface the elasticity of the ground is important as well as its porosity and flow resistivity (see, for example, Ref. 23). The seismic profile of the ground is required as input to the SAFARI code,[21] originally developed for use in underwater acoustics. It has been found necessary to specify sufficient parameters to allow inputs to any of the codes that are compared in this paper even though the resulting complication may not be needed simply for predicting sound in the atmosphere. A series of careful and comprehensive measurements have been carried out by various authors on a site at Wezep, The Netherlands[7] revealing a homogeneous and isotropic layer of porous sand, 2-m deep, above a nonporous substrate. The resulting values are given in Table I. For those codes that require input only of surface impedance, a four parameter model[8] has been used. The specific characteristic impedance of the ground, and hence the specific surface impedance where the ground is acoustically homogeneous, may be calculated from

$$Z_c = \omega \rho_b(\omega)/(k_b \rho_0 c_0), \quad (2)$$

where

TABLE I. Parameters used to characterize the ground and air at surface.

Parameters	Value
Flow resistivity (σ)	366 000 Pa s m^{-2}
Porosity (Ω)	0.27
Pore shape factor (s_p)	0.25
Grain shape factor (n')	0.5
(N.B. tortuosity $q^2 = \Omega^{-n'}$)	(tortuosity=1.925)
N_{Pr} (Prandtl number)	0.724
γ (ratio of specific heats)	1.4
ρ_0 (air density at 20 °C)	1.205 kg m^{-3}
c_0 (speed of sound at 20 °C)	343.23 m s^{-1}
Upper p velocity	270 m s^{-1}
Upper s velocity	190 m s^{-1}
Lower p velocity	500 m s^{-1}
Lower s velocity	330 m s^{-1}
Solid bulk modulus	4.6×10^{11} Pa
Upper layer thickness	2 m
Bulk density	1700 kg m^{-3}
p,s damping	0.02 dB per wavelength

$$k_b^2=(q\omega/c_0)^2[1-2(\lambda\sqrt{i})^{-1}T(\lambda\sqrt{i})]^{-1}[1+2(\gamma-1)\times(N_{\mathrm{Pr}}{}^{1/2}\lambda\sqrt{i})^{-1}T(N_{\mathrm{Pr}}{}^{1/2}\lambda\sqrt{i})],$$

$$\rho_b(\omega)=(q^2/\Omega)\rho_0[1-2(\lambda\sqrt{i})^{-1}T(\lambda\sqrt{i})]^{-1},$$

$$q^2=\Omega^{-n'}, \quad T(x)=J_1(x)/J_0(x),$$

and

$$\lambda=(1/2s_p)[8\rho_0 q^2\omega/\Omega\sigma]^{1/2}.$$

Using the values in Table I, the real and imaginary parts of impedance given by Eq. (2) are (38,79,38.41), (12.81,11.62), and (5.96,2.46) at 10, 100, and 1000 Hz, respectively.

D. Speed of sound profiles

1. General

All test cases assume the following general boundary conditions:

i. Atmospheric density and pressure are assumed constant.

ii. The density ρ_0 of the air at ground level is 1.205 kg/m^3 at $T=20$ °C and $p=1$ atm.

iii. The attenuation coefficient, α for atmospheric absorption is taken at a relative humidity (RH) of 70% and at a temperature $T=20$ °C. The frequency dependence of α is as defined by ANSI S1.26.[9]

iv. Under the conditions as above the sound velocity c_0 at the surface will be $c_0=343.23\approx343$ m/s.

2. Test cases

Illustrations of the test cases are depicted in Fig. 1.

Case 1 is the above mentioned "simple" case to study numerical effects of propagation codes. The profile simply consists of an isovelocity (homogeneous) medium with a constant sound speed c_0. For numerical reasons, the predictions for this case are sensitive to the numerical truncation of the medium to at a finite height.

Case 2 is a strong positive sound speed gradient as occurs under "downwind" conditions. The real world conditions were idealized to a linear profile with a constant gradient of 0.1 s^{-1}. No assumptions are made with respect to an upper boundary of the linear increase of sound speed. This case is intended to stimulate investigations on the effects of profile height on the sound field at a given range.

Case 3 depicts another idealized real world situation: Sound propagation under "upwind" conditions. For ease of systematic investigations only the sign of the gradient was changed from case 2. All other aspects discussed for case 2 apply to case 3 as well.

Case 4, as can be seen in Fig. 1, is a composite profile. This ducting type of sound profile is a rather typical observation in many real situations. To keep all cases systematic, case 4 includes only elements of earlier discussions. The profile starts at the surface with a positive gradient up to a height of 100 m. The gradient is constant at a rate of 0.1 s^{-1}. At a height of 100 m an inversion occurs yielding a constant negative gradient of -0.1 s^{-1}. The negative gradient continues to 300 m height. At this point the sound speed assumes a constant value to infinite height.

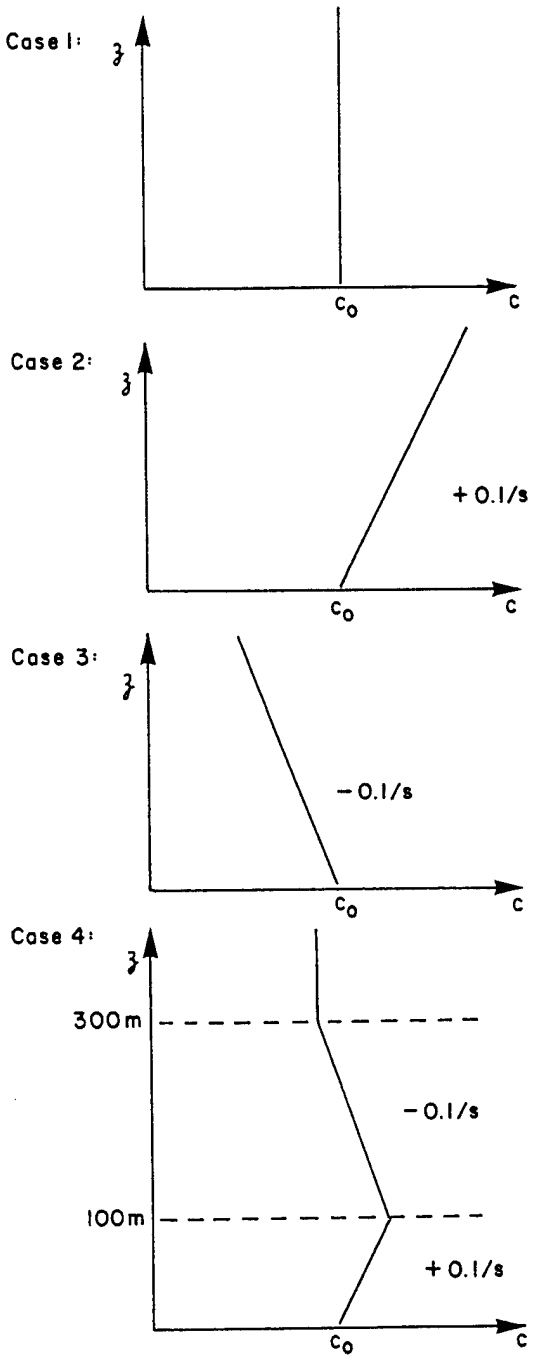


FIG. 1. Speed of sound profile for the four test cases.

3. Contour plots

Plots of height versus range with TL contours at 6-dB increments, maximum height being 1000 m, were also produced for the four cases.

II. DESCRIPTION OF THE MODELS

In this section the different models and the different implementations of the same model used in the benchmarks are briefly summarized. A complete and detailed description of each model can be found in the references cited in the text below. The analytical solutions for cases 1, 2, and 3 are first

described. No analytical solution for case 4 is discussed. Ray tracing is extensively used in atmospheric sound propagation. Therefore, next, three ray tracing based models are described. Third, the numerical models are discussed. These include four implementations of the FFP and two implementations of the PE.

A. Analytical wave solutions

We start with the classical wave equation for the acoustic pressure:[10]

$$\left(\nabla^2-\frac{1}{c^2(z)}\frac{\partial^2}{\partial t^2}\right)p(r,z,t)=-4\pi\delta(x,y,z-z_s,t), \quad (3)$$

where δ represents a delta-function source of unit strength, $c(z)$ is the speed of sound, and z_s is the height coordinate of the source.

Assuming simple harmonic time dependence $\exp(-i\omega t)$, Eq. (3) becomes the Helmholtz equation

$$(\nabla^2+k^2)p(r,z)=-4\pi\delta(r,z-z_s), \quad (4)$$

where the wave number $k=\omega/c(z)$.

In cylindrical coordinates (r,θ,z) and for cylindrical symmetry so that there is no variation with θ, the Helmholtz equation becomes

$$\frac{\partial^2 p}{\partial r^2}+\frac{1}{r}\frac{\partial p}{\partial r}+\frac{\partial^2 p}{\partial z^2}+k^2 p=-\frac{2}{r}\,\delta(r)\,\delta(z-z_s), \quad (5)$$

where the source has been assumed to be at $r=0$.

Equation (5) can be solved with a zero-order Hankel transform[10]:

$$p(r,z)=-\int_{-\infty}^{\infty} H_0^1(Kr)P(K,z)K\,dK, \quad (6)$$

where H_0^1 is the Hankel function of the first kind and of order 0 and $P(K,z)$ satisfies

$$\frac{d^2P(K,z)}{dz^2}+[k^2(z)-K^2]P(K,z)=-\delta(z-z_s). \quad (7)$$

In Eq. (7), K represents the horizontal component of the wave number. The function $P(K,z)$ must satisfy an impedance boundary condition at the ground, must be continuous at the source, have a discontinuous derivative with height at the source and must satisfy a radiation boundary condition at large height.

Analytical solutions for $P(K,z)$ can be found when the sound-speed variation with height is approximated by

$$c(z)=\frac{c(0)}{\sqrt{1-2az}}\approx c(0)(1+az) \quad (8)$$

in the case of downward refraction and

$$c(z)=\frac{c(0)}{\sqrt{1+2az}}\approx c(0)(1-az) \quad (9)$$

in the case of upward refraction. In the above a is defined by

$$a=\frac{1}{c(0)}\left(\frac{dc}{dz}\right)_0,$$

and $c(0)=c_0$ is the sound speed at the ground surface. As noted, these profiles approximate linear variations near the ground.

If $k^2(z)$ is linear, $P(K,z)$ can be expressed in terms of Airy functions. Airy functions are desirable since they have no branch cuts in the complex plane, which simplifies the analysis and the search for poles. To obtain the acoustic pressure $p(r,z)$, the expression for $P(K,z)$ in terms of Airy functions is substituted into Eq. (6), the residue of the integrand is calculated at each pole of the integral, and the results summed to form the total solution. Note that the effect of the ground impedance on these solutions, as well as the effect of the ground impedance on the normal modes is discussed in detail in Refs. 2 and 3.

1. Normal modes solutions for downward refraction (case 2)

The normal mode solution[3] is the residue series for the acoustic pressure corresponding to downward refraction (case 2):

$$p(r,z)=\frac{i\pi}{l}\sum_n \frac{H_0^1(k_n r)\mathrm{Ai}(\tau_n+z_s/l)\mathrm{Ai}(\tau_n+z/l)}{\tau_n[\mathrm{Ai}(\tau_n)]^2-[\mathrm{Ai}'(\tau_n)]^2}, \quad (10)$$

where $\tau_n=(k_n^2-k_0^2)l^2$ are the zeros of $\mathrm{Ai}'(\tau_n)+\gamma\,\mathrm{Ai}(\tau_n)=0$, $k_0=\omega/c(0)$ and k_n is the wave number of the nth mode. The abbreviations are

$$\gamma=(ik_0 l\rho c)/Z, \quad l=(r_c/2k_0^2)^{1/3},$$

where Z is the acoustic impedance and r_c is the radius of curvature of the ray paths (which are arc of circles in the case of a linear profile). The downward refracting case does not converge rapidly since the poles mostly lie close to the real axis. The number of modes necessary to accurately evaluate the downward refraction integral may be approximated by

$$n_{\max}=\frac{2}{3}\left(f\Big/\left|\frac{dc}{dz}\right|_0\right),$$

where f is the frequency. If the gradient expressed in Eq. (8) is truncated at a given height, the series will not contain modes which are reflected from the gradient above this height. In most cases, the nth mode height, h_n, may be approximated by $h_n\approx(3\pi n^{2/3}/2)l$. At long ranges, the largest contributions to the integral Eq. (6) result from poles of $P(K,z)$ which arise from the ground reflected term. Hence the solution is accurate at sufficient range for the direct wave to be negligible with respect to the sum of the residues of the integral.

2. Residue series for upward refraction (case 3)

The residue series for the acoustic pressure for upward refraction (case 3) is[10]

$$p(r,z)=\frac{\pi e^{i\pi/6}}{l}\sum_n \frac{H_0^1(k_n r)\mathrm{Ai}[b_n-(z_s/l)e^{2i\pi/3}]\mathrm{Ai}[b_n-(z/l)e^{2i\pi/3}]}{[\mathrm{Ai}'(b_n)]^2-b_n[\mathrm{Ai}(b_n)]^2}, \quad (11)$$

where $b_n=(k_n^2-k_0^2)l^2\exp(2i\pi/3)$ are the zeros of $\mathrm{Ai}'(b_n)+\gamma e^{2i\pi/3}\mathrm{Ai}(b_n)=0$. The other quantities are the same. The upward refracting case converges rapidly since the absorption increases very rapidly for higher-order poles.

3. Homogeneous medium (case 1)

In the absence of refraction (case 1), the basic problem is described in Fig. 2. The sound emitted by a monopole source S travels to a receiver along a direct ray path and a ray path reflected from the ground with a grazing angle ψ. The ray paths are straight. In this case the solution to Eq. (3) for the acoustic pressure at the receiver can be written in the following form[11]

$$p(r)=[A(R_1)/R_1]\exp(ikR_1)+Q[A(R_2)/R_2]\exp(ikR_2), \quad (12)$$

where R_1 and R_2 are the path lengths of the direct and reflected paths respectively, and the amplitude $A(R)$ accounts for atmospheric absorption. The quantity Q is a reflection coefficient which has been modified for spherical waves reflecting from a complex plane boundary. A good approximation for Q is

$$Q=R_p(\psi)+B[1-R_p(\psi)]F(w), \quad (13)$$

where $R_p(\psi)$ is the plane-wave reflection coefficient and the boundary loss function $F(w)$ is defined in terms of the numerical distance w and complimentary error function:

$$F(w)=1+i(\pi)^{1/2}we^{-w}\,\mathrm{erfc}(-iw). \quad (14)$$

In Eqs. (13) and (14), $B\approx 1$ is a correction term close to unity which can be set to unity at near grazing incidence. Expressions for $R_p(\psi)$, B, and w in the general case of a ground of extended reaction can be found in Ref. 11.

For most purposes it is sufficient to assume locally reacting ground and to set $B=1$. In this case

$$R(\psi)=\frac{\sin\psi-1/Z}{\sin\psi+1/Z} \quad (15)$$

and

$$w^2=\tfrac{1}{2}ikR_2(\sin\psi+1/Z)^2, \quad (16)$$

where Z is the normalized acoustic impedance of the ground. We note that both R_p and w are functions of the grazing angle ψ and that $R_p\to -1$ when $\psi\to 0$. Further, for high frequencies $Q=R_p$.

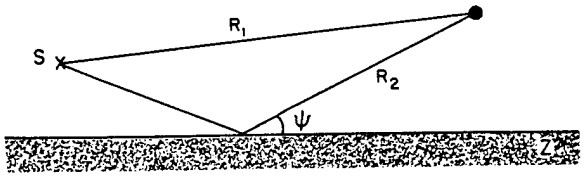


FIG. 2. Schematic showing geometrical definitions for direct and reflected ray paths in the absence of refraction (case 1).

B. Ray tracing and hybrid solutions

1. Inhomogeneous atmosphere

The effects of curved ray paths on the total sound field can be described from general principles even for complex sound speed profiles. Figure 3 illustrates curved direct and reflected ray paths for weak refraction in the case of downward and upward refraction, respectively. The total path lengths R_1 and R_2 are modified thus changing the path length difference between direct and reflected paths. More importantly, the effective grazing angle ψ for the reflected ray is different. In the case of downward refraction, the grazing angle for the reflected ray is greater than in the absence of refraction. This has the consequence that the reflection coefficient $R(\psi)$ deviates further from -1, the destructive interference between direct and reflected waves becomes less complete and their geometrical interference is shifted to higher frequencies. In the case of upward refraction, the grazing angle for the reflected ray is smaller than in the absence of refraction. The reflection coefficient tends more toward -1, the destructive interference is enhanced and the geometrical interference is shifted to lower frequencies.

In the cases of extreme downward refraction or at longer ranges, there are usually many ground reflected paths. There is no general ray tracing solution for this case. In the case of upward refraction at long ranges, the receiver can be beyond the shadow boundary and according to ray theory there is no sound. In this situation, wave theory, or wave extensions of ray theory must be used.

The following describes three models based on ray theory. The first model is restricted to a single bounce but assumes an arbitrary sound speed profile. This model was designed to predict sound propagation from aircraft and does

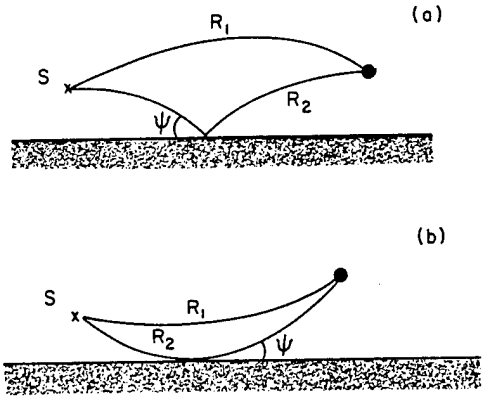


FIG. 3. Curved ray paths in the presence of a (a) downward refracting atmosphere (case 2) and (b) upward refracting atmosphere (case 3).

not account for multiple reflections from the ground. The second model assumes a linear sound speed profile but can account for the effects of multiple bounces. The third model is an extension of ray theory to account for wave effects.

2. The single bounce arbitrary profile model

In the cases of mild downward or upward refraction, the model ASOPRAT[12] uses a raytrace program to obtain the total path lengths, R_1 and R_2, and the grazing angle. Also, propagation constant and attenuation coefficients are computed from the raytrace program information about the paths R_1 and R_2. These calculated total path lengths and modified grazing angle are then used in Eqs. (15) and (16) and the resulting numerical distance and reflection coefficient are used to calculate the total field using Eq. (12).

When there is strong downward refraction or at long ranges where there are multiple reflected ray paths, ASOPRAT only accounts for one reflected path. The shortest ground-reflected path for R_2 is chosen since it is usually strongest in amplitude. The reflected paths that strike the ground more than once are disallowed under the assumption that ground reflections produce a loss in amplitude and coherence and paths with fewer reflections usually have larger amplitudes.

In ASOPRAT, the atmosphere is divided into horizontal layers which have different values of sound speed. The sound speed and wind speed are assumed to vary linearly with height between interfaces. This assumption permits solution of the acoustic ray equation in closed form but can sometimes lead to false caustics caused by sound speed slope discontinuities. An image atmosphere has also been inserted so reflected eigenrays can be treated as direct eigenrays.

For horizontally stratified media, the ray paths are given by Snell's law,

$$\frac{\cos\,\theta(z)}{c_z}=\xi=\frac{\cos\,\theta_s}{c_s}, \tag{17}$$

where c_s is the sound speed at the source and ξ is a constant. The initial angle measured from horizontal is θ_s with upward-sloping rays having positive launch angles. If the sound speed gradient is constant ($g=dc/dz$), the ray paths are arcs of circles with radius of curvature $r_c=1/\xi g$.

3. The multiple bounce linear profile model

This heuristic model[13] assumes that many realistic sound speed profiles can be approximated by a linear sound speed profile. The assumption of a linearly varying sound speed profile permits solutions of the acoustic ray equations in closed form. In the case of weak refraction there is the direct ray path and one reflected ray path (Fig. 3) and one can calculate the effective geometrical parameters of these two rays. In Eq. (12), the geometrical interference resulting from the path length difference between direct and reflected rays is represented by $k(R_2-R_1)$. However, because the sound speed varies with the height, the wave number k is not constant over all the rays. To consider such phenomena, the model uses the difference in the travel times between the rays instead of the difference in path lengths. Applying Eq. (12), the mean square sound pressure becomes

$$p^2(r)=\frac{A(R_1)^2}{R_1^2}+\frac{A(R_2)^2|Q|^2}{R_2^2}+\frac{2A(R_1)A(R_2)|Q|}{R_1R_2}\times\cos[2\pi f(\tau_2-\tau_1)+\mathrm{Arg}(Q)], \tag{18}$$

where τ_1 and τ_2 are the effective travel times for R_1 and R_2 computed using equations found in Ref. 13.

In the presence of a strong positive sound speed gradient, and/or for larger propagation distances, additional rays that go through n reflections on the ground appear between the source and the receiver. These additional rays can be determined using the following 4th order equation

$$n(n+1)x^4-(2n+1)Rx^3+[b_r^2+(2n^2-1)b_s^2+R^2]x^2-(2n-1)b_s^2Rx+n(n-1)b_s^4=0, \tag{19}$$

where $b_i^2=(z_i/a)(2+az_i)$ for $i=s$ or r and R is the horizontal distance between source and receiver. In Eq. (19), n is the number of reflections and the unknown x is the horizontal distance between the source and the first reflection on the ground. This equation must be solved successively from $n=0,1,2,3,\ldots$ (the number of reflections on the ground), until there is no real solution for x, i.e., as long as $0<x<R$. In presence of a positive gradient, there is at least one direct ray (for $n=0$) and one reflected ray (for $n=1$). When the receiver moves closer to the ground and further from the source, or when the gradient increases, the complex conjugate of the roots may become real, which means that additional reflected rays have to be considered. For $n=2$, two additional reflected rays may appear, and for $n>2$, four additional reflected rays may appear for each n.

Defining N as the total number of rays reaching the receiver (including the direct ray), the total sound pressure at the receiver can be calculated by summing up the contributions of all the N rays involved. This is done in the model using the following expression.

$$p^2(r)=\sum_{i=1}^{N}\frac{A_i^2|Q_i|^2}{R_i^2}+2\sum_{i=2}^{N}\sum_{j=1}^{i-1}\frac{A_i|Q_i|A_j|Q_j|}{R_iR_j}\times\cos\left[\omega(\tau_j-\tau_i)+\mathrm{Arg}\left(\frac{Q_j}{Q_i}\right)\right], \tag{20}$$

where $i=1$ denotes the direct ray, and therefore $Q_1=1$. A_i represents the standard attenuation of a single ray due to atmospheric absorption, computed using the refracted length of the path, R_i. τ_i is the travel time of the ray, Q_i is the equivalent reflection coefficient on the ground of this ray calculated with

$$Q_i=Q(\psi_i)^{n_i}, \tag{21}$$

where ψ_i is the angle of reflection on the ground, and n_i is the number of reflections.

The case of a weak sound speed gradient is therefore defined by putting $N=2$; in this case there is only one direct and one reflected ray, and Eq. (20) reduces to Eq. (18).

4. Gaussian beam approach

The basic concept of the theory is to launch a fan of beams from the source and to trace the propagation of these beams through the medium. The wave equation is solved in the immediate vicinity of each ray and the acoustic pressure at a receiver is obtained by summing the contribution of each of the individual beams. The solution is everywhere uniform thus removing the singularities at caustics found in the traditional ray tracing methods. The procedure is not sensitive to the exact details of the medium since, contrary to ray methods, the Gaussian beam provides a local averaging. It is not necessary to find the rays that exactly intercept the receiver, thus reducing computation time.

The first step of the Gaussian beam method is to solve the classical ray equations to obtain the central ray of the beam:

$$\frac{d\mathbf{r}}{dt}=\mathbf{V}_e+[\mathbf{k}/|k|]\cdot c \tag{22}$$

$$\frac{d\mathbf{k}}{dt}=-|k|\text{grad } c-\sum k_i \text{ grad } V_e, \tag{23}$$

where $\mathbf{V}_e$ is the vector wind speed, k is the wave number, c is the speed of sound, r is the position vector along the ray, and d/dt is the derivative with respect to time. These equations can be solved by standard numerical techniques or analytically in the case of a linear profile.

The second step of the Gaussian beam method is to solve the wave equation locally in ray-centered coordinates using the parabolic approximation. The ray-centered coordinate system is an orthogonal curvilinear system that follows a particular ray. In a two-dimensional medium, the ray centered coordinates can be specified by the unit vector $\mathbf{t}$ tangent to the ray in the (x,z) plane and the unit vector $\mathbf{n}$ normal to the ray. In the case of a point source, the solution of the local parabolic equation (see Sec. II D) is[14]

$$u(s,n,\omega)=\left(\frac{c(0)q(0)}{c(s)q(s)}\right)^{1/2}\times\exp\{-i\omega[\tau(s)+\tfrac{1}{2}M(s)n^2]\}, \tag{24}$$

where s is the distance along the central ray from the source and n represents the length in a direction perpendicular to the ray at s, $c(s)$ is the sound speed, $\tau(s)$ is the travel time along the central ray, and the function $M(s)$ is given by

$$M(s)=p(s)/q(s), \tag{25}$$

where

$$p(s)=\epsilon p_1(s)+p_2(s), \tag{26}$$

$$q(s)=\epsilon q_1(s)+q_2(s), \tag{27}$$

and

$$\epsilon=\epsilon_1+i\epsilon_2,\quad \epsilon_2\geqslant 0. \tag{28}$$

The complex number ϵ is the Gaussian beam parameter and $\{p_i, q_i, i=1,2\}$ are the two real, independent solutions of the following ray tracing system

$$\frac{dp_i}{ds}=-[c_{nn}/c^2(s)]q_i(s) \tag{29}$$

and

$$\frac{dq_i}{ds}=c(s)p_i(s), \tag{30}$$

where $c_{nn}=\partial^2 c(s)/\partial n^2$ denotes the second normal derivative of the sound speed $c(s)$. The condition that $\epsilon_2\geqslant 0$ in Eq. (28) guarantees that the energy is localized in the vicinity of the central ray. Equations (29) and (30) are solved step by step along each ray by standard numerical techniques in the same way as the ray equations. The choice of the Gaussian beam parameter ϵ and the initial condition at $s=0$ in the case of atmospheric sound propagation are discussed in Ref. 15.

The third step of the Gaussian beam method is a superposition of all Gaussian beams passing in the neighborhood of the receiver. We designate α as the launch angle of a ray with respect to an arbitrary axis passing through the source. The total field at a point located at the receiver is

$$p(r,\omega)=\int \phi(\alpha,\omega)u(s,n,\omega)d\alpha, \tag{31}$$

where $\varphi(\alpha,\omega)$ is the weight function determined by expanding the wave field at the source and matching the high frequency asymptotic behavior of the integral in Eq. (31) to the exact solution for the source in a homogeneous medium.

C. FFP models

1. Introduction

The fast field program (FFP) technique was developed for the prediction of underwater sound propagation[16–18] and has been adapted to propagation in the atmosphere by several authors. Four such adaptations are called CERL-FFP,[4,19] CFFP,[20] SAFARI,[21] and FFLAG.[23]

Full details of there adaptations will not be given here. Descriptions will be limited to outlines of basic formulations and differences. Further information on standard FFP methods may be found in two tutorial articles.[24,25] Fast field programs permit the prediction of sound-pressure level in a horizontally stratified atmosphere at an arbitrary point on or above a flat continuous ground, with range-independent properties, from a point source somewhere in the same half space. They allow specification of effective sound speed as arbitrary functions of height above the ground.

The basis of the FFP method is to work numerically from exact integral representations of the sound field within each layer in terms of coefficients that may be determined from boundary conditions. The method gets its name from the discrete Fourier transform used to evaluate these integrals.

2. Basic formulation

By taking the Hankel transform of Eq. (5) it is possible to remove the r dependence. Writing the zero order Hankel transform of p and its inverse as

$$p(r,z)=\int_0^\infty P(K,z)J_0(Kr)K\,dK$$

and

$$P(K,z)=\int_0^\infty p(r,z)J_0(Kr)r\,dr \tag{32}$$

it is possible to obtain

$$\frac{d^2P}{dz^2}+[k^2(z)-K^2]P=-2\delta(z-z_s), \tag{33}$$

where K is the horizontal wave number.

Equation (33), known as the height (or depth)-dependent transformed wave equation, reduces the problem to one that is one dimensional and forms the starting point of the FFP or Green's function method. The solution ($P(K,z)$) to Eq. (33) is the sum of a particular solution ($\hat{P}(K,z)$) and any linear combination of the two independent solutions [$P^-(K,z)$ and $P^+(K,z)$] to the corresponding homogeneous equation (where the right-hand side is zero). Hence

$$P(K,z)=\hat{P}(K,z)+A^-(K)P^-(K,z)+A^+(K)P^+(K,z), \tag{34}$$

where $A^-(K)$ and $A^+(K)$ are arbitrary coefficients to be determined from the boundary conditions. The most convenient choice for the particular solution is the field produced by the source or sources in the absence of boundaries. When the unknown coefficients have been found, the total field at angular frequency ω is found at any range r by carrying out the inverse Hankel transform according to Eq. (32).

The various FFP methods differ initially according to whether the sound speed gradient within each layer is assumed to be zero (as in the CERL-FFP and FFLAGS) or constant (as in CFFP and SAFARI). The form assumed for the transformed pressure potential within homogeneous layers, $\phi(K,z)$,

$$\phi(K,z)=A^-e^{-\alpha_i z}+A^+e^{\alpha_i z} \tag{35}$$

and within constant gradient layers in which $c(z)=[1/\rho(az+b)]^{1/2}$,

$$\phi(K,z)=A^+\mathrm{Ai}(\zeta)+A^-\mathrm{Bi}(\zeta), \tag{36}$$

where Ai and Bi are Airy functions of the first and second kind, respectively, and ζ is a transformed height variable.

CERL-FFP and CFFP treat the ground surface as an impedance boundary, while SAFARI allows the ground to be layered and elastic and FFLAGS permits ground layering, elasticity and porosity. Consequently SAFARI introduces two potentials in the form of Eq. (36) in each elastic ground layer and FFLAGS introduces three potentials in the form of Eq. (35) in each porous and elastic ground layer. Boundary conditions are solved to determine the unknown coefficients. CERL-FFP and CFFP use a transmission line method[24] to do this whereas FFLAGS and SAFARI use a global matrix method.[21]

After application of the appropriate number of boundary conditions to determine the unknown coefficients it remains to evaluate the Hankel transform integrals of the form of Eq. (32). Typically the Bessel function is replaced by the sum of two Hankel functions and the outgoing wave Hankel function is replaced by its asymptotic form.

This results in integrals of the form

$$G(r,z)=\sqrt{\frac{1}{2\pi r}}\,e^{i\pi/4}\int_0^\infty g(K,z)e^{-ikr}\sqrt{K}\,dK. \tag{37}$$

The indefinite integrals are then replaced by finite sums using discrete Fourier transforms. If the maximum value of wave number in the sum is $K_{\max}$ and N discrete values of K are introduced, then the wave number intervals are given by $\Delta K=K_{\max}/(N-1)$ and correspond to range intervals $\Delta r=2\pi/N\Delta K$, so for example,

$$p(r_m,z)=2(1-i)\sqrt{\pi/r_m}\Delta K\sum_{n=0}^{N-1}P(K_n)\sqrt{K_n}e^{2i\pi nm/N}, \tag{38}$$

where $K_n=n\Delta K$ and $r_m=m\Delta r$ (or $r_0+m\Delta r$, where r_0 is the desired starting range). Various numerical difficulties follow from the truncation of the integral to a finite sum and from the behavior of the integrand. For a detailed discussion of these see Refs. 21 and 22.

Different methods of dealing with these difficulties are used in CERL-FFP and in SAFARI or FFLAGS. In the FFT implementations employed in CERL-FFP, SAFARI, and FFLAGS, the relationship $\Delta K\Delta r=2\pi/N$ must hold. This does not allow the user to specify the desired range points independently of the step sizes used for the horizontal wave number K (i.e., the frequency range). CFFP uses the following range intervals

$$\Delta r=2n\pi/K_{\max}, \tag{39}$$

where n is any real number and evaluates the sums in Eq. (38) by means of the chirp-z transform.[20] This allows freedom both to select the desired ranges and to decrease the ΔK adaptively for accurate evaluation of the integral transform without changing the output range values.

3. Execution times (CPU times)

It is beyond the scope and purpose of this paper to provide a detailed discussion of computational effort since each code has been run under different conditions on different machines. Nonetheless the following qualitative remarks can be made. The CPU time is directly proportional to the number of integration points and therefore the CPU time is almost proportional to the range. Increasing the frequency requires that the integration limit be increased and therefore one has to increase the number of integration points to obtain a given range. Thus in an indirect way the frequency also influences the CPU time. The CPU time also depends on the number of air layers (and ground layers for FFLAGS) used in the FFP. Increasing the number of layers requires more CPU time for evaluation of the height-dependent Green's function. Typically the codes takes minutes to run on a fast workstation.

D. PE models

1. Introduction

The parabolic equation (PE) method has been a very useful technique for solving a variety of wave propagation problems. It has been used in optics and electromagnetic studies,[26] seismic prediction problems,[27] underwater acoustics,[28] and more recently in atmospheric sound propagation.[5,29–31] The technique employs an assumption that wave motion for a particular problem is always directed away from the source or that there is very little backscattering. The advantage in making this assumption is that it reduces a boundary value problem to an *initial* boundary value problem that results in a differential equation that is often much easier to solve.

Several PE methods have been developed for outdoor sound propagation, with many similarities between them. Departures from the usual atmospheric model include sound propagation over ground whose impedance varies with range,[32] sound propagation over a large ridge,[33] propagation through large-scale turbulence[34] and through randomly inhomogeneous media.[35] In this paper, we compare results from two PE-type models for the four benchmark cases.

2. Basic formulation

Here we present some important steps in the derivation of the PE. Precise details of the derivation may be found in the many references cited above. The boundary condition at the ground surface (assuming flat ground) is the local reaction condition:

$$\left(\frac{\partial p}{\partial z}+ik\beta p\right)_{z=0}=0, \tag{40}$$

where β is the complex surface admittance, which has been normalized by the characteristic impedance (ρc) of the air just above the surface; i.e., $\beta=1/Z$ where Z is the specific acoustic impedance of the ground divided by ρc. Proceeding with Eq. (5), the change of variables $U=pr^{1/2}$ and the far-field assumption ($kr\gg 1$) lead to the well-known Helmholtz wave equation for the field U in two dimensions (r,z):

$$\frac{\partial^2 U}{\partial r^2}+\frac{\partial^2 U}{\partial z^2}+k^2U=0. \tag{41}$$

Let $\tilde{Q}$ denote the operator in the last two terms of Eq. (41); that is

$$\tilde{Q}=\frac{\partial^2}{\partial z^2}+k^2. \tag{42}$$

Assuming k to be independent of range, Eq. (41) can be written

$$\left(\frac{\partial}{\partial r}+i\sqrt{\tilde{Q}}\right)\left(\frac{\partial}{\partial r}-i\sqrt{\tilde{Q}}\right)U=0. \tag{43}$$

The factors represents propagation of incoming and outgoing waves respectively, if we use a time dependence $\exp(-i\omega t)$. Considering only the outgoing wave Eq. (43) reduces to

$$\frac{\partial U}{\partial r}=i\sqrt{\tilde{Q}}U. \tag{44}$$

Most implementations of the PE can be traced back to Eq. (44). The approach for advancing the field in range is the point of departure for the two PE methods we describe. The software implementation of the two approaches are called the FINITE-PE[5,30] and FAST-PE,[31] respectively.

The FINITE-PE method numerically integrates the differential equation using a Crank–Nicolson approach. The operator $\sqrt{\tilde{Q}}$ is cast in the form of a rational Padé expansion[27] for small argument $\tilde{q}$. That is, we write $\sqrt{\tilde{Q}}$ as $k_0\sqrt{1+\tilde{q}}$, where $\tilde{q}=(\tilde{Q}/k_0^2-1)$, and

$$\sqrt{1+\tilde{q}}\approx\frac{1+\frac{3}{4}\tilde{q}}{1+\frac{1}{4}\tilde{q}}. \tag{45}$$

The wave number k_0 is chosen as some constant, mean value of $k(z)$.

The FAST-PE uses a Green's function approach and a split operation that factors propagation effects into an operator for a homogeneous medium and another operator for propagation through the inhomogeneous perturbation. The operator $\sqrt{\tilde{Q}}$ is evaluated using a spectral decomposition, leading naturally to a Green's function implementation which, besides being relatively easy to compute, directly incorporates boundary conditions such as finite ground impedance. Defining $\tilde{Q}_0=\tilde{Q}-\delta k^2$, where $\delta k^2=k^2(z)-k_0^2$ and using an expansion for $\sqrt{\tilde{Q}_0}$ gives

$$\sqrt{\tilde{Q}_0}=\sqrt{\tilde{Q}}+\frac{\delta k^2}{2k_0}. \tag{46}$$

This results in the so-called split step approximation.

3. Notes on implementation

The FINITE-PE uses a spacing between points of $1/k_0$ on its computational grid in height and the same spacing for range step intervals. A very small spacing was chosen near the ground to correctly match the boundary conditions. A Gaussian-shaped amplitude function was used to model the source and its image in the ground, as a starting field for the model. The amplitude of the image Gaussian was modified to satisfy the reflection coefficient at the ground. The wave number was made to be complex-valued to accommodate loss in the media. An absorbing layer approximately 30 wavelengths thick was used to attenuate reflections from the upper boundary, and correctly simulate radiation boundary conditions at infinity. The vertical grid height was fixed to either 500 or 1000 m, to ensure that no unwanted signals would arrive at the receiver, following the 45-deg beamwidth of the wide-angle PE.

In the FAST-PE, all of the range-stepping is accomplished via an FFT algorithm applied to the vertical field at each step. The computational grid must be of equal spacing in height to satisfy constraints placed on the computation by the FFT. Since the boundary condition is modeled explicitly in the algorithm (there are explicit terms for the field reflected from the ground), the grid size is not limited to a minimum spacing that might be needed to model the large variation of the field close to the boundary. Thus, the range step size may be considerably larger than the FINITE-PE

(however, to accurately track rapid deterministic spatial variations in the medium, the FAST-PE generally must use reduced range and height steps, as must the FINITE-PE). An absorbing layer of 30 wavelengths and fixed grid height was also used in the FAST-PE.

4. Execution time (CPU time)

On one computer we used, the finite-PE computational time rose according to the square of frequency and the square of the range of the most distant receiver. This was, of course, to be expected from the equal height and range spacing and the placement of the top boundary. The short-range, low-frequency cases usually ran faster with the finite-PE than with the CERL-FFP model we tested. On the other hand, there was generally not enough computational time available to complete the long-range high-frequency cases, using the finite-PE.

On another computer the FAST-PE and FINITE-PE were exercised on the test cases and the run times were logged.[31] For a constant sound speed (case 1) the FAST-PE calculation was 400, 90, and 80 times faster than the FINITE-PE at 1000, 100, and 10 Hz, respectively. The speed advantage was essentially the same for upward refraction (case 3): 450, 120, and 80 times faster, respectively, at 1000, 100, and 10 Hz. For downward refraction (case 2) and ducted propagation (case 4) the speed advantage was somewhat less. For case 2, at 1000, 100, and 10 Hz the FAST-PE speed advantage was 163, 40, and 86, respectively. For ducted propagation (case 4) and the same three frequencies the speed advantage was, respectively, 70, 44, and 81.

IV. RESULTS

From a practical point of view, downward refraction (case 2) represent propagation conditions for which large detection ranges are possible. Further, agreement between the various models and experimental data is generally good for downward refraction. On the other hand, although upward refraction (case 3) is as commonly occurring, detection ranges are shorter and the agreement of the models with experimental data is not good (for example, see the data in Ref. 30). There is no analytical solution for ducted propagation (case 4) and an isovelocity (homogeneous) medium (case 1) is rarely achieved. For these reasons, we only provide here a complete discussion of all the models for case 2. This is followed by a detailed discussion of all four cases for the FFP and PE.

A. Case 2; 200 m, 100 Hz

Case 2 for a horizontal propagation distance of 200 m and a frequency of 100 Hz has been selected for a detailed comparison of results. All the methods are applicable to this problem and all the wave solution equations perform best for lower frequencies. This case corresponds to propagation at a relatively short distance under downward refraction conditions. In the real world, this situation is expected to be relatively insensitive to atmospheric turbulence. Equation (19) predicts that there should only be one ground reflected ray.

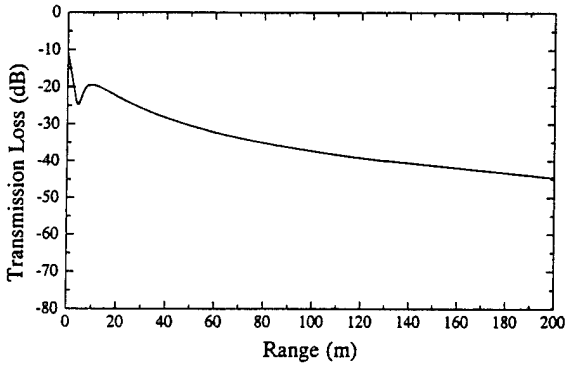


FIG. 4. Transmission loss versus range for case 2, 200 m, 100 Hz, obtained from all the models except SAFARI FFP.

1. Normal mode solution

The discrete normal mode solution [Sec. II A 1, Eq. (10)] for a sound speed gradient represented by Eq. (8) for the impedance of Eq. (2) is displayed in Fig. 4 for 100 Hz and ranges up to 200 m. Note that Eq. (8) represents a linear profile for az small, but becomes infinite at $z=1/2a$ and that the solution does not contain the direct wave. For this reason the discrete normal mode solution will not be accurate near the source. At long ranges, the direct wave decays spherically while the modes decay cylindrically so that the direct wave is negligible.

2. Ray tracing solutions

The same curve as Fig. 4 is also obtained from the single bounce ray tracing model (Sec. II B 2). The single bounce model is a hybrid that uses ray tracing to obtain the total path length of direct and reflected ray and the grazing angle. In general the ray tracing is not restricted to any particular form of atmospheric profile. The resulting grazing angle, when used in Eqs. (12)–(16), modifies the phase of the reflected ray and in turn modifies the transmission loss. The impedance was calculated using Eq. (2).

The same curve as in Fig. 4 is also obtained using the multiple bounce model [Sec. II B 3, Eq. (20)]. The multiple bounce model is also a hybrid model that assumes at the onset a linear sound speed profile and therefore derives the modified path lengths and grazing angle from analytical equations. For case 2, both hybrid models yield the same results since only one reflection occurs [$N=2$ in Eq. (20)].

The Gaussian beam solution [Sec. II B 4, Eq. (31)] yields a curve that is indistinguishable from the one in Fig. 4. This is expected since, in the case of a single reflection, there are no singularities where conventional ray tracing would fail.

3. CERL FFP

The curve in Fig. 4 is also obtained from the CERL FFP solution [Sec. II C 2, Eq. (35) with the ground surface as an impedance plane]. In using the CERL FFP, only the number and distribution of sound speed layers, the extra loss and the number of integration points were varied. Default values for other parameters (Kmax, layer cutoff tolerance, and number of points per FFT) were used. Sufficient layering to represent

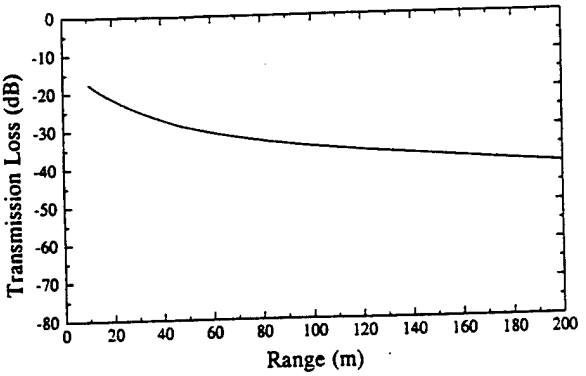


FIG. 5. Transmission loss versus range for case 2, 200 m, 100 Hz, obtained from SAFARI FFP.

the sound speed profile and sufficient number of sampling points to represent the integrand were provided. The number of layers and number of sampling points were continually adjusted where possible to achieve convergence of the TL within 0.5 dB. The run time for convergent solutions was minimized and, given convergence, the extra loss was minimized. Finally, the layer thickness was 2.5 m, with a 100 m isospeed cap.

4. CFFP

The CFFP solution [Sec. II C 2, Eq. (36) with the ground surface as an impedance plane] was implemented using 12 layers and a cap of 12 m. The artificial attenuation used to move poles off the integration path was 1×10^{-8} Np/m. For this case, 64 k points were required. The resulting curve is indistinguishable from the one in Fig. 4.

5. FFLAG FFP

For this implementation [Sec. II C 2, Eq. (35) for a poroelastic ground surface] the atmosphere is modeled by a 12 layer system each about 1 m thick. The air density is calculated by assuming that $\rho=\eta/c^2$ and the constant of proportionality η is calculated by substituting the surface values of ρ_0 and c_0.

The complete parameter set given in Table I is used to model the ground as a thick porous elastic layer over a porous elastic half-space. However, for this parameter set, the predictions are expected to be indistinguishable from those over a rigid porous half-space and the resulting curve is indistinguishable from the one in Fig. 4.

6. SAFARI FFP

This implementation [Sec. II C 2, Eq. (36) for an elastic surface, i.e., porosity is not included] characterizes the ground by compressional and shear wave velocities and compressional and shear wave attenuation from the relevant parameters listed in Table I. The result is the curve in Fig. 5. In the calculation the wave number spectrum is limited to a 4096 samples FFT length. The wave number resolution Δk and spatial resolution Δr are linked by minimum phase velocity: e.g., $\Delta k=2\pi f/c_{\min}M$ (m^{-1}), $\Delta r=c_{\min}/f$ (m) and $R_{\max}=M\Delta r$ (m), where f=frequency, $c_{\min}$=minimum phase velocity, M=number of range intervals and $R_{\max}$=maximum useable range.

As the range increases to 200 m, SAFARI predicts less transmission loss than the normal mode or ray tracing solutions because SAFARI assumes a nonporous elastic soil. In fact for ranges less than 200 m, the transmission loss is close to geometric spreading and the inversion does not affect the loss until larger distances at this frequency.

7. The FINITE PE

The result of the FINITE PE calculation [Sec. II D 1, Eq. (45) with the ground surface as an impedance plane] is indistinguishable from the curve in Fig. 4. The range and height step sizes were both 0.547 m (usually equal to $\frac{1}{4}$ wavelength), the upper edge of the field array was equal to 400 m, and the total number of grid points needed in each field array was 730.

8. FAST-PE

The numerical implementation of the FAST-PE {Sec. III D 2, Eq. (46) with the ground surface impedance plane] on a computer requires the computation of a forward and inverse Fourier transform. For efficiency, a fast Fourier transform (FFT) is used. A numerical Fourier transform requires finite limits on the integration. The lower end of the transform is truncated at the ground while the upper limit is truncated at a height which is defined as z_{top}. The calculated results are also indistinguishable from the curve in Fig. 4 but took less time to compute than the FINITE PE employing a Crank Nicolson method. A complete discussion of the speed advantage of FAST PE is found in Ref. 31 where run times were logged for the FAST-PE and FINITE-PE on the same computer.

9. Summary

In summary, at this short range, all the models (except SAFARI) agree to within the thickness of the line when the same set of parameters are used as input data. The curve obtained from SAFARI, predicts less transmission loss than the other models because of the assumption of an elastic soil with no porosity. Finally, although FFLAG also assumes an elastic soil, the addition assumption of porosity leads FFLAG to agree with models that assume the ground surface as an impedance plane.

B. Case 2; 10 000 m, 100 Hz

1. Normal mode solution

The curve in Fig. 6 is obtained for case 2, $f=100$ Hz up to a range of 10 000 using the normal mode solution, Eq. (10). Beyond about 500 m, the transmission loss shows a complex feature of dips and peaks. These dips and peaks are explained by the normal mode solutions in terms of modal density. We note that there are approximately 667 modes for this case.

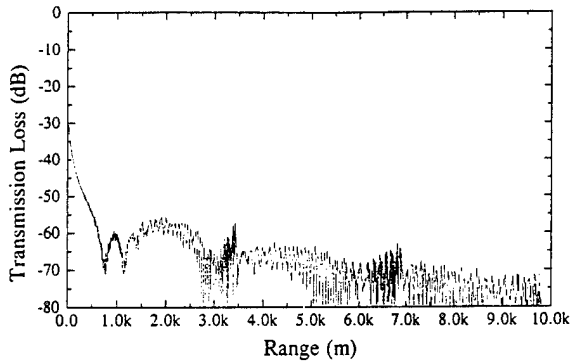


FIG. 6. Transmission loss versus range for case 2, 10 km, 100 Hz, obtained from the normal mode solution.

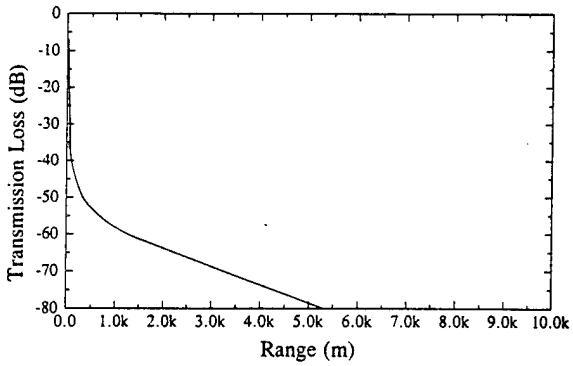


FIG. 8. Transmission loss versus range for case 2, 10 km, 100 Hz, obtained from the single bounce ray model.

2. Ray tracing solutions

At longer ranges, Eq. (19) predicts additional rays that go through multiple reflections on the ground. The curve in Fig. 7 is obtained using the multiple bounce ray model, Eq. (20). The discontinuity that appears at about 500 m corresponds to the sudden appearance of additional rays and is an artifact of the model. However the dip corresponds to physical reality and is explained by a ray model by interference between the various rays. As the range increases, more rays appear in the summation Eq. (20). The additional rays produce the interferences in the curve and contribute to reducing the transmission loss (increasing relative levels). At a range of about 10 km, Eq. (19) predicts up to 100 rays reflected from the ground.

The curve in Fig. 8 is the result of the single bounce ray model described in Sec. II B 2. Since the calculation is restricted to a single bounce, the dips corresponding to the additional rays are absent and at ranges beyond about 1500 m, the transmission loss is greater than the loss predicted when the additional rays are accounted for.

The curve in Fig. 9 was obtained from the Gaussian beam solution, Eq. (31). We note that the beam summation smooths the sharp discontinuity around 500 m that is produced by the multiple bound ray model.

It is interesting to note that on the whole the average TL predicted from the Gaussian beam solution agrees with the TL obtained from the multiple bounce ray model. Not surprisingly, though, the detailed fine structure in the dip and peaks differs due to the nature of the approximations inherent in the ray tracing solutions. The same comments apply to the comparison with the normal mode solution. We note, however, that in a real atmosphere, turbulence would smooth out most of the fine structure predicted by this test case.

3. FFP solutions

The curve calculated from three FFPs that assume a porous ground is shown in Fig. 10. The agreement between the three calculations is within the accuracy of implementation (typically 0.5 dB). The features, predicted as ray interferences according to ray theory and modal density according to the residue series, are reproduced by the numerical solution of the full wave equation.

The curve obtained from the FFLAG FFP is indistinguishable to within numerical implementation from the curves obtained from the CERL FFP and the CFFP. The FFLAG FFP allows for a layered porous and elastic ground using the full range of parameters in Table I. On the other hand the CERL FFP and the CFFP models the ground as a porous half-space. The agreement suggests that a complex impedance plane is a good approximation to the layered ground described in Table I for the benchmark cases. As expected, although not shown here, SAFARI yields a different result from the other FFPs.

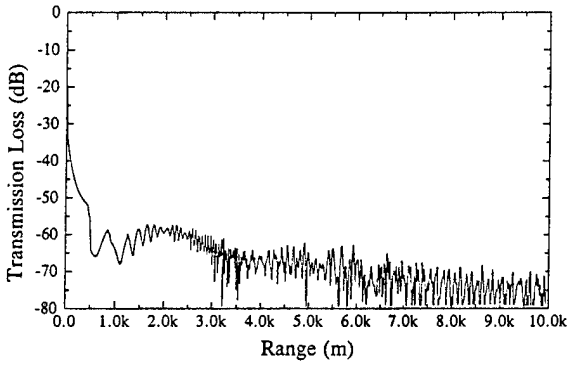


FIG. 7. Transmission loss versus range for case 2, 10 km, 100 Hz, obtained from the multiple bounce ray model.

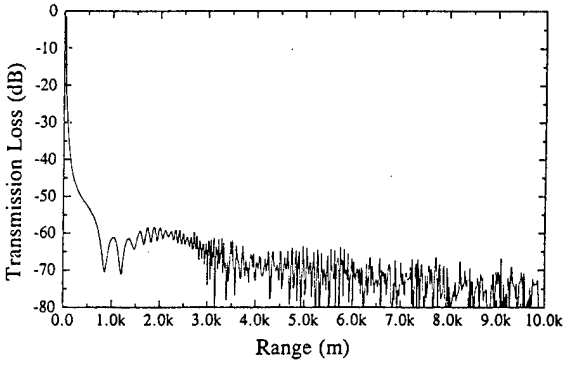


FIG. 9. Transmission loss versus range for case 2, 10 km, 100 Hz, obtained from the Gaussian beam model.

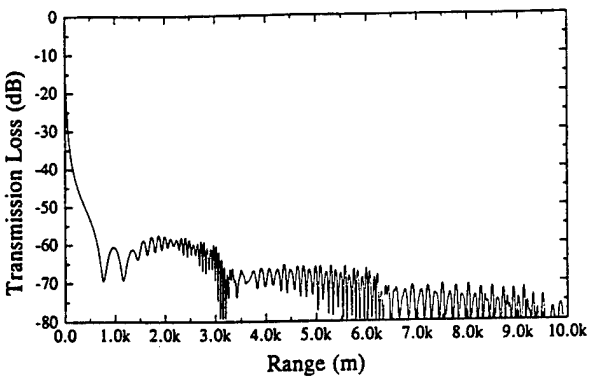


FIG. 10. Transmission loss versus range for case 2, 10 km, 100 Hz, obtained from the CERL FFP, CFFP, FFLAG FFP, and the FAST PE.

4. PE solutions

As noted earlier, computational time is excessive for this long range case using the FINITE PE. The FAST-PE, however, runs in just a few minutes. Details of the fine structure depend upon the height at which the calculation is truncated but the general form is the same as that in Fig. 10.

C. All cases for the FFP and PE

In this section the numerical techniques are used to calculate the transmission loss for the full range of atmospheric conditions described by cases 1 to 4. In the preceding section, results showed that three of the FFP solutions and the

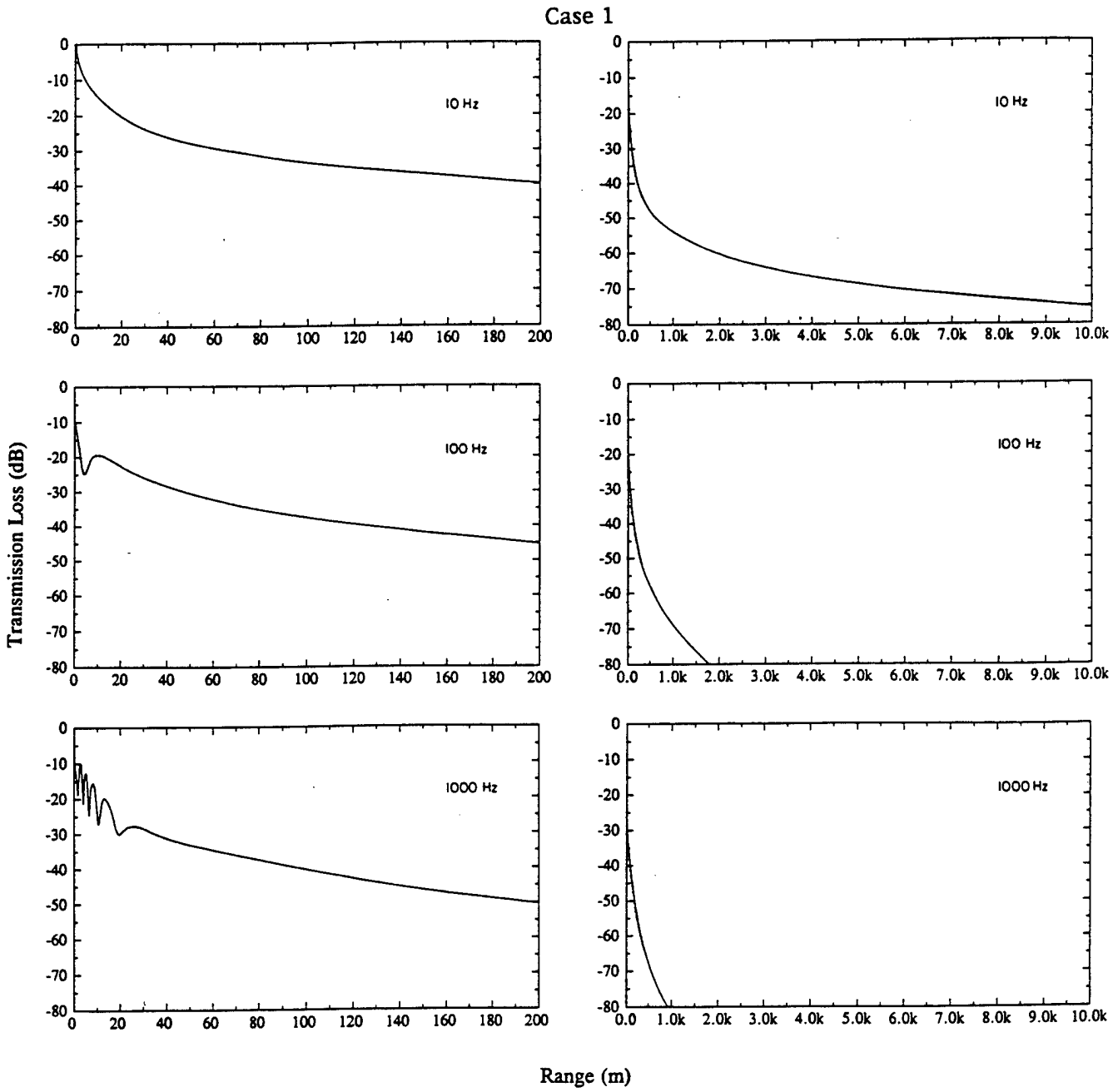


FIG. 11. Transmission loss for case 1 for the FFP and PE. The curves obtained from the two models are indistinguishable.

two PE methods agree with the analytical solution Eq. (10) for case 2, to within the accuracy of implementation (The results from the SAFARI FFP differed because of the assumption of an elastic ground surface). Although not shown here, the FFP and the PE also agree with the analytical solution Eq. (11) for case 3 and the analytical solution Eq. (12) for case 1. Therefore detailed comparison here are restricted to the FFP (except SAFARI) and the PE. In the FFP input parameters were adjusted to obtain an accurate prediction over the entire range of interest for each case, while minimizing run time. The PE used accurately propagates the phases for angles under 45° from horizontal. Consequently, very high angles modes in case 2 at 10 Hz, 10 km have phase errors in the PE solution. We note that case 1 is the most difficult to implement with the FFP's. Case 2 is most difficult for the PE due to the large number of trapped modes. The larger number of modes make the sampling in wave number space less sensitive for the FFP's, since the total solution is the sum of many contributions. Sampling errors of a few modes will not alter the total solution significantly. The FINITE-PE is generally prohibitively slow at longer range or high frequencies, though it needs much less parameter "tuning" than the FFP to obtain convergence.

1. Case 1

Case 1 is a model for sound propagation near the ground in a homogeneous atmosphere and the results are shown in Fig. 11. In case 1, spherical spreading dominates the solution, with a 6-dB enhancement of the field at low frequen-

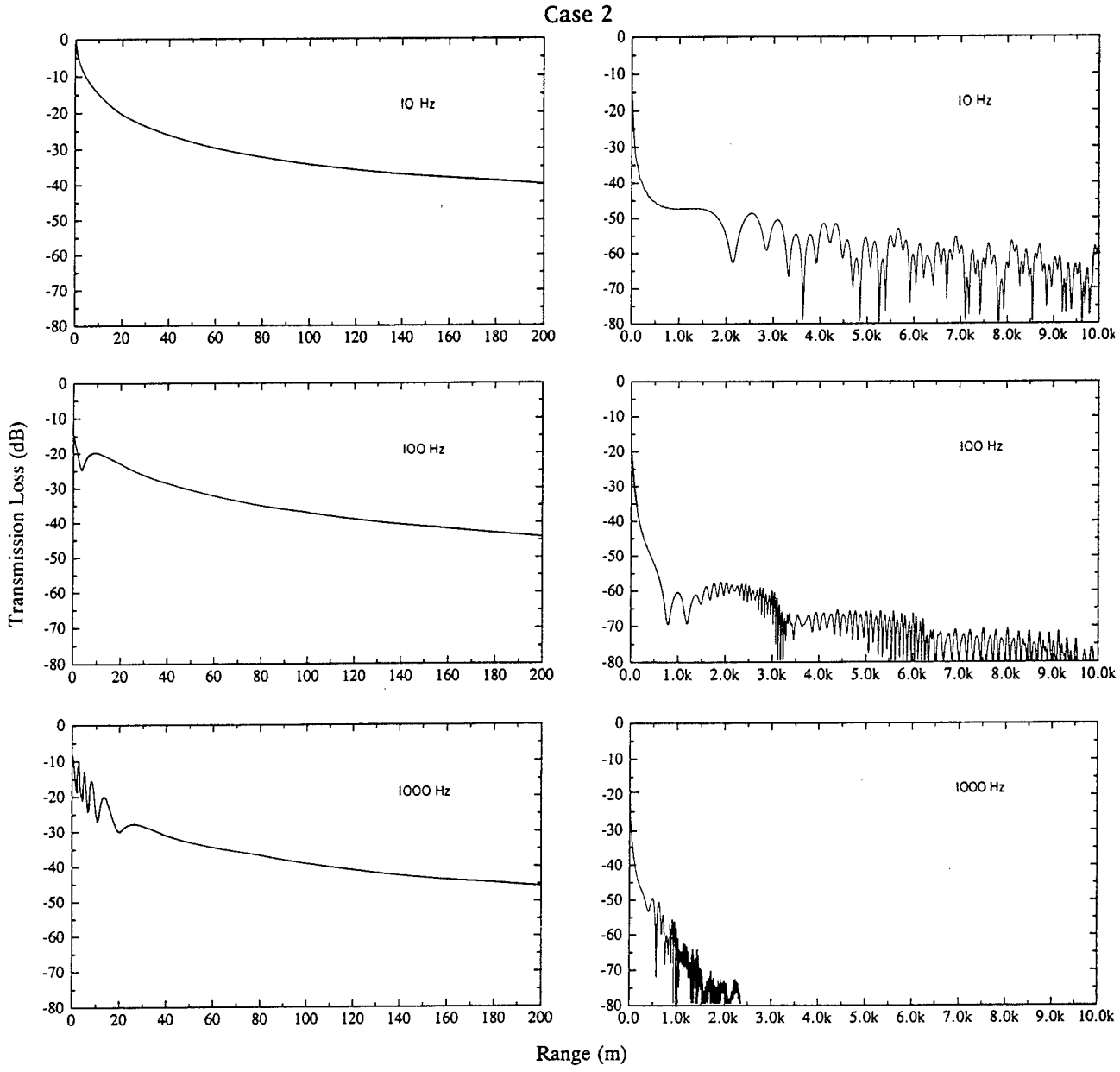


FIG. 12. Transmission loss for case 2 for the FFP and PE. The curves obtained from the two models are indistinguishable.

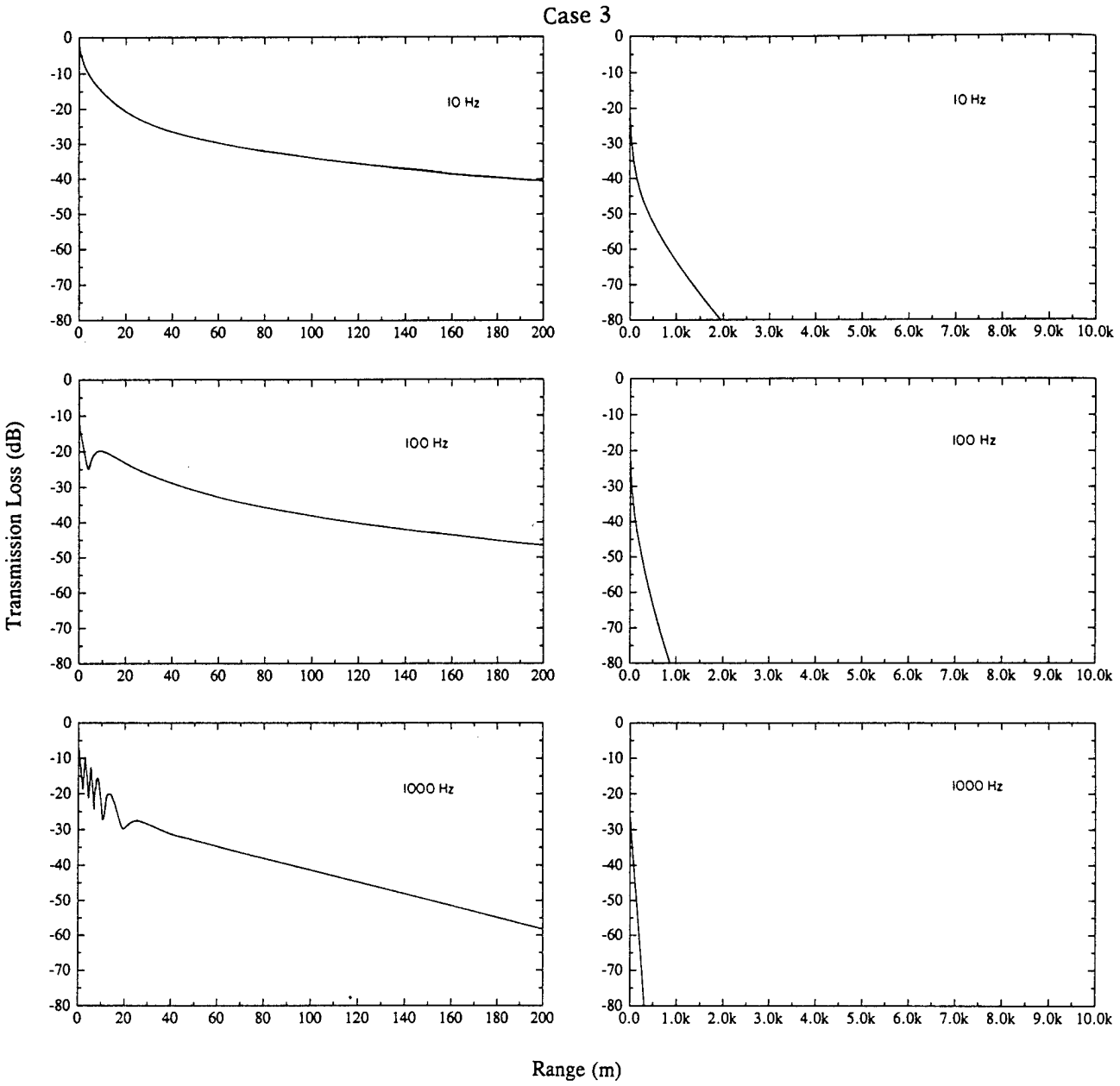


FIG. 13. Transmission loss for case 3 for the FFP and PE. The curves obtained from the two models are indistinguishable.

cies. There is negligible attenuation of the wave at 10 Hz from atmospheric absorption in the atmosphere. There is also negligible ground attenuation from the finite-impedance ground surface at 10 Hz. At 100 Hz, the atmospheric attenuation is 0.18 dB/km. A ground shadow appears at around 850 m and the signal decays at 46 dB per decade of distance (spherical spreading amounts to 20 dB per decade of distance). At 1 kHz, the atmospheric attenuation is 5.4 dB/km. Beyond 200 m a ground shadow occurs and the field decays at 56 dB per decade of distance, including spherical spreading, thus greatly exceeding the decay rate from atmospheric absorption. Although case 1 is the most difficult for the FFP's, when properly implemented, curves obtained from the FFP's and the PE's are indistinguishable. The analytical solution for this case is Eq. (12) and this solution agrees with all the numerical results.

2. Case 2

Case 2 models sound propagation near the ground in a downward refracting situation and was considered in detail in the previous section for all the models. The curves obtained in the case of the FFP and the PE are indistinguishable and a complete set of curves are shown in Fig. 12.

3. Case 3

Case 3 models sound propagation near the ground in an upward refracting environment. At short range (under 200 m)

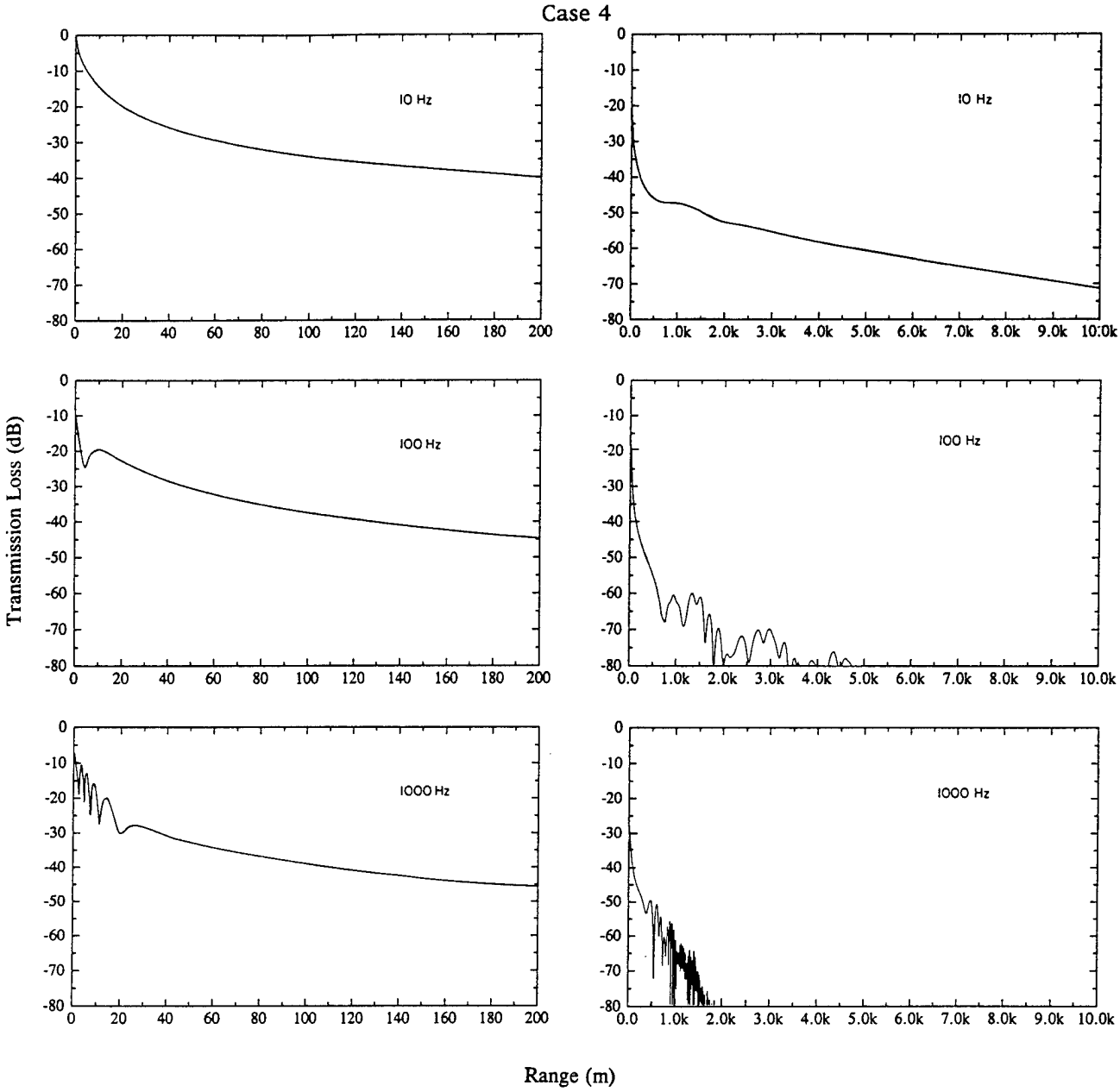


FIG. 14. Transmission loss for case 4 for the FFP and PE. The curves obtained from the two models are indistinguishable.

the predicted curves show very little variation from case 1 except for an increase in TL for 1000 Hz beyond about 130 m. At longer ranges, all curves rapidly decay in the shadow zone formed by the sound speed profile. The results of the PE are indistinguishable from the FFP. A complete set of curves are shown in Fig. 13.

The analytical solution for this case within the shadow zone is the residue series given by Eq. (11) and calculations using this expression, while not shown, agree with the results in Fig. 13. We note, however, that despite agreement between the calculations, the large attenuation predicted at the longer range is not supported by experimental data (for example, see the results shown in Ref. 30).

4. Case 4

Case 4 models sound propagation near the ground in a profile that traps energy near the ground, and refracts sound waves away from the ground at altitudes above 100 m. Above 300 m, the profile is constant. At short range (under 200 m) there is no difference between predictions for this situation and case 2 (downward refraction).

At longer ranges, higher angle modes of propagation interact with the upward refracting portion of the sound speed profile and "leak" energy out of the waveguide formed at the ground. At 100 Hz the curve shows less energy in the signal beyond 1.5 km than in case 2. At 1000 Hz, essentially all of the energy is trapped in the waveguide, and

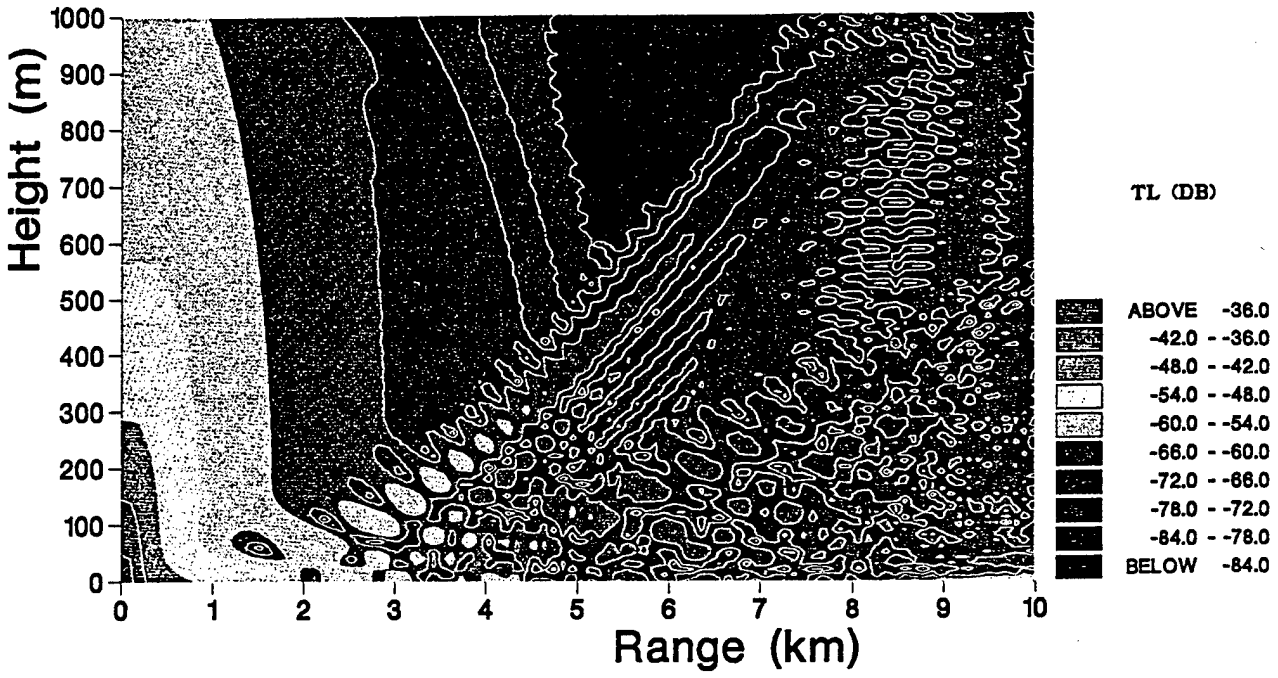


FIG. 15. Contour plot for case 2, 10 Hz.

there is no difference in the results from case 2. The complete set of curves are shown in Fig. 14. There is no analytical solution for this case.

D. Contour plot

For illustrative purposes a contour plot of transmission loss is shown in Fig. 15 for case 2, $R=10$ km, $f=10$ Hz which can be compared directly with the ray diagram in Fig. 16.

V. SUMMARY

Benchmarks were calculated using analytical models, ray tracing models, the FFP solution and the PE solution. The analytical models included two residue series for cases 2 and 3, respectively, and the result of a steepest descent integration for case 1. The ray tracing included a model restricted to a single bounce, a multiple bounce model and a Gaussian beam summation. Four different implementations of the FFP were considered. In one FFP (SAFARI), the ground is assumed to be a nonporous elastic layered interface. Two FFP's (CERL-FFP, CFFP) considers the ground to be a rigid semi-infinite porous surface while the fourth FFP (FFLAGS) assumed a porous elastic layer above a nonporous elastic substrate. The PEs considered included a Crank–Nicolson version (FINITE PE) and a fast Green's function method (FAST PE).

For all cases, results show that the FFP and the PE agree with the analytical solution, where an analytical solution is available, to within the accuracy of implementation. Any differences between the various implementations of the FFP and the PE can readily be attributed to different assumptions in the models. For example SAFARI predicts less transmission loss than the other versions because it assumes a nonporous ground.

The results from the ray tracing models depend on how many physical mechanisms are included and to what level of detail they are considered. For example a ray model that assumes a single bounce at all distances predicts larger transmission loss at long ranges for a downward refracting atmosphere than the other models. In the absence of upward refraction, for the cases considered, the ray trace algorithms give reasonable agreement with the more accurate techniques (FFP and PE).

None of the models considered here include the effects of turbulence, rough ground, or terrain features. These effects are all present to some degree in experimental data. For this reason, the results of the calculations presented here may not compare well to experimental data especially in refractive

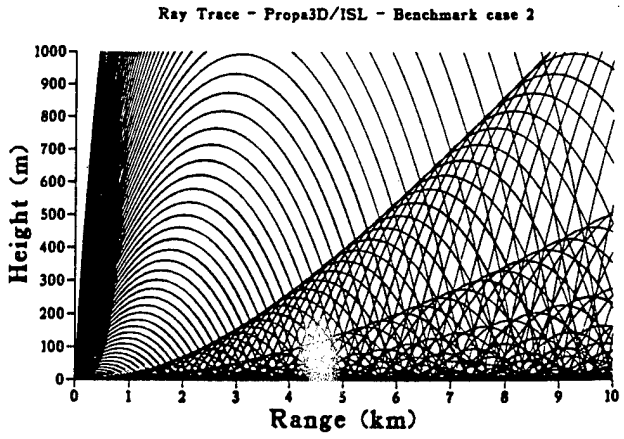


FIG. 16. Ray diagram for case 2.

shadow zones. The benchmark calculations presented here should be used to test theory and algorithms in the absence of these complications which are always present in the real world. Then the user can proceed to include other effects with a confidence that refraction, diffraction, and ground reflection have been properly accounted for.

ACKNOWLEDGMENTS

The authors are rigorously listed in alphabetical order with authors from the same institution grouped together. The editing of the manuscript and preparation of the figures was coordinated by the National Research Council of Canada. NATO DRG Panel 3 is acknowledged for sponsoring and providing the frame work for this investigation. Atlas Elektronik GmbH acknowledges MoD-Bonn FRG, RÜ T III/4, Mr. H. Wolff. Work at the National Research Council was supported in part by DND/DREV. Work at the Open University, England, was supported in part by the U.S. Army (WES) through its European Research Office.

[1] J. E. Piercy, T. F. W. Embleton, and L. C. Sutherland, "Review of noise propagation in the atmosphere," J. Acoust. Soc. Am. **61**, 1403–1418 (1977).

[2] R. Raspet, G. E. Baird, and W. Wu, "The relationship between upward refraction above a complex impedance plane and the spherical wave evaluation for a homogeneous atmosphere," J. Acoust. Soc. Am. **89**, 107–114 (1991).

[3] R. Raspet, G. Baird, and W. Wu, "Normal mode solution for low frequency sound propagation in a downward refracting atmosphere above a complex impedance plane," J. Acoust. Soc. Am. **91**, 1341–1352 (1992).

[4] S. W. Lee, N. Bong, W. F. Richards, and R. Raspet, "Impedance formulation of the fast field program for acoustic wave propagation in the atmosphere," J. Acoust. Soc. Am. **79**, 628–634 (1986).

[5] K. E. Gilbert and M. J. White, "Application of the parabolic equation to sound propagation in a refracting atmosphere," J. Acoust. Soc. Am. **85**, 630–637 (1989).

[6] L. B. Felsen, "Benchmarks: An option for quality assessment," J. Acoust. Soc. Am. **87**, 1497–1498 (1990).

[7] H. A. J. M. van Hoof and K. W. F. M. Doorman, "Coupling of airborne sound in a sandy soil," Laboratory for Electronic Developments TNO (1983). The Wezep soil is also measurement No. 12 in M. J. M. Martens, L. A. M. vander Heijden, H. H. J. Walthous, and W. J. J. M. van Rems, "Classification of soils based on acoustic impedance air flow resistivity and other physical soil parameters," J. Acoust. Soc. Am. **78**, 970–980 (1985).

[8] The expressions for k_b^2 and $\rho_b(\omega)$ are found in K. Attenborough, "Acoustical impedance models for outdoor ground surfaces," J. Sound Vib. **99**, 501–544 (1985), however, λ is written here as a function of the pore shape factor ratio s_p defined by Eq. (21) in K. Attenborough, "On the acoustic slow wave in air-filled granular materials," J. Acoust. Soc. Am. **81**, 93–102 (1987). A good appreciation of the role of the flow resistivity, tortuosity, and porosity of ground surface can be found in K. Attenborough, "Ground parameter information for propagation modeling," J. Acoust. Soc. Am. **92**, 418–427 (1992) and H. M. Hess, K. Attenborough, and N. W. Heap, "Ground characterization by short-range propagation measurements," J. Acoust. Soc. Am. **87**, 1975–1986 (1990).

[9] ANSI SI.26-1978, "Method for the calculation of the absorption of sound by the atmosphere" (Acoustical Society of America, New York, 1978).

[10] A. Berry and G. A. Daigle, "Controlled experiments on the diffraction of sound by a curved surface," J. Acoust. Soc. Am. **83**, 2047–2058 (1988).

[11] K. Attenborough, S. I. Hayek, and J. M. Lawther, "Propagation of sound above a porous half-space," J. Acoust. Soc. Am. **68**, 1493–1501 (1980). Typographical corrections can be found in the Appendix in K. B. Rasmussen, "Sound propagation over grass covered ground," J. Sound Vib. **78**, 247–255 (1981).

[12] Anon., "Advanced Sound Propagation in the Atmosphere," University of Mississippi (1991). ASOPRAT includes the CERL FFP as a default in shadow zones. Only the ray trace part is described here.

[13] A. L'Espérance, P. Herzog, G. A. Daigle, and J. Nicolas, "Heuristic model for outdoor sound propagation based on an extension of the geometrical ray theory in the case of a linear sound speed profile," Appl. Acoust. **37**, 111–139 (1992).

[14] V. Cerveny, M. M. Popov, and I. Psencik, "Computation of wave fields in inhomogeneous media-Gaussian beam approach," Geophys. J. R. Astron. Soc. **70**, 109–128 (1982).

[15] Y. Gabillet, H. Schroeder, G. A. Daigle, and A. L'Espérance, "Application of the Gaussian beam approach to sound propagation in the atmosphere: Theory and experiments," J. Acoust. Soc. Am. **93**, 3105–3116 (1993).

[16] F. R. DiNapoli, "A fast field program for multilayered media," Naval Underwater Systems Center, Tech. Rep. 4103 (1971).

[17] H. W. Kutschale, "The integral solution of the sound field in a multilayered liquid–solid half-space with numerical computations for low-frequency propagation in the Arctic Ocean," Tech. Rep. No. 1 (CU-1-70, ONR Contract N00014-67-A-0108-0016), Lamont–Doherty Geological Observatory, Columbia University, Palisades, New York (1970).

[18] H. W. Kutschale, "Rapid computation by wave theory of the propagation loss in the Arctic Ocean," Tech. Rep. No. 8, Lamont–Doherty Geological Observatory, Columbia University, Palisades, New York (1981).

[19] R. Raspet, S. W. Lee, E. Kuester, D. C. Chang, W. F. Richards, R. Gilbert, and N. Bong, "Fast-field program for a layered medium bounded by complex impedance surfaces," J. Acoust. Soc. Am. **77**, 345–352 (1985).

[20] Y. L. Li, S. J. Franke, and C. H. Liu, "Numerical implementation of an adaptive fast-field program for sound propagation in layered media using the chirp z transform," J. Acoust. Soc. Am. **89**, 2068–2075 (1991).

[21] A. Güdesen, "Application of the SAFARI model to sound propagation in the atmosphere," J. Acoust. Soc. Am. **87**, 1968–1974 (1990). For a detailed discussion of the use of SAFARI in underwater acoustics, see H. Schmidt, "SAFARI User's Guide," SACLANTCEN Report SR-114, NATO (1988) and C. H. Harrison, "Ocean Propagation Models," Appl. Acoust. **27**, 163–201 (1989).

[22] T. L. Richards and K. Attenborough, "Accurate FFT-based Hankel transforms for prediction of outdoor sound propagation," J. Sound Vib. **109**, 157–167 (1986).

[23] S. Tooms, S. Taherzadeh, and K. Attenborough, "Sound propagation in a refracting fluid above a layered fluid-saturated porous elastic material," J. Acoust. Soc. Am. **93**, 173–181 (1993).

[24] S. J. Franke and G. W. Swenson, Jr., "A brief tutorial on the Fast Field Program applied to Sound Propagation in the Air," Appl. Acoust. **27**(3), 203–216 (1989).

[25] M. West, R. A. Sack, and F. Walkden, "The fast field program. A second tutorial: Application to long range sound propagation in the atmosphere," Appl. Acoust. **33**, 199–228 (1991).

[26] G. D. Dockery, "Modeling electromagnetic wave propagation in the troposphere using the parabolic equation," IEEE Trans. Antennas Propag. **36**(10), 1464–1470 (October, 1988).

[27] J. F. Claerbout, *Fundamentals of Geophysical Data Processing* (McGraw-Hill, New York, 1976), p. 206.

[28] F. D. Tappert, "The parabolic approximation method," in *Wave Propagation and Underwater Acoustics*, edited by J. B. Keller and J. S. Papadakis (Springer-Verlag, New York, 1977).

[29] M. K. Myers and G. L. McAninch, "Parabolic approximation for sound propagation in the atmosphere," AIAA J. **16**, 836–842 (August 1978).

[30] M. J. White and K. E. Gilbert, "Application of the parabolic equation to the outdoor propagation of sound," Appl. Acoust. **27**(3), 227–238 (1989).

[31] K. E. Gilbert and X. Di, "A Fast Green's function method for one-way sound propagation in the atmosphere," J. Acoust. Soc. Am. **93**, 714–720 (1992). The software package that implements The Green's Function PE (GF-PE) is referred to as "FAST-PE" in this article.

[32] J. C. Craddock and M. J. White, "Propagation of acoustic waves past an impedance discontinuity: a parabolic equation approach," J. Acoust. Soc. Am. **91**, 3184–3191 (1992).

[33] J. S. Robertson, M. J. Jacobson, W. L. Siegmann, and D. P. Santandrea, "Acoustical effects of a large ridge on low-frequency sound propagation in stationary and moving atmospheres," RPI Math. Rep. No. 181, Dept. of Mathematical Sciences, Rensselear Polytechnic Institute, Troy, NY 12180-3590 (November 1989).

[34] J. M. Noble, H. E. Bass, and R. Raspet, "Effects of large-scale wind driven turbulence on sound propagation," J. Acoust. Soc. Am. **87**, S149 (1990).

[35] K. E. Gilbert, R. Raspet, and X. Di, "Calculation of turbulence effects in an upward refracting atmosphere," J. Acoust. Soc. Am. **87**, 2428–2437 (1990).

Editor's Note: *This is the fiftieth in a series of review and tutorial papers on various aspects of acoustics.*

Review of noise propagation in the atmosphere

J. E. Piercy and T. F. W. Embleton

Division of Physics, National Research Council, Ottawa, Ontario, Canada, K1A OS1.

L. C. Sutherland

Wyle Laboratories, El Segundo, California 90245

(Received 4 January 1977; revised 11 March 1977)

A general review is presented of most areas of sound-propagation outdoors that are of interest for the control of community noise. These areas are geometrical spreading, atmospheric absorption, ground effect, (near horizontal propagation in a homogenous atmosphere close to flat ground), refraction, the effect of atmospheric turbulence, and the effect of topography (elevation, hillsides, foilage, etc.) The current state of knowledge in each area is presented and suggestions made concerning research activities, applications of existing research, and practical problems which arise in the prediction of noise levels.

PACS numbers: 43.10.Ln, 43.28.Bj, 43.28.Fp, 43.50.Vt, 43.50.Sr

INTRODUCTION

The study of sound propagation in the atmosphere has a long and intriguing history. Between 1850 and 1950 there were a number of isolated investigations of good quality in response to specific needs of the times, such as fog signaling, the location of artillery pieces, etc. (see Ref. 1 for an annotated bibliography of work done prior to 1965). When the noise from jet aircraft and the testing of rockets became a social problem in the 1950's, there was a resurgence of research activity, adapting existing knowledge from other fields of physics to the problem of noise propagation, and several excellent reviews of sound propagation in the atmosphere appeared at this time.[2–5] The field then returned to its previous state of sporadic isolated activity, although at a substantially higher level. In recent years increased concern over noise has led once again to advances in the understanding of outdoor propagation by the application of knowledge from other fields of physics, which this time has been gained since the 1950's.

This review is an attempt to document the advances for most areas of sound propagation in the atmosphere which are relevant to the control of community noise. It was undertaken at the request of the Coordinating Committee on Environmental Acoustics of the Acoustical Society of America, and is part of the output of the working group on outdoor propagation (see also Refs. 6 and 7). The review aims to summarize the current state of knowledge in each area, and to suggest research activities, applications of existing research, and practical problems which arise in the prediction of noise levels. Because the field is so diffuse, each subject area is treated separately, accompanied by most of the suggestions relevant to the particular area. To add perspective, however, problems in predicting noise levels are discussed generally at the end (Sec. V) with recommendations for research.

The first two areas covered in the review are geometrical spreading (Sec. I A) and atmospheric absorption (Sec. I B). Together, these are the dominant mechanisms determining sound levels in air-to-ground sound propagation. Near-horizontal propagation in a homogeneous atmosphere close to flat ground (the ground effect) is then treated in Sec. II and the effect of surface meteorology in Sec. III. The latter includes the effects of refraction and atmospheric turbulence. The role of topography—elevation, hillsides, foliage, obstructions, etc.—is then discussed briefly in Sec. IV.

There is much useful material in the earlier reviews of a general nature[2–5] which has been excluded here for brevity. Furthermore, specifically excluded are multipath propagation in cities, and noise reduction by barriers, which have been the subjects of recent reviews.[6,7]

I. PROPAGATION AWAY FROM BOUNDARIES (AIR TO GROUND)

A. Geometrical spreading

The geometrical spreading of sound from a coherent source is normally well covered in textbooks—an attenuation of 6 dB per doubling of distance for spherical expansion from a point source, 3 dB per doubling of distance for cylindrical expansion from an infinite line source, and parallel loss-free propagation from an infinite area source. For sources of finite size there is a nearfield where the above is approximately true and a farfield where the expansion is spherical. In community noise, however, incoherent sources are often more important and the treatment of geometrical spreading from incoherent sources has been extended in recent years to cope with multivehicle problems, particularly road traffic and railway noise.

The description above of the spreading from coherent sources remains true for sources which are incoherent, but the size of the nearfield is much more restricted, and the propagation much less directional (compare as an illustration, the light from a light bulb to that from a laser). Treatments of the incoherent acoustic line source which are useful for highway or railway design are given in Refs. 8 and 9.

The background noise in a city may be modelled as an incoherent area source.[10–12] The effect of city boundaries—streets, barriers, and open areas—in channeling or attenuating this propagation is reviewed elsewhere.[6,7,10] A point of relevance to noise control is that the geometric spreading of different percentiles of

Copyright © 1977 by the Acoustical Society of America

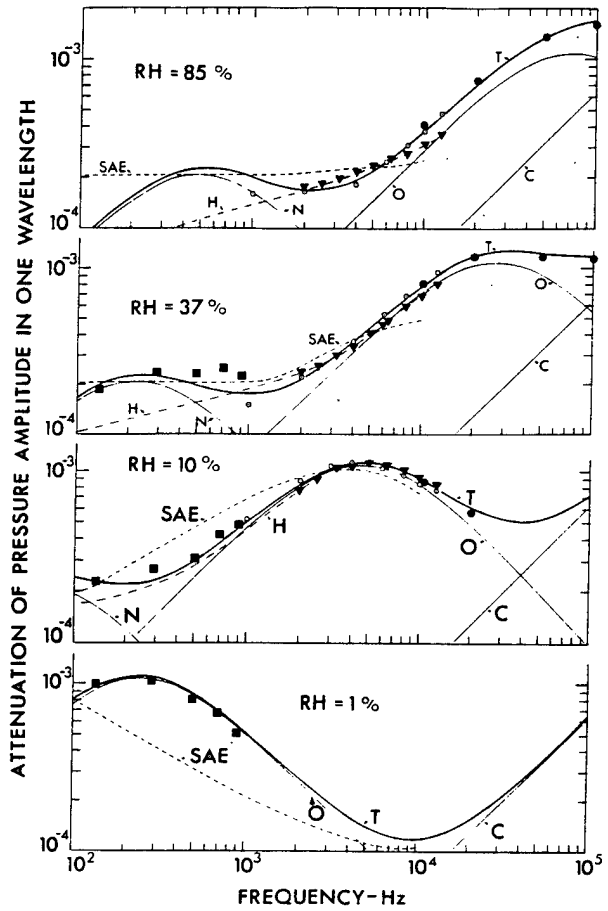


FIG. 1. The absorption of sound in a distance of one wavelength for four values of relative humidity RH, a temperature of 20 °C, and a pressure of 1 atm (Ref. 14). Line C gives the classical and rotational absorption, curves O and N the contributions from the vibrational relaxations of oxygen and nitrogen, respectively, and curve T is the sum of these three contributions. Experimental points ● are from Ref. 15, ○ from Ref. 16, ▼ from Ref. 17, and ■ from Ref. 18. The absorption predicted by Ref. 22 is curve SAE, by Ref. 17 is curve H, and by Ref. 23 is curve T.

noise from an (isolated) incoherent line source varies with the percentile.[13] There is a considerable conceptual and economic benefit in specifying noise criteria for the design of roadways in such a way that the spreading loss can be accounted for in a simple manner, such as a loss of 6 dB/doubling of distance for the maximum level of individual vehicles (i.e., levels close to L_1, the level exceeded only 1% of the time), or a loss of 3 dB/doubling of distance for the equivalent continuous sound level (L_{eq}).

B. Atmospheric absorption

Present knowledge of the rate at which acoustic energy is absorbed during propagation through the atmosphere comes from three sources—direct measurements in the field, measurements of air absorption in the laboratory, and general knowledge of the mechanisms. The current state of the latter two may be examined using Fig. 1. In this figure,[14] the frequency dependence of the absorption in a distance of one wavelength is shown for four different values of relative humidity (RH). The measurements are from four different investigations,[15–18] and the first point to note is the close similarity between the measurements by the different investigators. The second is the agreement for each value of relative humidity between the measurements and the top curve T, which is the theoretical curve representing the sum of the contributions from the different mechanisms of absorption (identified by the curves marked C, O, and N).

Line C gives the classical absorption caused by the transport processes of classical physics (shear viscosity, thermal conductivity, mass diffusion, and thermal diffusion), together with the absorption caused by rotational relaxation of the molecules in air. Curves O and N represent the contributions from the vibrational relaxation of oxygen and nitrogen molecules, respectively. The fit of curve T to the measurements shown in the figure has been achieved by the adjustment of two constants in the theory, but the basic theory[19] is now sufficiently firm to achieve a fit with an accuracy of approximately 25% without using any adjustment.

It is useful to compare the different methods available for predicting atmospheric absorption with the measurements shown in Fig. 1. Kneser first developed a theory based only on contributions C and O. While predictions with his original method[20] are substantially different from curves C and O shown in the figure, a later version by Evans and Bazley[16] gives positions for these curves which are close to those in the figure. Evans and Bazley's method provides, therefore, a fairly good approximation to experimental data for higher frequencies. At the higher humidities represented by the top two sets of measurements in Fig. 1, however, the vibrational relaxation of nitrogen[21] (curve N) is the principal mechanism of absorption for the range of frequencies below about 1–2 kHz which are of major importance for community noise: both the Kneser, and Evans and Bazley methods therefore grossly underpredict the absorption in this low-frequency range.

Two similar methods of calculating atmospheric absorption were proposed before the role of the nitrogen relaxation was understood, that of Committee A21 of the Society of Automotive Engineers[22] (1964) and Harris[17] (1966), both of which produced a decided improvement in the low-frequency range, and these two essentially empirical methods have seen widespread use. Curves marked H and SAE in Fig. 1 show the predictions of the method of Harris, and the SAE committee, respectively. Major difficulties using these empirical methods for predicting absorption appear mainly for abnormal measurement conditions, due to extrapolation with an incorrect dependence on the variables, as shown in the figure, for example, for low frequencies or dry air.

It is now possible to generate curve T in Fig. 1 using simple algorithms with a firm theoretical base, which can be handled easily, for example, by a programmable hand calculator. A new (draft) standard method for calculating atmospheric absorption has recently been proposed on such a basis.[23] The predictions of this new method have been compared with a large assembly of both laboratory and field data from the literature.[24]

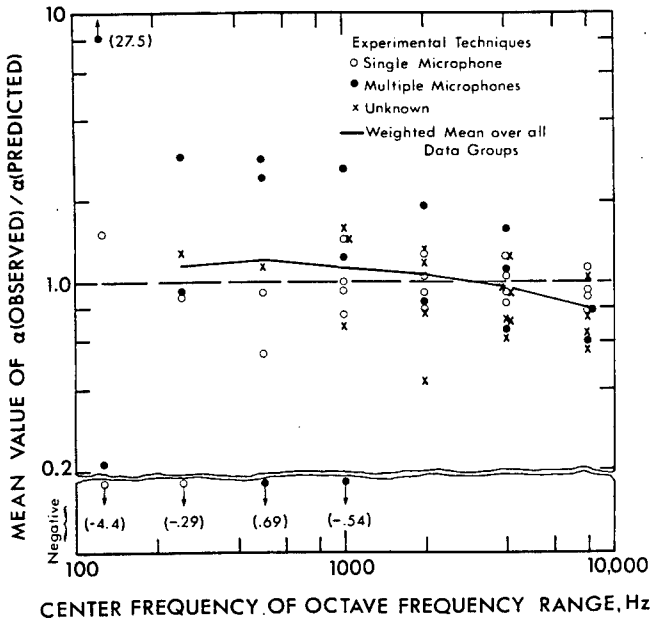


FIG. 2. Ratio of air-to-ground (aircraft sound propagation) measurements of air absorption loss coefficients to predicted values as a function of frequency (Ref. 24). Predicted values were based on weather conditions on the ground, or when data were available on average weather conditions along the propagation path.

The comparison of over 850 laboratory measurements with predictions shows that near 20 °C the predicted values agree within about 5%, with the average of the measurements throughout the audio frequency range and over a wide range of humidity. Although laboratory data at other temperatures are limited, the prediction model is estimated to be reliable within ± 10% from 0° to 40 °C. The field data included analyses of more than 750 measurements of aircraft flyover noise over wide ranges of frequency, temperature, and humidity from a number of different investigators. Although the scatter in the field results was much larger than that in the laboratory measurements, because the weather cannot be controlled, the agreement between predicted and measured absorption was good on the average, as shown in Fig. 2.

The absorption predicted by the new method, in the practical units of dB/100 m for a pressure of 1 atm, a temperature of 20 °C, and a relative humidity of 70%, is shown in Fig. 3. Note that the attenuation by absorption is constant for a given difference in propagation path lengths unlike geometrical spreading, where it is constant for a given ratio of propagation path lengths. Thus atmospheric absorption tends to become more important with increasing distance between the source and receiver. Note also from Fig. 3 that the total absorption increases sharply with increasing frequency.

With the new proposed method of predicting atmospheric absorption, the lag between current practice and knowledge has been largely eliminated. Recently, also, this method has been confirmed by extensive measurements at high frequencies[25,26] (4–100 kHz), and a few measurements at, in effect, very low frequencies[27] (down to 4 Hz). Nevertheless, there is still need for more data at low frequencies over a substantial range of temperature, and at low humidities. Continuing support is also needed for fundamental work on the mechanisms of air absorption. The measurements would provide needed understanding of the temperature dependence of the relaxation frequency of nitrogen, and lead in due course to the calculation of atmospheric absorption directly from fundamental knowledge[19] with enhanced accuracy.

It should be noted that many of the best measurements of outdoor propagation have appeared in the literature in corrected form, with the attenuation calculated for atmospheric absorption removed. The correction for absorption should now be recalculated using the new method, before the original interpretation of these measurements is accepted. In particular the failure to recognize earlier the contribution from the nitrogen relaxation has led to a number of difficulties some of which will appear in later sections.

II. PROPAGATION NEAR THE GROUND

The theory of sound propagation near or along the ground has been treated analytically in the literature, with at least three levels of complexity for the ground surface.

(a) The ground is treated as a locally reacting surface, and waves within the ground are not considered.

(b) The ground is treated as an isotropic fluid medium capable of transmitting dilatational waves in any di-

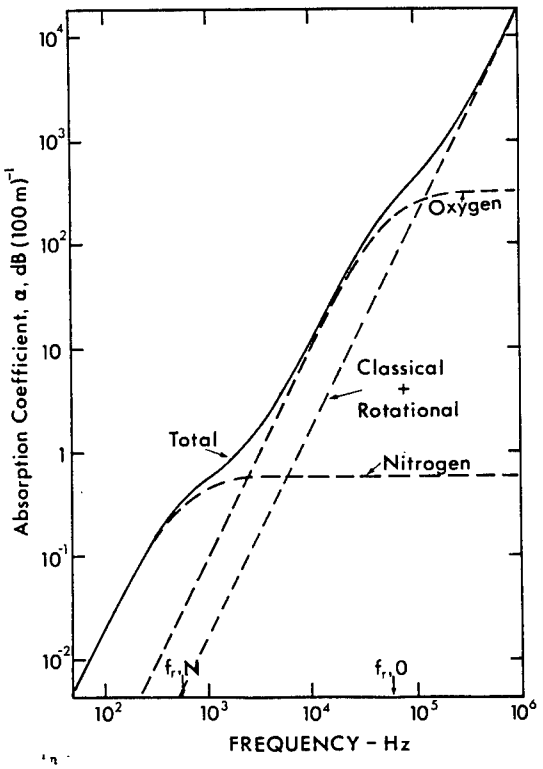


FIG. 3. Predicted atmospheric absorption in dB/100 m for a pressure of 1 atm, temperature of 20 °C and relative humidity of 70% (Ref. 23).

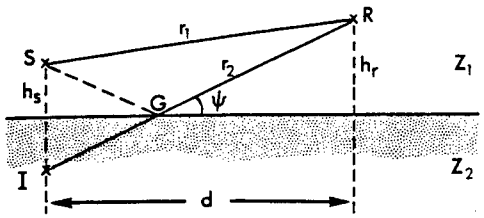


FIG. 4. Reflection of sound from flat ground with impedance Z_2.

rection as the result of an impinging wave on the surface.

(c) The ground is treated as an elastic solid medium capable of transmitting both dilatational and shear waves resulting from an impinging wave on the surface.

While theoretical models are available in the literature for the latter two[28,29] and other more sophisticated models,[30] there is very little evidence available as yet to show that propagation of sound over the ground is not adequately predicted by the use of the first, much simpler, "locally reacting" model for the ground surface. Therefore, this is the only model which will be treated in detail in this review. One exception to this general applicability of a "locally reacting" model will be covered only briefly.

A. Plane waves

The amplitude reflection coefficient R_p for a plane wave of sound incident obliquely on a plane locally reacting surface may be conveniently represented by the formula[31]

$$R_p = \frac{\sin\psi - Z_1/Z_2}{\sin\psi + Z_1/Z_2}, \tag{1}$$

where ψ is the grazing angle (Fig. 4), $Z_1 = \rho c$ is the characteristic impedance of air, and Z_2 is the acoustic impedance of the surface (i.e., the ratio of pressure to the normal component of velocity at a point on the surface). In order to include the change of phase as well as amplitude on reflection, complex notation is used for both R_p and Z_2. The characteristics of a particular locally reacting surface may then be represented completely by a complex impedance, $Z_2 = R_2 + jX_2$, which may be dependent on the frequency but not on the grazing angle.

For a perfectly reflecting hard surface, the phase change on reflection is zero, R_p is 1, and Z_2 must be infinite. In practice, however, Z_2 may be very large, but must always remain finite. For normal incidence in this case $\sin\psi = 1$, and R_p is effectively 1. It is always possible, however, to choose ψ sufficiently small to make the term $\sin\psi$ in Eq. (1) small compared to the fixed parameter Z_1/Z_2, and hence make R_p effectively -1. This value signifies a phase change of 180° on reflection and a cancellation of incident and reflected waves at grazing incidence, even though their path lengths are equal. Plane waves at grazing incidence over a plane surface with a finite acoustic impedance [illegible] forbidden mode of propagation.

The rest of Sec. II is a description of the implications of this statement for the propagation of noise over the ground, where values of the grazing angle less than 5° have primary importance. The cancellation for small ψ will be shown to represent, in practice, a shadow zone whose acoustical depth is related to the value of the ground impedance.

B. Spherical waves, direct and reflected

The propagation of spherical waves from a point source near a reflecting plane is an intricate and rambling subject, both mathematically and conceptually. In electromagnetics it has a long history, well chronicled by Baños,[32] from its start by Sommerfeld[33] in 1909 to its present state, as reviewed by Wait.[34] In acoustics there is also a substantial history, from Rudnick[35] in 1947 to Wenzel,[36] and Donato,[37] which is chiefly one of adoption of the ideas and solutions from electromagnetic theory. The portion of this history most relevant to the propagation of noise over the ground has recently been surveyed by Embleton, Piercy, and Olson.[38] The treatment here will follow the latter study, to which the reader is referred for details.

1. Shadow zone—source or receiver on the ground

Consider, again, the basic configuration shown in Fig. 4. There is a point source S and receiver R situated above a plane boundary with acoustic impedance Z_2. The pressure amplitude p at point R may be represented as follows[39,40]:

$$p/p_0 = (1/r_1)\exp(-ikr_1) + (R_p/r_2)\exp(-ikr_2) + (1-R_p)(F/r_2)\exp(-ikr_2)\ . \tag{2}$$

Parameter p_0 is the amplitude of the pressure at unit distance from point source S in the absence of the ground surface. Equation (2) is one form of what is often referred to as the Weyl–Van der Pol solution.

The first term on the right of Eq. (2) expresses, in complex notation, the contribution from the wave proceeding directly from S to R. It is only necessary to recognize that the distance r_1 in the denominator is an expression of the inverse square law. The second term is the familiar one for the reflected wave, which appears at point R to have come a distance r_2 from the image source I. The first two terms together give the behavior shown in Fig. 5, which is essentially as described in the previous section for plane waves, since R_p is the plane wave reflection coefficient given by Eq. (1). [The significance of the third term in Eq. (2), and in particular the amplitude factor F, which gives the behavior of ground and surface waves, will be described later in Sec. IID.] Propagation upwards in Fig. 5, away from the point source on the surface [large ψ in Eq. (1)], is spherical expansion: compare, for example, the 30-dB contour to the dashed circle in the figure. Along the surface, however, a shadow zone forms gradually as the sound propagates outward. In this region ($\sin\psi \ll Z_1/Z_2$), Eq. (1) and the first two terms in Eq. (2) indicate an attenuation A_e in excess of that from inverse square law and atmospheric absorption (and called hereafter just excess attenuation) given by[39]

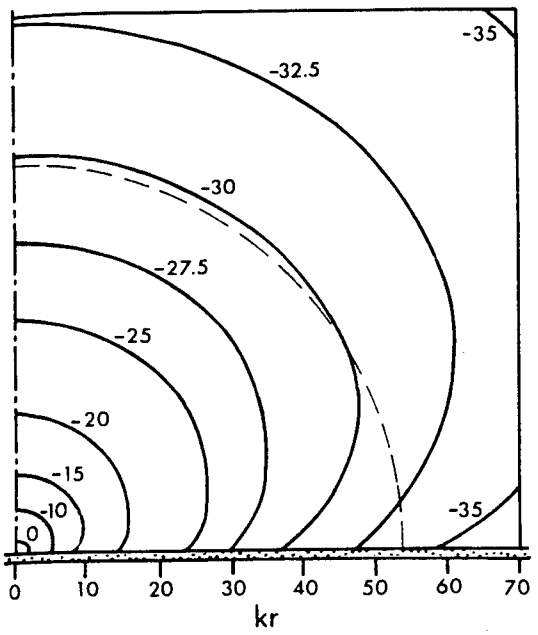


FIG. 5. Pressure distribution around a point source located on a plane with impedance ratio $Z_2/Z_1 = 10$ (Ref. 40). The dashed line gives, for comparison, spherical expansion over a rigid plane.

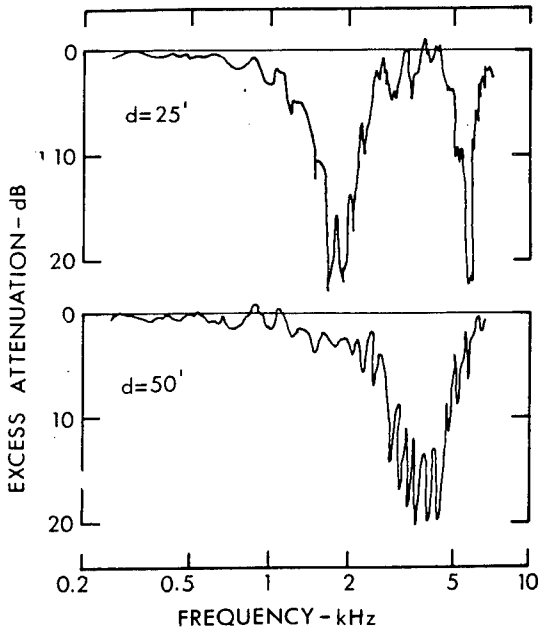


FIG. 7. Measured excess attenuation for propagation from a point source over asphalt, $h_s = 0.3$ m, $h_r = 1.2$ m. The excess attenuation is relative to that for the point source placed on a perfectly hard surface (Ref. 41).

$$A_e = 20\log_{10}[2\sin\psi(Z_2/Z_1)], \quad \text{dB} . \tag{3}$$

A comparison of this prediction with measured levels at distances out to 240 m for one well-chosen example is shown in Fig. 6. Note that the excess attenuation in the shadow zone is dependent only on ψ and the magnitude of the ratio Z_2/Z_1, and reaches values of ~ 30 dB. It is the result of cancellation between direct and reflected waves in the immediate vicinity of the source, produced by the phase change of nearly 180° upon reflection, as described previously for plane waves.

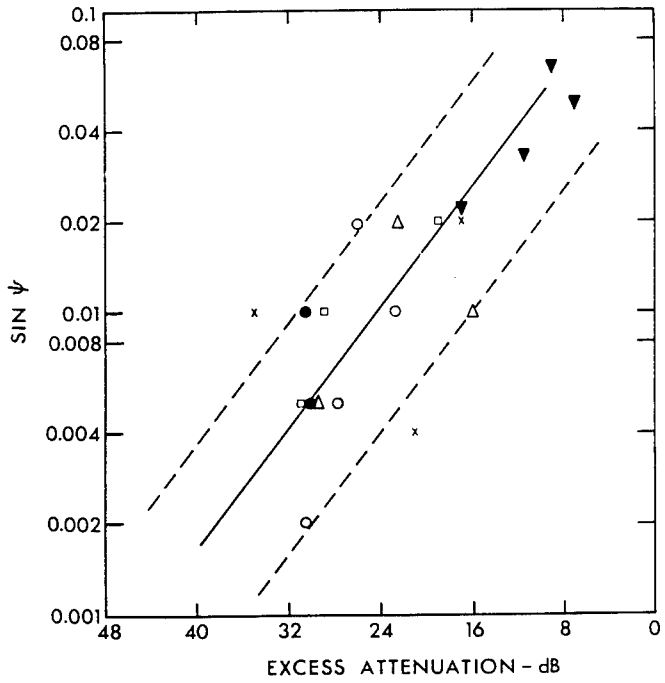


FIG. 6. Measured excess attenuation for a point source on the surface ($h_s = 0$) of a flat grassy site and the receiver at various heights ($0 \le h_r \le 5$ m) and distance ($d = 5$ m for ▼, 15 m for X, 30 m for ○, 60 m for △, 120 m for □, and 240 m for ●) for a frequency of 3 kHz. There was a brisk crosswind to eliminate the gradients of wind and temperature which cause refraction. The measurements are compared to the attenuation predicted (solid line) via Eq. (3) from the value of the ground impedance determined from measurements at short range. The dashed lines encompass 90% of the measured values (Ref. 38).

2. Effect of path-length differences—source and receiver both above a hard ground

When both the source and receiver are above the ground there is a phase change caused by the difference in length between propagation paths r_1 and r_2 in Fig. 4, in addition to the phase change on reflection described in the previous section. To show the effects of the former, free from the latter, consider propagation in the configuration of Fig. 4 above a hard, smooth surface, whose reflection coefficient R_p is effectively 1. Spectra of excess attenuation measured[41] for two specific configurations relevant to standard vehicle tests are shown in Fig. 7. The major dips in the spectra (of ~ 20 dB) are the result of cancellation between direct and reflected rays for path-length differences (PLD's) of an odd number of half wavelengths. The interference between direct and reflected sound over the rest of the spectra is approximately coherent addition. (Note that the minor dips and peaks are caused by turbulence, and will be considered later.) An elementary geometric calculation shows

$$\text{PLD} \approx 2h_s h_r/d \tag{4}$$

for small values of ψ. Note therefore that the frequency $f_{\min}$ of the first minimum will increase with increasing d, as shown in Fig. 7, and also decreasing h_s and h_r.

Spectra such as those shown in Fig. 7 for propagation

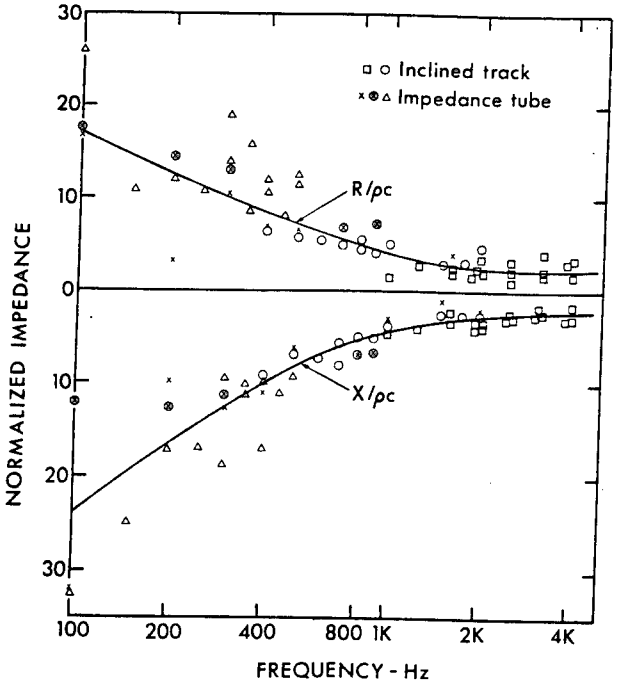


FIG. 8. Real component R_2 and imaginary component X_2 of the acoustic impedance of grass-covered flat ground (different areas of two different sites) obtained from measurements either with an impedance tube or of interference at oblique incidence (Ref. 38). (Note that the imaginary component corresponds to a stiffness reactance.)

over a hard surface may be divided into two regions: $f < \sim \frac{1}{2} f_{min}$ where the transmission characteristic is reasonably uniform, and $f > \frac{1}{2} f_{min}$ where it alternately shows the effects of constructive and destructive interference. The former is highly desirable for testing noise sources (see Sec. V A), and the importance of both regions for outdoor propagation in general will be discussed later in Sec. II E.

C. Ground impedance

Measurements of the acoustic impedance of the surface of the ground are scarce: only for mown grass, as found on airports and around public highways and buildings, is a substantial amount of data available.[38,42,43] The most detailed information on the impedance of mown grass[38] is shown in Fig. 8. It was obtained from measurements either with a vertical impedance tube, or from interference phenomena, at oblique angle of incidence, similar to those described in the previous section. The values of impedance shown in Fig. 8 confirm the grassy surface to have the acoustic properties of a locally reacting porous medium, which are well understood from the study of acoustic building materials. The impedance is determined by the presence of a thin porous layer on the surface that must have holes small enough to offer substantial resistance to the flow of air. Thus the impedance of short grass as in a lawn is not changed appreciably by mowing, as the main resistance to flow appears to be either in the turf itself or, more likely, in the soil near the surface kept loose by the roots. Different stretches of institutional or mowed airport grass are found to have similar values of impedance, and this impedance was not [illegible] weather.[42] It is important to note that the reactive component corresponds to a stiffness and not to a mass.

There is little information about the acoustic impedance of other ground surfaces. A stubble field,[43] or the ground under a pine plantation[44] both appear to have an impedance similar to that of grass, presumably because of the roots loosening the soil. Snow,[45] and ground recently loosened by discing[44] appear to have a much lower impedance at low frequencies. The impedance of asphalt,[46] although much higher than grass, is by no means infinite.

D. Ground and surface waves—source and receiver both on the ground

The third term in the Weyl–Van der Pol solution of the wave equation for spherical waves [Eq. (2)] arises mathematically from the need to match the boundary conditions, in particular the variation of the curvature of the wavefronts with distance along the boundary. Parameter F in this term represents a complicated mathematical function[35,39] of a variable w called the numerical distance. This function is plotted[34] versus w in Fig. 9 for various values of the phase angle $\phi = \tan^{-1}(X_2/R_2)$ of the surface impedance Z_2. The numerical distance is given by the expression[38]

$$w = (\tfrac{1}{2} ikr_1)(\sin\psi + Z_1/Z_2)^2 , \qquad (5)$$

where $k = 2\pi/\lambda$ is the propagation constant in air, and the other variables are defined in Secs. II A and II B. It is useful to consider the numerical distance w to represent the propagation distance r, scaled for a given value of frequency f, and grazing angle ψ, by the impedance Z_2, according to Eq. (5).

The physical interpretation of the third term in Eq. (2) is more elusive than that of the first and second terms which are the direct and reflected waves discussed in Sec. II B. The physical interpretation is most clear for both source and receiver on the boundary ($h_s = h_r = 0$ in Fig. 4, and $\psi = 0$) and the boundary resistive ($\phi = 0$), because of the direct analogy to the propagation of electromagnetic waves above the earth (for

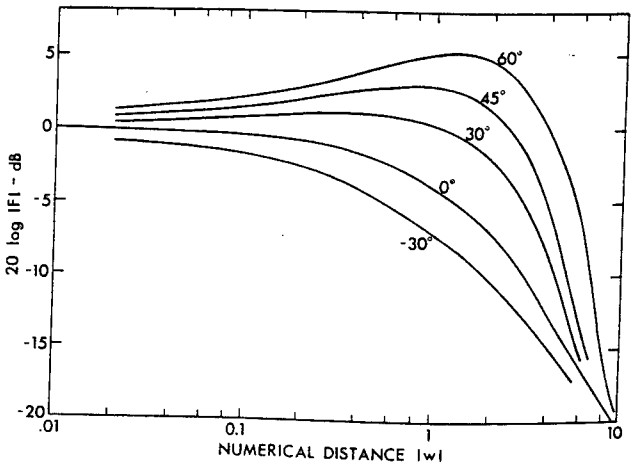


FIG. 9. Amplitude factor $F(w)$ of the ground and surface waves vs numerical distance w for various values of the phase angle

TABLE I. The propagation distance d_w for numerical distance = 1, and height z_{sw} for the amplitude of the surface wave to decrease by $1/e$ from its value at the boundary, calculated by Eqs. (5) and (6), respectively, from the impedances shown in Fig. 8 for mown grass.

f (Hz)	d_w (m)	z_{sw} (m)
50	4000	58
100	1500	23
200	270	7.3
500	24	0.9
1000	3.6	0.5
2000	1	0.15

which the Weyl–Van der Pol solution was originally derived). Under these conditions the direct and reflected waves cancel completely to form a shadow zone, as described earlier in Secs. II A and II B 1. This shadow zone is penetrated by the ground wave which is represented by the third term in Eq. (2) for $\phi = 0$. The amplitude of this ground wave is indicated by the curve for $\phi = 0$ in Fig. 9. Although the ground wave is important for radio communication, carrying the electromagnetic energy from the antenna of the local AM station to one's radio, its exact physical nature is still obscure.[32] The dependence on the variables, however, is well defined. The curve for $\phi = 0$ in Fig. 9 indicates that at short distances ($w \ll 1$) the ground wave suffers no excess attenuation compared to propagation over an infinitely hard surface, but for longer distances ($w \gg 1$) exhibits a loss of 6 dB per doubling of distance in addition to that provided by the inverse square law.

The data in Fig. 8 indicate, however, that the phase angle ϕ for the acoustic impedance of a grassy surface varies between approximately 45° and 60° over the audible range of frequencies. (By the sign convention used in this paper, a positive imaginary term for impedance corresponds to a stiffness reactance.) The amplitude of the function F for this range of ϕ in Fig. 9 differs from that for $\phi = 0$ (and hence differs also from the values used for radio wave propagation in the atmosphere) by the presence of a substantial increase in the vicinity of $w = 1$. This increase in the function F we now know to be mainly the contribution of a surface wave[36,37] in the air: this wave is coupled to the ground surface owing to the latter's stiffness reactance but propagates in the air, with an amplitude that decreases exponentially with height z above the boundary according to the formula

$$p = p_b \exp[-X_2^2 (R_2^2 + X_2^2)^{-1} kz], \qquad (6)$$

where p_b is the amplitude at the boundary. For $w \ll 1$, e.g., for d small, the contribution of the surface wave in Fig. 9 is relatively smaller than that of the ground wave because the surface-wave amplitude decreases at only 3 dB per doubling of distance (cylindrical expansion from a point source) compared to 6 dB for the ground wave. For $w \gg 1$ the amplitude of the surface wave again becomes smaller than the ground wave because of its attenuation by viscous losses in the pores of the boundary, which is exponential with distance r along the boundary.

In summary, theory indicates that the propagation of sound energy between a point source and receiver which are both placed on a grassy surface ($h_s = h_r = 0$) is by a ground wave, as in electromagnetic propagation, augmented by a surface wave. The scale of the distances and heights of these waves is given in Table I for a grassy surface whose impedance is that shown in Fig. 8. Some of the properties of these waves have been verified directly by measurements over grass[37,38] at short distances, where disturbing phenomena such as turbulence and refraction are absent, and by model experiments indoors.[39,47] The relevance of ground and surface waves to long range propagation when both source and receiver are above the ground will be discussed in the following section.

E. All phenomena together—source and receiver both above the ground

All the phenomena described previously in Secs. II A–II D must be considered when both point source and receiver are above the ground. The curves shown in Fig. 10, for propagation in the configuration of Fig. 4 with $h_s = 1.8$ m, $h_r = 1.5$ m, and various values of d, have been calculated[48] using the theory of Donato[37] with values of impedance taken from the curves for mown grass in Fig. 8. The individual contributions to these curves from direct D, reflected R, ground G, and surface S waves are illustrated[48] for three of these distances in Fig. 11, together with measurements of the propagation of jet noise for the same values of h_s and h_r and comparable d (110 and 615 m) under neutral atmospheric conditions at an airport.[45] (These measurements will be described in greater detail in Sec. IV B.)

Consider first the curves for a distance of 31.2 m in Fig. 11. Here the grazing angle ψ is sufficiently large for D and R waves alone [first two terms in Eq. (2)] to be a good approximation to the complete solution. Note

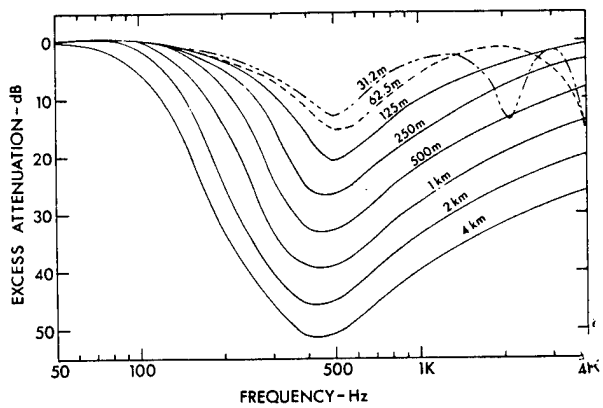


FIG. 10. Excess attenuation calculated for propagation from a point source over mown grass for $h_s = 1.8$ m, $h_r = 1.5$ m, and the distances of propagation d indicated. The attenuation is calculated for values of impedance given in Fig. 8. The excess attenuation is relative to that for the point source placed on a perfectly hard surface (Ref. 48).

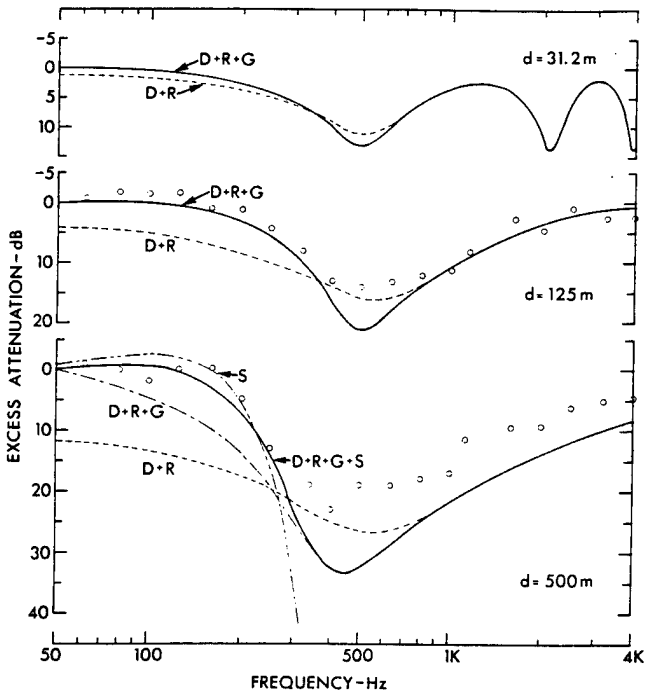


FIG. 11. Excess attenuation for propagation from a point source over mown grass, $h_s = 1.8$ m, $h_r = 1.5$ m. The calculated curves show the contributions from the various waves—direct D, reflected R, ground G, and surface S. The points are measurements of jet noise from Ref. 45 at comparable distances. The excess attenuation is relative to that for a point source placed on a perfectly hard surface (Ref. 48).

that the latter includes contributions from D, R, and G waves only, because the angle ψ is too large for an S wave to be significant. For frequencies greater than about 1 kHz there is a series of maxima and minima caused by path length differences (PLD's) between direct and reflected waves, that is an interference pattern similar to that described in Sec. IIB for a hard surface (see Fig. 7). Unlike the pattern for a hard surface, however, the minima are for PLD's which are approximately even multiples of $\frac{1}{2}\lambda$, indicating that the surface is effectively soft for this configuration and frequency range. As the frequency decreases, however, the ground impedance increases (see Fig. 8) causing the surface to give a response which is effectively hard for low frequencies ($f < \sim 150$ Hz) as in Fig. 7. Unlike the response of the hard surface shown in Fig. 7, however, there is a broad minimum between these two ranges centered at about 500 Hz, which is characteristic of the propagation for short distances over soft ground. It is the result of cancellation between direct and reflected waves caused primarily by the phase change on reflection. The dip shown in Fig. 11 for a distance of 31.2 m, therefore (for source and receiver now both raised above the surface), is the remains of the shadow zone described previously in Sec. IIB, for the source (or receiver) on the ground. It has caused major problems for the testing of noise sources such as snowmobiles at short ranges, because the cancellation and hence the noise, is dependent on the impedance of the surface of the particular site.[49]

Because the curvature of the wavefront from a point source decreases with increasing distance, the behavior for long ranges must be asymptotic to that described in Sec. IIA for plane waves. The curves in Fig. 10 indicate that the dip broadens and deepens with increasing distance until most of the audible frequency range is included in the shadow zone in accordance with this principle. Comparison of the curves in Fig. 11 for a distance of 125 m with those for 31.2 m shows how this broadening comes about. The interference pattern, which occurs at high frequencies for $d = 31.2$ m, has, for $d = 125$ m moved upward in frequency beyond 4 kHz, by virtue of the d^{-1} term in Eq. (4), leaving only the rising slope to the first maximum in Fig. 11. At the low frequency end, the increasing cancellation between D and R waves produced by decreasing ψ [as in Eq. (1)] has revealed a substantial contribution from the ground wave. The measurements are in reasonable agreement with theoretical prediction for 125 m, using impedances obtained from a different site.

Consider now the curves for 500 m in Fig. 11. The broadening of the shadow zone to higher frequencies evident in the curves for 125 m has continued with increasing distance. There has also been a broadening to the lower frequencies, which leaves most of the energy in the frequency range 50–200 Hz in the surface wave, and the latter fact remains true for longer distances. The measurements for 500 m at low frequencies closely follow theory, showing a small enhancement (negative excess attenuation) due to the surface wave.[36] There is, however, a tendency at the high frequencies for the measured excess attenuation to be consistently less than predicted.

The reason for the high-frequency discrepancy probably has to do with the use of coherent acoustic theory for the predicted curves. Turbulence in the atmosphere is known to reduce coherence between different propagation paths. While it would be premature to attempt to predict the size of this effect from existing knowledge, to be realistic one must expect significant departures from the curves shown in Fig. 10 due to turbulence (see Sec. IIIB) particularly at higher frequencies, longer distances, and when the excess attenuation due to interference between direct and reflected waves would otherwise become large (greater than ~20 dB, see Sec. IIIC).

In practice, for broadband noise sources, such as a jet engine, and for distances greater than about 1 km, the high frequencies will be attenuated sufficiently by atmospheric absorption, as shown in Fig. 3, and the midfrequencies will be attenuated by the ground shadow, as shown in Fig. 10, so that the main contribution to the measured A-weighted sound level will come from the surface wave at frequencies below 200 Hz, as illustrated in Fig. 11. It is likely also that the background roar from distant traffic, is transmitted mainly via this surface wave. It is interesting to note that the attenuation of the surface wave is mainly by the viscous flow of air in the pores of the ground.

F. Summary

In the preceding Secs. IIA–IIE the present understanding of the near-horizontal propagation of sound in

an acoustically homogeneous atmosphere close to flat ground has been outlined. The primary effect is a shadow zone caused by the finite acoustic impedance of the ground surface. This shadow zone is penetrated at low frequencies by the ground and surface waves. For both source and receiver above the boundary, the shadow zone is also penetrated at high frequencies by constructive interference between direct and ground-reflected waves. Sound levels can be calculated with reasonable precision using known theory within the constraints imposed by the present knowledge of the ground impedance and the use of coherent wave theory. The effect of inhomogeneity of the atmosphere produced by surface meteorology, as well as surface topography, will be discussed in Secs. III and IV.

G. Prediction schemes

It appears from the analysis in Sec. II E, that the modern theory of the propagation of sound from a point source has good potential for predicting the excess attenuation of noise produced by the ground surface within the constraints reviewed in Sec. II F above. There are two equivalent formulations, by Wenzel[36] and Donato,[37] with the latter more suitable for computation. The theory and the mechanisms of propagation are complicated but the solutions are programmable, so that this method appears worthy of development.

A useful but less general approach has been suggested by Delany and Bazley.[50,51] They adapted the limited exact mathematical solution of Wise[52] for electromagnetic propagation to the propagation of sound. This solution is equivalent to taking only the first two terms in Eq. (2). The simplified method of Delany and Bazley has the additional constraint, therefore, to regions where the contributions from the ground and surface waves may be neglected. It appears from the discussion in Secs. II D and II E (and Fig. 11) that this simplified scheme should not be used for low frequencies at small grazing angles. It should, however, be useful for evaluating the noise from aircraft in flight, the purpose for which it was designed, because here two of the other major constraints (weather and topography) are often not a problem.

III. EFFECT OF SURFACE METEOROLOGY

A. Relevant meteorological phenomena

The variation of the average speed of the wind V_w with height z in the vicinity of the ground for a large flat open area is approximately as shown in Fig. 12.[53–55] For heights greater than z_0, V_w may be represented by the formula

$$V_w = K_v \log(z/z_0) \ . \tag{7}$$

Parameter z_0 is determined by the roughness of the surface, and is often very approximately the height of a consistent obstacle, such as grass or corn. The region of logarithmic variation shown by Eq. (7) is caused by the viscous drag of the surface and is known as the viscous boundary layer. The constant K_v is determined by the roughness of the surface, and the wind velocity above this boundary layer, which is usually not greater than ~ 10 m thick, the height at which wind is normally measured at an airport. In practice it is often necessary to consider deviations from the velocity profile given by Eq. (7) caused by the buoyancy introduced by the temperature profile.

In the vicinity of the ground, the variation in average temperature T with height for a large flat area may be represented during the daytime by the analogous expression

$$T = T_0 - K_t \log(z/z_0) \ , \tag{8}$$

where T_0 is the temperature for $z < z_0$. The thermal boundary layer given in Eq. (8), which normally has a thickness and value of z_0 similar to that of the viscous layer, is caused by the heating of the surface by the sun. It normally coincides with a condition known to meteorologists as lapse from measurements well above the boundary layer, an example of which is given[56] in Fig. 13. Shown also in Fig. 13 is an example of an inversion caused by the cooling of the surface of the ground through radiation to the night sky. The term "neutral" will be used later in the discussion to signify a third meteorological condition, where the dependence of temperature with height is small: This condition may arise typically under heavy clouds.

B. Refraction

1. Basic forms

Forms of refraction produced by different meteorological conditions are shown in Fig. 14. While the effects of wind and temperature gradients appear similar in the figure, the following differences should be noted. Because temperature is a scalar quantity the refraction of sound produced by lapse or inversion conditions is the same in all horizontal (compass) directions. Wind, however, produces refraction nonuniform in direction according to the vector component in the direction of propagation. Thus the refraction produced by the wind is zero when the sound propagates directly crosswind, and increases progressively as the direction of propagation deviates from this condition.

The major acoustic effect in refraction upwards as shown in Fig. 14(b) is the production of a refractive shadow zone, shown crosshatched in the figure, where, according to the ray picture (geometric acoustics) no sound may enter. Refraction downwards, as shown in

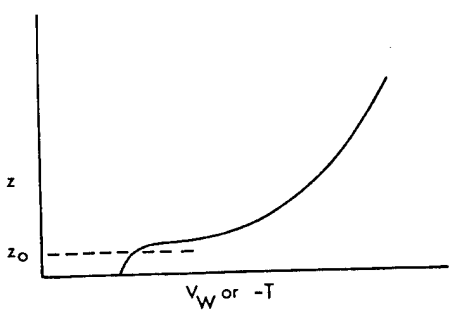


FIG. 12. Variation of wind velocity and temperature in the vicinity of a flat ground surface ($z < \sim 10$ m).

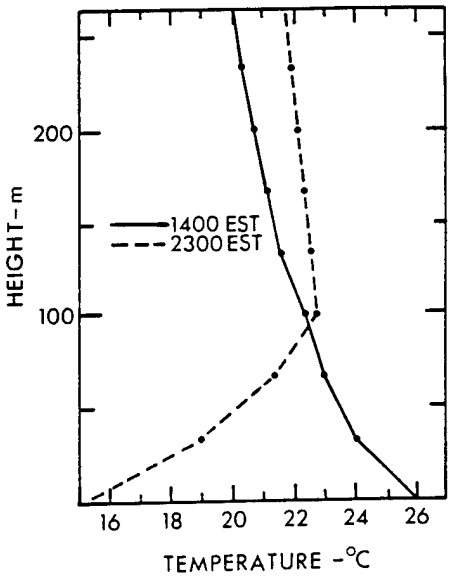


FIG. 13. Variation of temperature aloft—examples of lapse (afternoon) and inversion (evening) conditions from Ref. 56.

Fig. 14(a) tends in practice to enhance the sound level, by mechanisms which will be discussed in the next section.

2. *Refraction effects for distances less than 1500 m*

The work of Refs. 3, 45, 53, and 57–65 may be fitted together to provide a consistent picture of the effect of refraction on the propagation of sound over distances up to 1.5 km using the meteorological and refraction phenomena described in the two previous sections and the propagation phenomena described in Sec. II. These phenomena will be described with the help of the detailed measurements of Parkin and Scholes[45] shown in Fig. 15 for the propagation of jet noise over mown grass on a reasonably level airport.

Measured spectra of attenuation in excess of that caused by spherical spreading and atmospheric absorption are shown in Fig. 15 for two different distances and a number of different meteorological conditions. The source height is 1.8 m and the receiver height 1.5 m. The curves each represent the average of a number of measurements on different days. The dotted, dashed, and solid curves for each distance are designated +5, 0, and −5 to indicate the vector component of wind velocity in the direction of propagation in meters per second.

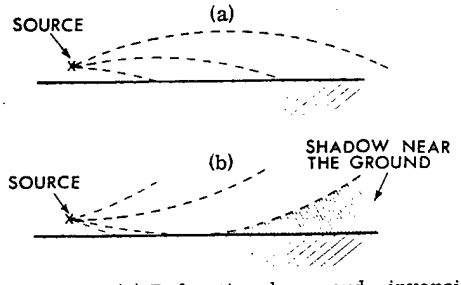


FIG. 14. (a) Refraction downward—inversion or downwind propagation. (b) Refraction upwards—lapse or upwind propagation.

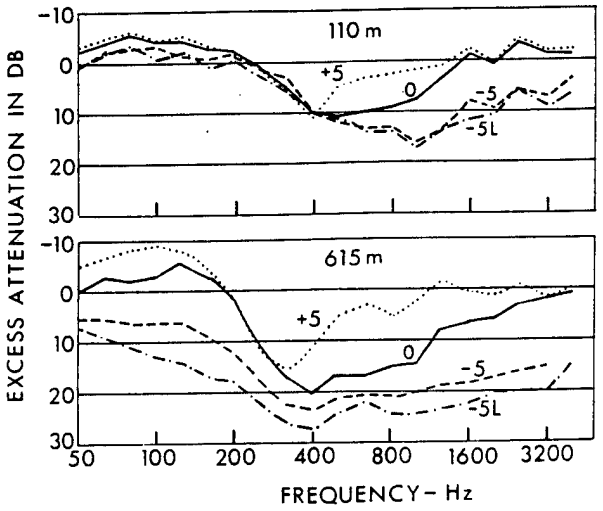


FIG. 15. Observed attenuation of aircraft noise in a ground-to-ground configuration under a variety of weather conditions (Ref. 45). Calculated losses from atmospheric absorption and spherical spreading have been subtracted from the attenuation measured in 1/3-octave bands for distances of 110 m and 615 m. The numbers on the curves indicate the vector component of the wind velocity in the direction of propagation in m/sec. All curves are for neutral conditions of temperature except for those marked *L*, which are for lapse.

For these three curves the temperature conditions are classed as neutral, as described in Sec. III A. Curve 0 for each distance therefore indicates the attenuation produced by the ground effect, as described in Sec. II E, and in fact is composed of straight lines joining the experimental points in Fig. 11. Curve −5 (in Fig. 15) indicates the additional attenuation produced by a wind shadow zone, and curve +5 the enhancement during downwind propagation.

Parkin and Scholes resolve their measurements into two further categories according to the gradient of temperature—either lapse or inversion. The result is generally to change the labels on their curves for all distances by one step (exceptions will be discussed later). Thus their lapse +5 curve is close to the neutral zero curve in Fig. 15, and the lapse zero curve is equivalent to the neutral −5 curve. The lapse −5 curve, which is the bottom curve (−5*L*) in the figure, represents one of the exceptions. Inversion conditions could only be attained for light winds, and their inversion zero curve is very close to the neutral +5 curve in the figure.

The first conclusion to note, therefore, from the measurements of Parkin and Scholes is that refraction due to vertical gradients of wind and temperature in practice produces equivalent acoustic effects; these are additive, moreover, within the limits imposed by a saturation phenomenon to be described presently. To understand the physical nature of these effects it is useful to divide the frequency range for the spectra shown in Fig. 15 into three parts: a small central region from about 200–500 Hz where the effect of refraction is smallest, and separate regions for frequencies above and below this central region.

(*a*) *High-frequency region* ($f > 500$ *Hz*). In this re-

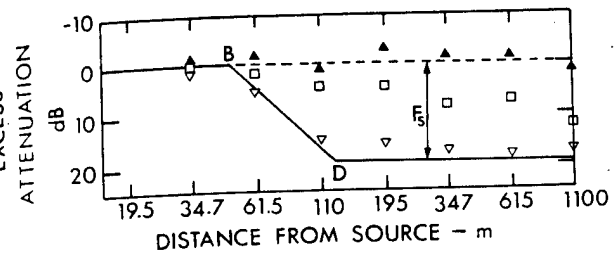


FIG. 16. Excess attenuation measured for aircraft noise in the 1.2-kHz 1/3-octave band for the ground-to-ground configuration (Ref. 45). The vector component of the wind velocity in the direction of propagation for ▲ is +5 m/sec, □ is 0 m/sec, and ▽ is −5 m/sec. The temperature profile is neutral. F_s is the shielding factor, B is the shadow boundary.

gion the measurements show a strong tendency towards two extreme values for the excess attenuation, approximately 0 dB in conditions of downward refraction (inversion or downwind propagation) and −20 dB in conditions producing upward refraction (lapse or upwind propagation), seemingly independent of frequency or distance (when the latter is sufficiently large). This pattern is reasonably clear in the measurements for a distance of 615 m shown in Fig. 15, and obvious in the measurements for a frequency of 1.2 kHz over a range of distances shown in Fig. 16.

The pattern for conditions of upward refraction is idealized by the solid lines shown in Fig. 16. There is effectively no excess attenuation in propagation out to the shadow boundary B in the figure, which is typically 50 m away from the source. The excess attenuation then increases rapidly to position D, but then stays independent of further increases in distance. In effect the signal well within the shadow zone follows the inverse square law but is also reduced in level by a fixed attenuation given by the shielding factor F_s in Fig. 16.

It should be noted that the pattern of behavior described above is independent of the magnitude of excess ground attenuation. Thus the measurements for a distance of 615 m in Fig. 15 indicate at a frequency of 3–4 kHz, where the ground effect is small, there is effectively no enhancement of the signal during downwind propagation, and for a frequency of ~600 Hz, where the ground effect is large, the increase in attenuation in upwind propagation indicated by the −5 curve is small. Furthermore, the $-5L$ curve indicates that the additional refraction produced here by adding a temperature gradient is also small. This effect has been described previously as a saturation.

To explain these saturation characteristics, the following mechanism has been proposed for the solid line in Fig. 16 (upward refraction). The behavior well within the shadow zone is as if the signal were from a flanking path. The refractive shadow boundary is known to make a small angle, typically $< 5°$, with respect to the ground (see Sec. IV A), and it has been widely suggested[3,10,45,58] that the flanking signal is energy scattered down across this boundary by atmospheric turbulence, as indicated by the dashed arrows in Fig. 17. However, this model for the mechanism of the shielding factor remains untested by experiment or quantitative analysis.

The method by which the excess attenuation becomes 0 dB for downward refraction, as given by the dashed line in Fig. 16, is also not clear. The only *a priori* likely way is by destruction of the ground effect. The curvature in the ray path increases the grazing angle ψ in Fig. 4, hence reduces the cancellation which causes the ground effect.[38]

(*b*) *Low-frequency region* ($f < 200$ *Hz*). The curves for $d = 110$ m in Fig. 15 indicate that the effects of refraction are much smaller for low frequencies than for high: the physical reason is that the scale of the strong gradients in wind and temperature which occur close to the ground (see Fig. 12 and Sec. III A) become small compared to the wavelength of sound at the lower frequencies.[3,58,62] Refraction still occurs for the low frequencies but at much larger distances. The curves marked −5 and $-5L$ for $d = 615$ m in Fig. 15, for example, indicate that at this distance the shadow zone is still only partially formed, i.e., the excess attenuation falls part way along line BD in Fig. 16, similar to the curves for −5 and $-5L$ at high frequencies and a distance of 110 m in Fig. 15.

The +5 curve for a distance of 615 m in Fig. 15 indicates that during downwind propagation or in an inversion only low-frequency signals are enhanced well above that expected by inverse square law. A likely mechanism for this enhancement is provided by the unique mode of propagation at these frequencies, as described in Sec. II E. The surface wave may be attributed to the porosity of the ground retarding the velocity of propagation in the atmospheric layer immediately above the surface. This retardation is enhanced by the vertical gradient of wind velocity during propagation downwind, or the vertical gradient of temperature during an inversion.

(*c*) *Central frequency region* ($f = 350 \pm 150$ *Hz*). A study of curves similar to those in Fig. 15 for other sites[59,61] and geometrical configurations[61] (h_s, h_r, d) shows the position of this limited region, where the effects of refraction are small, to be in the vicinity of the frequency of maximum ground attenuation, which is somewhere in the range 200–500 Hz. The −5 and $-5L$ curves in Fig. 15 for a distance of 615 m indicate that the small effect of upward refraction in this region is the result of saturation in the shadow zone, as described above for high frequencies.

The persistence of the ground shadow indicated by the coincidence of curves 0 and +5 for a distance of

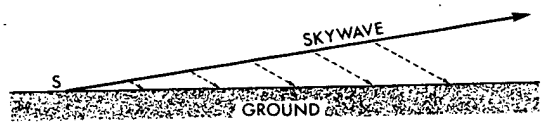


FIG. 17. A mechanism for the shielding factor, F_s in Fig. 16. The dashed lines represent the continuous scattering of energy down from the sky wave, which propagates without excess attenuation from source S.

615 m is remarkable, however, and almost certainly reflects the unusual mechanism of propagation described in Sec. II D and II E. Thus, the sharp increase in excess attenuation with frequency between 200 and 300 Hz at a distance of 615 m was shown to be caused by the absorption of energy in the surface wave by the viscous flow of air in the pores of the ground. The coincidence of the +5 curve with the zero curve in Fig. 15, therefore, signifies that this loss cannot be replaced by refraction.

3. *Prediction schemes*

Wiener and Keast[57] have provided an empirical prediction scheme for the effects of refraction by wind and temperature which is applicable to the high-frequency region described above. In this scheme the distance to the shadow boundary B in Fig. 16 is calculated assuming linear vertical gradients of wind and temperature. Delany[58] has improved this scheme by calculating the distance for the logarithmic profiles in the boundary layer indicated by Eqs. (7) and (8). Kriebel[66] has also provided an interesting analysis of the propagation in the boundary layer.

It should be noted, however, that measurements in the low frequency region described above show the excess attenuation to depend on the surface impedance, which is outside the scope of Wiener and Keast's method.

It should also be noted that for propagation over distances much greater than 1.5 km, different factors become important. There are scattered measurements[67,68] which indicate a much higher attenuation in the shadow zones than predicted by Wiener and Keast. In downward refracting conditions the relevant portion of the vertical profiles of wind and temperature become higher than the logarithmic boundary layer given by Eqs. (7) and (8). Inversions then become more important (see Fig. 13) and a theoretical analysis of the excess attenuation during propagation in inversions is available.[69]

Empirical models have also been developed[70] to explain systematic trends observed during a very large number of measurements of low-frequency sound propagation over flat wooded terrain at distances up to 8 km. The average observed excess attenuation was primarily determined by the sign and magnitude of the vertical gradient in the speed of sound, and secondly by the wind speed.

C. Turbulence

Large eddies are formed in the atmosphere by instabilities in the thermal and viscous boundary layers at the surface of the ground, described in Sec. III A. Further instability causes these eddies to break down progressively into smaller and smaller sizes until the energy is finally dissipated by viscosity in eddies approximately 1 mm in size. A statistical distribution of eddies, which we call turbulence, is therefore present in the atmosphere at all times.[54,71] The intensity of the turbulence, however, is strongly dependent on meteorological conditions (and also height above the ground), being high, for example, on a windy summer afternoon, and low under nocturnal inversions.

The effect of atmosphere turbulence on wave propagation has been studied extensively during the last twenty years,[71] most notably on the propagation of light and microwaves, but also of sound.[71,72] For sound, the condition most studied is near-vertical propagation, well away from the surface of the ground, to permit the "acoustical sounding" of meteorological conditions.[73] There are a small number of investigations of near-horizontal propagation of sound close to the ground[74,75] which are of direct relevance to noise problems, but for most questions concerning the effect of turbulence on noise propagation it is necessary to fall back on knowledge accumulated in these related areas.

It is known that atmospheric turbulence produces fluctuations in the amplitude and phase of a pure tone during propagation which increase with increasing distance until a point is reached where the phase fluctuations have a standard deviation comparable to 90°. At this point the signal effectively becomes uncorrelated with the source and the amplitude fluctuations are limited to a standard deviation[75] of ~6 dB. The distance to this point for horizontal propagation of sound near the ground on a summer day is very approximately 700λ, for frequencies in the range 500–5000 Hz.[75] An effect of this magnitude brings into question the common use of coherent acoustic theory for predicting noise propagation in a number of critical regions, some of which have already been mentioned (Secs. II E and III B).

Perhaps the condition most sensitive to atmospheric turbulence is interference phenomena: the sharp spikes clearly visible in the minima of the interference patterns shown in Fig. 7, for example, are caused by fluctuations in the phase difference between direct and reflected waves, and the effective depth of the minima is determined by the fluctuations.[74] While a good start has been made on the theory of interference in a turbulent medium,[74] much more work will be needed on partially coherent propagation before a satisfactory method is achieved for predicting the ground effect at the longer ranges and higher frequencies shown in Fig. 10.

Another role of atmospheric turbulence which is important for the prediction of noise levels, is in determining the saturation effect in shadow zones, as described in terms of a shielding factor in Sec. III B. The role is in fact even more general than described there: it should be expanded to include the acoustic shadow provided by walls,[7,9] buildings,[6] the ground effect,[10] etc. We are very limited at present in our ability to predict levels in shadow zones because of an almost complete lack of understanding of the basic phenomena which determine the shielding factor.

A less profound aspect of atmospheric turbulence, but still a nuisance, is the need to cope with amplitude fluctuations in specifying noise levels[76] from relatively distant sources such as aircraft. One recent experimental study of sound propagation to the ground from a 152 m tower indicated the potential significance of the Richardson's number on the magnitude of these fluctuations.[77]

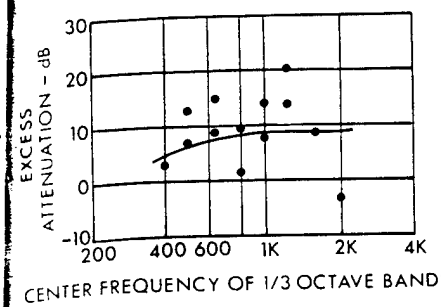


FIG. 18. Excess attenuation derived from hilltop to hilltop transmission measurements over a distance of about 3 km. The points are measurements which were corrected for inverse square law and molecular absorption using the method of Ref. 16 by the original investigators (Ref. 57). The curve is molecular absorption from the nitrogen relaxation, calculated by the method of Ref. 23, which was neglected in the method of Ref. 16.

There is one role of atmospheric turbulence which used to be considered important[2] in predicting noise levels, but now is generally believed to be minor[78]—that is the direct attenuation of sound by turbulence. If the sound is in a highly directed beam, as for example a vertical atmospheric sounder,[73] then turbulence attenuates the beam by scattering energy out of it (a phenomenon converse to the filling of the narrow shadow zone by scattered energy described in Sec. III B). This attenuation can now be calculated with reasonable certainty.[79] In practice, however, the propagation of noise is usually much closer to spherical expansion from a point source. Because the scattering from turbulence is elastic, and mostly forward through a small scattering angle, the attenuation presented to a spherically expanding acoustic field is negligible. The attenuation of sound due to scattering from a moderately directional source, such as a jet or rocket engine, must lie between these two extremes, but has never been evaluated thoroughly. It is generally believed to be negligible[24,45,57,58] for most applications. However, where very low frequencies are of importance, for example in the case of rocket noise, the attenuation by scattering may be significant[80] compared to the very low attenuation produced by atmospheric absorption.

IV. EFFECT OF TOPOGRAPHY

The acoustic phenomena considered previously in Secs. II and III have all been described for a particular topography, primarily large flat areas covered with short grass, such as an airport, with a receiver placed approximately ear height for a human being standing on the ground. In practice, noise levels need to be predicted for different heights above the ground, for hillsides, for various types of foliage, houses, walls, etc. The data available for such sites is generally much less extensive than for flat ground, and present knowledge is therefore more sketchy. The aim of this section, however, is to outline briefly what is known about the effects of topography. Propagation in city streets, and the attenuation due to barriers will be specifically excluded because these subjects have been covered in recent reviews[6,7] which this article is designed to accompany.

A. Elevation

As an extreme case, consider the propagation from hilltop to hilltop with most of the propagation path several hundreds or thousands of feet above the valley floor. Several sites of this nature have been investigated,[57,81] and the measured attenuation is in reasonable agreement with that predicted by inverse square law and atmospheric absorption, when the latter includes the vibrational relaxation of nitrogen (see, for example, Fig. 18 and the discussion on this point in Sec. I B). This attenuation is found to be independent of wind and temperature except in so far as the latter enters into the calculation of absorption.

Studies of noise from low-flying aircraft—effectively sources at low glancing angles ψ (Fig. 4)—over flat ground are informative.[82] They indicate, for example, that refraction by wind gradients is negligible for $\psi > 5°$ out to a distance of about 6000 m. There is currently a controversy, however, over the prediction of aircraft noise levels for small ψ, and this problem would probably be well served by an assembly of existing data from many investigators and a fit to the principles outlined in Sec. II.

Calculations for the effect of elevation produced by a shallow hillside ($\psi < 1°$) bordering an airport[38,83] have indicated a significant reduction of the ground shadow for aircraft on the ground. There were no supporting measurements but the results correlated well with subjective observations (the structure of complaints).

B. Foliage

The effect of stands of trees, corn, reeds, etc. have been documented in several studies.[44,53,84–87] Apparently the main effect at low frequencies is to enhance the ground effect,[44] the roots making the ground more porous, as described in Sec. II C. At high frequencies, where the dimensions of leaves become comparable with the wavelength, there is also a significant attenuation caused by scattering.[85,86] A formula for the latter effect has been developed.[86] The propagation in a vegetative layer over the ground has also been investigated theoretically,[30] but there were no measurements by which to judge the significance of this theory.

In a forest, the vertical gradients of wind and temperature are small,[56] the foliage making z_0 in Eqs. (7) and (8) approximately the height of the trees. Presumably, therefore, the effects of refraction are small also.

C. Obstructions (walls, houses, bushes, etc.)

The effect of refraction on the propagation of sound in city streets[88] as within forests, is small, and for the same reason: the obstruction to the flow of air is sufficient to raise the viscous and thermal boundary layers to the vicinity of the tops of the obstacles (buildings) (see Sec. III A). The attenuation of sound by the ground

effect in the city[6] is also much smaller than in the country,[45] partly because of the paved ground, but more importantly because obstacles to the propagation of sound produce an interference pattern much different from the simple interference between direct and ground reflected waves described in Sec. II. These effects have been studied both for cities[6,88] and open flat land,[45,57] with results that are very different. There has been very little work, however, on intermediate sites such as suburbs. Questions such as how many obstacles are needed to destroy the effect of the ground, or refraction, as measured for flat land, cannot be answered at present. As a result the prediction of the propagation of noise from airports, freeways, etc., out into the suburbs is based largely on empirical methods not well founded on either analytical models or extensive experimental data.

V. PROPAGATION PROBLEMS SPECIFIC TO COMMUNITY NOISE

A. Testing and certification of sources

1. *Vehicles and other ground based sources*

There are several propagation problems in the testing of ground based noise sources such as road vehicles, lawn mowers, snowmobiles, etc.[46,49] The noise is usually measured with a microphone 1.2 m above the ground at a distance of 7–16 m with the propagation path over the surface on which the machine is designed to operate. If the surface is porous, for example grass or snow, then there can be a problem of variability as a result of the incipient ground shadow described in Sec. IIE. The solution here is a compromise between measurement configurations, a surface of predictable impedance, and relevance to the use of the machine. If the source is stationary, most of the relevant scientific knowledge[38,49] is available, with the exception of good data on the impedance of relevant surfaces. A convenient method of measuring the impedance of surfaces over a range of grazing angle (ψ in Fig. 4) and frequency is badly needed.

If the surface is hard, there are problems of repeatability for sources with large-amplitude pure-tone components when the transmission path introduces interference maxima and minima at similar frequencies. This phenomenon is reasonably well documented[46,49] (see also Sec. IIB). Variability due to atmospheric turbulence has also been recorded.[41,89]

There is also an unresolved problem of variability for moving sources such as vehicles.[41] There is a systematic variation in A-weighted pass-by levels with ambient temperature[41,90] of -1 dB/10 °C, and a run-to-run variation of approximately 1 dB, both of which appear to be associated with the movement of the vehicle. The most obvious mechanism here is refraction by the movement of air in the wake of the moving vehicle.[41]

2. *Aircraft*

The new method for calculating atmospheric absorption[23,24] described in Sec. IB should bring the physics of the propagation from overhead aircraft under reasonable control. However, the factors which influence sideline propagation are much less well understood. A substantial amount of work has been done on the acoustic effects of the ground plane during ground tests of jet engines, for which a recent review is available.[91]

B. Prediction of environmental impact

The basic physics underlying the propagation of noise at long range close to the ground described in Secs. I–IV is obviously intricate, and current knowledge of it patchy. As a result it is usually not possible in practice to predict noise levels with reasonable reliability at sites distant (hundreds or thousands of meters) from a strong source near the ground using basic knowledge alone. It is possible for a specific site, however, to take a completely empirical approach, measuring the noise from a source at positions of interest under various weather conditions, and predicting, for example, the levels to be expected at these positions from a highway or airport proposed for this site with tolerable precision.[92] After several of these individual investigations, it is tempting to generalize the empirical results for average levels and apply them to other sites. This procedure has been popular in recent years, through the development and use of empirical design guides for highways and airports. These prediction methods generally ignore conditions of topography and weather which are known to have a strong influence on sound propagation, as described in Secs. II–IV. Such design guides may be expected to provide reasonable predictions on the average, but for a specific site their unaided use may well result in serious error. Consider, for example, the problems in the prediction of noise from airports reported by Piercy and Embleton,[83] and Dickinson,[93] and also the considerable difference between empirical rules for predicting noise levels from highways based on different (but large) sets of data reported by King and Gordon,[94] Hajek,[95] and Plotkin and Kunicki.[96] In a field which is physically complex and only partly understood, such as outdoor sound propagation, there is still a decided need for human judgment based on specialist's knowledge of the current state of the art.

C. Recommendations

It is recommended, in the light of the analysis in the preceding section, as well as the history of research outlined in the introduction, that long-term support be developed at a modest level for the field of outdoor sound propagation. The aim should be to obtain the depth of understanding required to gradually improve the basic knowledge in this complex field. Attainment of the necessary depth of understanding is not compatible with short-term research studies.

ACKNOWLEDGMENTS

The authors wish to acknowledge the many ideas and extensive support of their colleagues, both professional and technical, in the Acoustics Section of the Physics Division of the National Research Council of Canada, and in Wyle Research, El Segundo, California.

[1]J. W. Wescott and S. S. Kushner, "Propagation of Sound in Air—A Bibliography with Abstracts," Report of Geophysics Lab., University of Michigan (1965) (AD 465 678).
[2]K. U. Ingard, "Review on the Influence of Meteorological Conditions on Sound Propagation," J. Acoust. Soc. Am. 25, 405–411 (1953).
[3]K. U. Ingard, "The Physics of Outdoor Sound," in *Proceedings of the Fourth Annual Noise Abatement Symposium* (Publisher, City, 1955), pp. 11–25.
[4]W. L. Nyborg and D. Mintzer, "Review of Sound Propagation in the Lower Atmosphere," WADC Technical Report 54–602 (1955).
[5]I. Rudnick, "Propagation of Sound in Air," in *Handbook of Noise Control*, edited by C. M. Harris (McGraw–Hill, New York, 1957), pp. 3-1 to 3-17.
[6]R. H. Lyon, "Role of multiple reflections and reverberations in urban noise propagation," J. Acoust. Soc. Am. 55, 493–503 (1974).
[7]U. J. Kurze, "Noise reduction by barriers," J. Acoust. Soc. Am. 55, 504–518 (1974).
E. J. Rathe, "Note on Two Common Problems of Sound Propagation," J. Sound Vib. 10, 472–479 (1969).
[9]U. J. Kurze, "Sound Propagation Outdoors," in *Noise and Vibration Control*, edited by L. L. Beranek (McGraw–Hill, New York, 1971), pp. 164–193.
[10]E. A. G. Shaw and N. Olson, "Theory of Steady-State Urban Noise for an Ideal Homogenous City," J. Acoust. Soc. Am. 51, 1781–1793 (1972).
[11]R. B. Tatge, "Noise Radiation by Plane Arrays of Incoherent Sources," J. Acoust. Soc. Am. 52, 732–736 (1972).
[12]L. C. Sutherland, "Ambient noise level above plane with continuous distribution of random sources," J. Acoust. Soc. Am. 57, 1540–1542 (1975).
[13]G. J. Thiessen, "Community Noise-Surface Transportation," J. Acoust. Soc. Am. 53, 303(A) (1973).
[14]J. E. Piercy, "The Absorption of Sound in the Atmosphere," J. Acoust. Soc. Am. 52, 1310(A) (1972).
[15]W. Pohlman, "Die Schallabsorption von Luft verschiedener Feuchtigkeit zwischen 10 and 100 kHz," *Proceedings of the Third International Congress on Acoustics, Stuttgart*, edited by I. L. Cremer (Elsevier, Amsterdam, 1961), pp. 532–535.
[16]E. J. Evans and E. N. Bazley, "The Absorption of Sound in Air at Audio Frequencies," Acustica 6, 238–244 (1956).
[17]C. M. Harris, "Absorption of Sound in Air Versus Humidity and Temperature," J. Acoust. Soc. Am. 40, 148–162 (1966).
[18]C. M. Harris and W. Tempest, "Absorption of Sound in Air below 1000 cps," J. Acoust. Soc. Am. 36, 2390–2394 (1964).
[19]L. B. Evans, H. E. Bass, and L. C. Sutherland, "Atmospheric Absorption of Sound; Theoretical Predictions," J. Acoust. Soc. Am. 51, 1565–1575 (1972).
[20]H. O. Kneser, "A Nomogram for Finding the Sound Absorption Coefficient in Air," Ak. Zeit 5, 256–257 (1940).
[21]J. E. Piercy, "Role of the Vibrational Relaxation of Nitrogen in the Absorption of Sound in Air," J. Acoust. Soc. Am. 46, 602–604 (1969).
[22]SAE Committee A-21, "Standard Values of Atmospheric Absorption as a Function of Temperature and Humidity for Use in Evaluating Aircraft Flyover Noise," Society of Automotive Engineers, ARP866 (1964).
[23]L. C. Sutherland, J. E. Piercy, H. E. Bass, and L. B. Evans, "Method for Calculating the Absorption of Sound by the Atmosphere," J. Acoust. Soc. Am. 56, S1(A) (1974). This method with a summary of the theoretical and experimental background is due to be published in the series *Physical Acoustics*, edited by W. P. Mason (Academic, New York, to be published).
[24]L. C. Sutherland, "Review of Experimental Data in Support of a Proposed New Method for Computing Atmospheric Absorption Loss," DOT Report TST-75-87 (1975).
[25]F. D. Shields and H. E. Bass, "A Study of Atmospheric Absorption of High Frequency Noise and Application to Fractional-Octave Bands," NASA report CR 2760 (1976).
[26]E. N. Bazley, "Sound Absorption in Air at Frequencies up to 100 kHz," NPL (Great Britain) Report AC74 (1976).
[27]M. C. Lee and L. C. Sutherland, "Laboratory measurements of atmospheric absorption in air," J. Acoust. Soc. Am. 60, S31(A) (1976).
[28]K. V. Mackenzie, "Reflection of Sound from Coastal Bottoms," J. Acoust. Soc. Am. 32, 221–231 (1960).
[29]C. B. Officer, *Introduction to the Theory of Sound Transmission with Application to the Ocean*" (McGraw–Hill, New York, 1958).
[30]S. P. Pao and L. B. Evans, "Sound Attenuation over Simulated Ground Cover," J. Acoust. Soc. Am. 49, 1069–1075 (1971).
[31]For a good discussion of locally reacting surfaces see P. M. Morse, *Vibration and Sound* (McGraw–Hill, New York, 1958).
[32]A. Banõs, *Dipole Radiation in the Presence of a Conducting Half-Space* (Pergamon, New York, 1966).
[33]A. N. Sommerfeld, "Propagation of Waves in Wireless Telegraphy," Ann. Phys. (Paris) 28, 665 (1909); 81, 1135–53 (1926).
[34]J. R. Wait, *Electromagnetic Waves in Stratified Media* (Pergamon, New York, 1970), Chap. 2.
[35]I. Rudnick, "Propagation of an Acoustic Wave Along a Boundary," J. Acoust. Soc. Am. 19, 348–356 (1947).
[36]A. R. Wenzel, "Propagation of Waves along an Impedance Boundary," J. Acoust. Soc. Am. 55, 956–963 (1974).
[37]R. J. Donato, "Propagation of a spherical wave near a plane boundary with complex impedance," J. Acoust. Soc. Am. 60, 34–39 (1976).
[38]T. F. W. Embleton, J. E. Piercy, and N. Olson, "Outdoor Propagation over ground of finite impedance," J. Acoust. Soc. Am. 59, 267–277 (1976).
[39]R. B. Lawhead and I. Rudnick, "Acoustic Wave Propagation Along a Constant Normal Impedance Boundary," J. Acoust. Soc. Am. 23, 546–549 (1951).
[40]K. U. Ingard, "On the Reflection of a Spherical Wave from an Infinite Plane," J. Acoust. Soc. Am. 23, 329–335 (1951).
[41]J. E. Piercy, T. F. W. Embleton, and N. Olson, "Mechanisms Causing Variability in the Noise Testing of Light Motor Vehicles," National Research Council of Canada Report APS 560 (1976).
[42]P. J. Dickinson and P. E. Doak, "Measurements of the Normal Acoustic Impedance of Ground Surfaces," J. Sound Vib. 13, 309–322 (1970).
[43]H. G. Jonasson, "Sound Reduction by Barriers on the Ground," J. Sound Vib. 22, 113–126 (1972).
[44]D. Aylor, "Noise Reduction by Vegetation and Ground," J. Acoust. Soc. Am. 51, 201–209 (1972).
[45]P. H. Parkin and W. E. Scholes, "The Horizontal Propagation of Sound from a Jet Engine Close to the Ground at Hatfield," J. Sound Vib. 2, 353–374 (1965).
[46]R. A. Bettis and M. Z. Saxton, "Effect of Test-Site Topography on Vehicle Noise Measurements," J. Acoust. Soc. Am. 54, 332(A) (1973).
[47]K. M. Ivanov-Shits and F. V. Rozhim, "Investigation of Surface Waves in Air," Sov. Phys.–Acoust. 5, 510–512 (1960).
[48]J. E. Piercy, R. J. Donato, and T. F. W. Embleton, "Near-horizontal propagation of sound over grassland," J. Acoust. Soc. Am. 60, S2(A) (1976).
[49]J. E. Piercy and T. F. W. Embleton, "Effect of Ground on Near-horizontal Sound Propagation," Trans. Soc. Auto. Eng., Sec. I 83, 928–932 (1974).
[50]M. E. Delany and E. N. Bazley, "Monopole Radiation in the Presence of an Absorbing Plane," J. Sound Vib. 13, 269–279 (1970).
[51]M. E. Delany and E. N. Bazley, "A Note on the Effect of Ground Absorption in the Measurement of Aircraft Noise," J. Sound Vib. 16, 315–322 (1971).

[52]H. Wise, "Asymptotic Dipole Radiation Formulas," Bell System Tech. J. 8, 662 (1929).
[53]F. M. Wiener and D. N. Keast, Bolt, Beranek and Newman Final Report 466, (June 1957) (AD-136319).
[54]J. L. Lumley and H. A. Panofsky, *The Structure of Atmospheric Turbulence*, (Wiley, New York, (1964).
[55]O. G. Sutton, *Micrometeorology* (McGraw–Hill, New York, 1953).
[56]R. E. Munn, *Descriptive Micrometeorology* (Academic, New York, 1966).
[57]F. M. Wiener and D. N. Keast, "Experimental Study of the Propagation of Sound over Ground," J. Acoust. Soc. Am. 31, 724–733 (1959).
[58]M. E. Delany, "Range Prediction for Siren Sources," NPL Aero Special Report 033 (1969).
[59]P. H. Parkin and W. E. Scholes, "The Horizontal Propagation of Sound from a Jet Engine close to the Ground at Radlett," J. Acoust. Vib. I, 1–13 (1964).
[60]W. E. Scholes and P. H. Parkin, "The Effect of Small Changes in Source Height on the Propagation of Sound over Grassland," J. Sound Vib. 6, 424–442 (1967).
[61]W. E. Scholes, A. C. Salvidge, and J. W. Sargent, "Field Performance of a Noise Barrier," J. Sound Vib. 16, 627–642 (1971).
[62]E. A. Dean, "The Wavelength Limits of Atmospheric Ray Acoustics," Proc. Symp. Acoust. Propag. Texas, 1961, p. 7 (AD-408716).
[63]P. Baron, "Propagation du son dans l'atmosphère et audibilité des signaux avertisseurs dans le bruit ambient," Ann. Telecomm. 9, 258–270 (1954).
[64]H. Goydke, W. Kallenbach, and J. J. Schroeder, "Untersuchungen zur Shallausbreitung von Sirenensignalen," Acustica 20, 276–288 (1968).
[65]P. A. Franken and D. E. Bishop, "The Propagation of Sound from Airport Ground Operations," NASA Report CR767 (1967).
[66]A. R. Kriebel, "Refraction and Attenuation of Sound by Wind and Temperature Profiles over a Ground Plane," J. Acoust. Soc. Am. 51, 19–23 (1972).
[67]F. Ingerslev and C. Svane, Proc. 6th Int. Cong. Acoust. F93–96 (1968).
[68]I. A. Dneprovskaya, V. K. Joffe, and F. I. Levitas, Sov. Phys.–Acoust. 8, 235–239 (1963).
[69]T. F. W. Embleton, G. J. Thiessen, and J. E. Piercy, "Propagation in an inversion and reflection at the ground," J. Acoust. Soc. Am. 59, 278–282 (1976).
[70]R. C. Polly and R. N. Tedrick, "Measured Acoustic Propagation Parameters in the Mississippi Test Operations Area," NASA MSFC, TMX-1132 (August 1965).
[71]V. I. Tatarski, *The Effects of the Turbulent Amosphere on Wave Propagation* (Keter, Jerusalem, 1971).
[72]P. A. Mandics, "Line-of-Sight Acoustical Probing of Atmospheric Turbulence," Tech. Rep. 4502-1 Radio Science Lab., Stanford University, 92PP (1971).
[73]C. G. Little, "Acoustic Methods for the Remote Probing of the Lower Atmosphere," Proc. IEEE 57, 571–578 (1969).
[74]K. U. Ingard and G. C. Maling, "On the Effect of Atmospheric Turbulence on Sound Propagation over the Ground," J. Acoust. Soc. Am. 35, 1056–1058 (1963).
[75]T. F. W. Embleton, N. Olson, J. E. Piercy, and D. Rollin, "Fluctuations in the Propagation of Sound Near the Ground," J. Acoust. Soc. Am. 55, 485(A) (1974).
[76]T. F. W. Embleton and J. E. Piercy, "Relation between steady and normally distributed sound levels," J. Acoust. Soc. Am. 58, S108 (A) (1975).
[77]P. K. Kasper, R. S. Pappa, L. R. Keefe, and L. C. Sutherland, "A Study of Air-to-Ground Sound Propagation Using an Instrumented Meteorological Tower," NASA Report CR-2617, (1975).
[78]But note for an opposite view R. Deloach, NASA Tech. Note D7823 (1975).
[79]E. H. Brown and S. F. Clifford, "On the attenuation of sound by turbulence," J. Acoust. Soc. Am. 60, 788–794 (1976).
[80]L. C. Sutherland, "Scattering Attenuation of Sound in the Lower Atmosphere," J. Acoust. Soc. Am. 49, 129(A) (1971).
[81]L. P. Delsasso, "The Attenuation of Sound in the Atmosphere," USAF contract W-28-099-ac-228 (AD89256) (1953).
[82]R. W. Benson *et al.*, "Sound Propagation Near the Earth's Surface as Influenced by Weather Conditions," WADC Tech. Rep. 57-353, Pt. I (1958), Pt. II (1960), Pt. III (1961).
[83]J. E. Piercy and T. F. W. Embleton, "Effect of Weather and Topography on the Propagation of Noise—Vancouver Airport," NRC (Canada) Report APS 526.
[84]C. F. Eyring, "Jungle Acoustics," J. Acoust. Soc. Am. 18, 257–270 (1946).
[85]T. F. W. Embleton, "Sound Propagation in Homogenous Deciduous and Evergreen Woods," J. Acoust. Soc. Am. 35, 1119–1125 (1963).
[86]D. Aylor, "Sound Transmission Through Vegetation in Relation to Leaf Area Density, Leaf Width, and Breadth of Canopy," J. Acoust. Soc. Am. 51, 411–414 (1972).
[87]D. A. Dobbins *et al.*, "Jungle Acoustics I: Transmission and Audibility of Sounds in the Jungle," U. S. Army Tropic Test Center Report 7 (1966) (AD 647 804).
[88]F. M. Wiener, C. I. Malme, and C. M. Gogos, "Sound Propagation in Urban Areas," J. Acoust. Soc. Am. 37, 738–747 (1965).
[89]J. F. Hemdal *et al.*, "A Study of Repeatability at Motor Vehicle Noise Measurement Sites," Environmental Res. Inst., University of Michigan (1974).
[90]R. M. LaBreche, "Variation in Measured Automobile Passby Noise due to Operator and Environmental Factors," G. M. Proving Ground Eng. Publ. 4587 (1974).
[91]Society of Automotive Engineers, "Acoustic Effects Produced by a Reflecting Plane," Aerospace Information Report AIR 1327 (15 January, 1976).
[92]See, for example, Refs. 67 and 83.
[93]P. J. Dickinson, "Temperature Inversion Effects on Aircraft Noise Propagation," J. Sound Vib. 47, 438–443 (1976).
[94]F. D. King and C. C. Oliver, "Effects of wind and temperature gradients on the propagation of highway traffic noise," J. Acoust. Soc. Am. 57, S32(A) (1975).
[95]J. J. Hajek, "Ontario Highway Noise Prediction Method," Ministry of Transport (Ont.) Report RR197 (1975).
[96]K. S. Plotkin and R. Kunicki, "Comparison of Highway Noise Prediction Models," Wyle Research Report WR76-25 (1976).

Physical Acoustics Summer School
21-28 June 1996

Sensor Physics: Signals and Noise

Thomas B. Gabrielson
Code 4554 MS 07
NAWC Aircraft Division
Warminster, PA 18974-0591

ABSTRACT

Various techniques for converting acoustic or vibration signals to electronic or optical signals (and vice versa) are treated in detail in many papers and books. Frequently, a superficial understanding of the conversion processes is all that is required; however, there is a depth to the physics of transduction that can only be appreciated by closer examination. While the physics of electromechanical signal conversion is rich enough, there are two roads less travelled but equally rewarding: sensor noise analysis and sensor calibration.

Because the subject of sensor self-noise integrates concepts in mechanics, solid-state physics, quantum mechanics, and thermodynamics, the study of self-noise would be fascinating even if there were no direct application of the principles. For high-performance sensors, however, the achievable noise floor is a critical parameter and so an understanding of sensor self-noise is essential for design, construction, and application of such sensors.

Equilibrium thermal noise is an inevitable consequence of the Second Law of Thermodynamics and, as such, is perhaps the most fundamental source of noise in any system. Equilibrium thermal noise is the noise produced by the normal thermal motion of the molecules that make up the sensor structure or the surrounding liquid or gas. It has a long history of both theoretical and experimental study and so it is remarkable that this mechanism is frequently ignored in the design and analysis of new sensor types. Presumably the reason for this omission is that there are many conventional sensors for which the limiting noise is some other mechanism and the familiarity with this other mechanism has led to the assumption that it is the only important mechanism. This belief does not cause problems until a significant technology shift is introduced and the relative importance of different noise mechanisms changes.

A critical aspect of equilibrium thermal noise is that *every resistance-like term contributes a fluctuating force* in accordance with Nyquist's Theorem. This includes fluid damping, structural damping, acoustic radiation resistance, an acoustic or mechanical load having a real component in the impedance, or thermal radiation. Any path that permits energy transfer from the sensor to the environment, whether it involves ordered motion (*e.g.*, radiation) or disordered motion (*e.g.*, viscous damping), introduces a fluctuating force.

Sometimes this noise component is called Brownian noise; this can lead to the impression that its sole source is the collision of gas or liquid molecules with the moving element. Even if the element were operated in a vacuum, though, there would still be an equilibrium thermal noise associated with damping in the structure. Frequently, "thermal noise" is cited as a component of the ambient noise (particularly in descriptions of underwater ambient noise). This component is completely accounted by associating the proper fluctuation force with the radiation resistance of the transducer. This component should not be assumed to contain all "thermal noise" fluctuations; it only accounts for those fluctuations associated with the radiation path.

Equilibrium-thermal noise is not only observed in miniature sensors. It can be important in any system designed for ultimate detection performance. At one extreme is the advanced Laser Interferometric Gravitational-Wave Observatory (LIGO), with masses expected to be 10 000 kilograms on suspensions with Q's from 10^7 to 10^9. Because of the extremely small signals anticipated, molecular agitation of those large masses is still a significant source of noise. At the other extreme, the 50 picogram "proof mass" in the geotactic protozoan *Loxodes* is just large enough to permit distinguishing of up from down in the background of equilibrium noise.

Because equilibrium thermal noise is a consequence of classical thermodynamics, its calculation does not depend on any particular model of the physical process. The noise that results from molecular bombardment of a suspended mass in thermal equilibrium with a fluid can be described with complete accuracy by continuum models from fluid dynamics (in terms of the mechanical damping resulting from viscosity in the fluid, or in terms of the continuum radiation resistance into the surrounding fluid, for example). There is no need to accurately describe the microscopic behavior of the fluid. In contrast, *shot noise* is critically dependent on two specific aspects of a particular physical model. First, the carriers (molecules, electrons, photons) must be discrete objects and, second, these discrete objects must act independently: the action of any one carrier must not depend on the presence or absence of any other carrier. Low levels of electron emission from the cathode of a vacuum tube, low levels of current flow across a reverse-biased semiconductor junction, current flows in nondegenerate semiconductors, photoelectric excitation of photodiodes, and tunneling of electrons represent instances in which the carriers can be considered to be acting independently. Molecular impacts at normal gas pressures (because of the very small mean free path) and electron flow in metals (because of the interdependence forced by the exclusion principle) are situations in which there is a very high degree of dependence between carriers; applying simple shot-noise analyses to these phenomena leads to substantial overestimation of the noise.

Shot noise can be an equilibrium phenomenon or a nonequilibrium phenomenon. In equilibrium shot noise, the calculated noise will be identical to that obtained by an equilibrium thermal noise analysis. For example, a piece of semiconductor with no applied voltage (hence, no *measurable* average current) exhibits a noise voltage that can be calculated in two ways: (1) by considering the Johnson noise associated with the semiconductor's resistance, or (2) by considering two equal (on average) but oppositely directed random currents related to the independent (thermally excited) motion of the carriers and then calculating the root-mean-square value of the voltage produced by these two noise currents. Often, there is no *observable* average current that can be associated with equilibrium shot noise. (An important exception to this is the case of pressure fluctuations in the free-molecular regime; here, the noise is directly related

to the observable pressure.) In the equilibrium case, the equilibrium-thermal analysis is fundamental; equilibrium shot noise is a special case and must, in any event, lead to the same answer, not to an additional component.

For current in a semiconductor, application of an external voltage forces the system away from equilibrium, and the forward current (applied plus random) dominates over the reverse current (random). The current noise expression rapidly approaches the standard shot-noise expression in terms of the measurable current for applied voltages greater than k_BT/q (about 25 mV at room temperature). This is a classic example of nonequilibrium shot noise.

Since the voltage (25 Mv) beyond which shot noise would dominate over Johnson noise is relatively small, serious errors can be introduced by ascribing a shot-noise component to a material with an inherently high degree of dependence in the carrier flows. Devices such as inductors, capacitors, resistors (carbon, carbon-film, metal-film, wirewound), and degenerate (very highly doped) semiconductors have well-populated conduction bands. Since no two electrons can occupy the same spin-state (the exclusion principle) and most electrons in a well-populated band will be surrounded by occupied states, only those few electrons near the band edge are capable of moving to vacant states. Therefore, only a very small fraction of the conduction electrons exhibit fluctuations. The resultant noise can be much lower than that predicted by the usual shot-noise expression.

Currents associated with potential jumps (currents through PN junctions, gate leakage currents in FETs, tunneling currents) or currents composed of carriers that are thermally excited out of their "rest" bands into sparsely populated conduction bands (currents in intrinsic or normally doped semiconductors) will exhibit full shot noise. In these cases, there are plenty of unoccupied states near the occupied ones and the carriers can act independently even if they are electrons obeying the exclusion principle.

If a system is in thermal equilibrium, then its noise is completely and accurately described by equilibrium thermal noise (mechanical, electrical, or otherwise). Once a system is forced away from equilibrium, the noise frequently increases over the equilibrium value. In addition to non-equilibrium shot noise, another common form of this excess noise is *1/f noise*, so called because its power spectrum goes at least roughly f^{-1} (f being frequency). There is no satisfactory unifying theory for 1/f noise as there is for equilibrium thermal noise or shot noise. Besides the spectral dependence, one frequently observed characteristic of 1/f noise is that the power spectrum of that noise is proportional to the applied power. While processes are observed in which a 1/f behavior is observed to extremely low frequency, if the noise power spectrum were 1/f all the way to zero frequency, the total noise power would be infinite. Consequently, a universal model for 1/f noise must not only account for the 1/f spectral shape, it must also predict a very-low-frequency rolloff. Unfortunately, since the power is concentrated at the low frequencies, small errors in the location and nature of the rolloff can lead to large errors in the predicted magnitude of the noise. In addition, the rolloff itself is poorly understood because of the difficulty of making measurements at extremely low frequencies.

As interesting as it is, noise is not the whole story: in general, the sensor's response to signal must be determined as well. Most transducer calibrations are done by comparison with a

reference transducer having a known response but it is particularly useful to understand some more direct methods of calibration. (For example, how are the reference transducers calibrated?) Crude calibrations can sometimes be performed by using the equilibrium thermal noise of a sensor (since that is related absolutely to temperature) but rarely to better than 10 percent accuracy and then only when the relevant part of the noise spectrum is dominated by equilibrium thermal noise.

Reciprocity is an especially powerful technique for determining the response of a transducer. While often considered to be arcane, reciprocity is actually a relatively straightforward method by which difficult-to-measure quantities such as forces or velocities can be translated into easier-to-measure quantities such as masses or voltage ratios. Another way of considering reciprocity calibration is to view the transducer as a two-port device, one port of which is an electrical input or output, the other port of which is a mechanical input or output. Reciprocity provides the framework for sets of measurements using several transducers in such a way that only quantities at the electrical ports of the transducers need be made.

Although many arrangements of transducers can be contrived for reciprocity calibration, the classic arrangement consists of three transducers: a source, a receiver, and a transducer that is reciprocal (and is used as both source and receiver). Three sets of measurements of drive current and received voltage are made: source to receiver, source to reciprocal transducer, and reciprocal transducer to receiver. This gives the necessary data to set up three equations in the four unknowns (the receiving responses of the receiver and the reciprocal transducer, and the transmitting responses of the source and the reciprocal transducer). The fourth equation is obtained by applying reciprocity to find a relationship between the transmitting and receiving responses of the reciprocal transducer (normally in terms of the transfer impedance of the medium to which the transducers are attached or in which the transducers are immersed).

KEY TO THE INCLUDED REFERENCES

Usher, Buckner, and Burch, "A miniature wideband horizontal-component feedback seismometer," J. Phys. E: Sci. Instrum. 10, 1253, 1977. The Introduction is the important section from the point of view of fundamentals. But, if you have the time, read the entire paper. This is an excellent example of a well-designed sensor; several of the highest performance miniature accelerometers available today are based closely on this design.

Rudnick, "Unconventional reciprocity calibration of transducers," J. Acoust. Soc. Am. 63, 1923, 1978. Don't be misled by the title. This is an excellent introduction to the application of reciprocity to the calibration of transducers.

Uhlenbeck and Goudsmit, "A problem in Brownian motion," Phys. Rev. **34**, 145, 1929. Don't get bogged down in the math. Read the text and find the thought processes and the conclusions. Footnote number 5 is very important. Take note of the importance of independence in applying the methods they use.

A miniature wideband horizontal-component feedback seismometer

M J Usher†, I W Buckner‡ and R F Burch§
† Department of Cybernetics, University of Reading, Whiteknights, Reading RG6 2AL, UK
‡ Garrard Ltd, Swindon, UK (formerly of Department of Cybernetics, University of Reading)
§ Blacknest Seismological Centre, AWRE, Aldermaston, Reading RG7 4PR, UK

Received 1 February 1977, in final form 9 June 1977

Abstract A miniature, wideband, horizontal-component feedback seismometer has been developed and compared with conventional seismometers. The instrument employs an inverted pendulum of mass 0·04 kg supported with a natural period of 0·6 s and a Q factor of about 20. Displacements of the mass are sensed by a differential capacitive transducer and feedback is applied via an electromagnetic system to maintain the mass stationary with respect to its supports. The instrument has a response defined by feedback from DC to 100 Hz and has a Brownian noise level of 5×10^{-10} m s^{-2}. Overall dimensions, including an evacuated jacket, are 15 cm × 10 cm diameter making it suitable for borehole applications. The instrument has been operated for several weeks with conventional long-period seismometers and has been found to compare satisfactorily.

1 Introduction

The spectrum of earth motion of interest in seismology covers the range from about 0·01 to 10 Hz. A large peak known as microseismic noise occurs at about 0·15 Hz. It is due to effects on the earth's surface and seismometers have conventionally been designed to avoid it, short-period instruments operating over the range 1–10 Hz and long-period instruments from 0·1 to 0·01 Hz.

Seismometers measure the relative displacement x_r between a suitably supported mass and the instrument frame (assumed to follow the required ground motion). The equation of motion of the mass M when excited by a sinusoidal ground acceleration $\ddot{x}_i$ is

$$\ddot{x}_r + \frac{R}{M}\dot{x}_r + \frac{1}{MC}x_r = \ddot{x}_i \qquad (1)$$

where R represents the viscous damping resistance and C is the compliance of the supporting spring. The natural angular frequency is $\omega_0^2 = 1/MC$. It can be seen that for excitation frequencies $\omega < \omega_0$ the device behaves as an accelerometer with $x_r = \ddot{x}_i/\omega_0^2$, whereas for high excitation frequencies it becomes a displacement meter with $x_r = x_i$. Conventional seismometers have employed suspensions with natural periods of the order of 1–3 s for short-period instruments, and of the order of 10–30 s for long-period instruments in order to obtain the maximum response from the mechanical system. Various types of 'standard responses' have been adopted, often based on existing instruments, obtained by feeding the output to a suitable filter or by adjusting the period and damping of the instrument.

The fundamental limit to the detection of ground motion by a seismometer is set by the Brownian motion of the mass. It can be shown (Usher 1973) that the noise-equivalent acceleration $(\ddot{x}_i)_{ne}$ for a bandwidth Δf is given by

$$(\ddot{x}_i)_{ne}^2 = \frac{4RkT\Delta f}{M^2} = \frac{4kT}{M}\frac{\omega_0}{Q} \qquad (2)$$

where kT is the equipartition energy and Q the quality factor of the suspension. A small mass may be used provided that the damping is low, though seismometers have conventionally employed large masses of several kilograms, usually nearly critically damped. A mass of 0·01 kg with a natural period of 1 s and a Q of 100 has $(\ddot{x}_i)_{ne} \approx 3 \times 10^{-10}$ m s^{-2} in a bandwidth of 1 Hz at room temperature, which is of the order required in practice (see later).

The application of negative force-feedback to a seismometer produces a number of advantages and is in fact necessary when a small mass is suspended with a high Q, in order to achieve a satisfactory transient response. Feedback may affect any of the three terms of the left-hand side of equation (1). The most useful form is negative displacement feedback, which tends to keep the mass fixed in position with respect to its supports and affects only the term in x_r, making the suspension appear more stiff and increasing the natural frequency. A general block diagram of such a seismometer system is

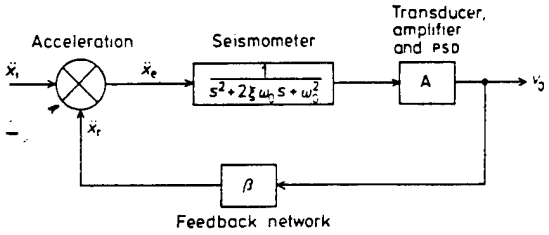


Figure 1 A force-feedback seismometer.

shown in figure 1, in which the transfer function of the seismometer itself is represented as

$$\frac{x_r}{\ddot{x}_i} = \frac{1}{s^2 + 2\xi\omega_0 s + \omega_0^2}$$

where $s = j\omega$ is the Laplace operator and ξ is the damping ratio. The closed-loop transfer function is

$$\frac{v_0}{\ddot{x}_i} = \frac{A}{s^2 + 2\xi\omega_0 s + (\omega_0^2 + A\beta)}$$

where v_0 is the output voltage, A is the gain in the forward path and β is the transfer function of the feedback path, and becomes equal to $1/\beta$ when $A\beta$ is the dominant term.

If it is assumed that $A\beta$ is independent of frequency, the DC loop gain L is $(1/\omega_0^2)A\beta$ and the natural frequency is increased by a factor $L^{1/2}$, the damping being reduced by the same factor. The response is essentially flat (to acceleration) from DC to the new natural frequency and the transient response

7943

can be controlled by a compensation network in the feedback path.

A small mass of, say, 0·01 kg, suspended with a Q of 100 and employed in a feedback system of suitable loop gain (say 100), can thus provide a flat response and adequate detectivity over the whole range of interest in seismology. In addition, advantages over conventional open-loop instruments are obtained in linearity, dynamic range and calibration.

Mechanical design requirements are eased and the desired wideband response is simply determined by the feedback parameters. The signal-to-noise ratio is unaffected by feedback. The response is controlled by applying forces to the mass: this does not affect its Brownian motion, whereas adding damping to control the response in an open-loop system increases the dissipation and therefore increases the Brownian motion.

2 Review of developments in seismometry

The earliest seismometers employed very large masses, sometimes of several tonnes, in order to achieve a sufficient momentum of the mass to drive the recording equipment. Conventional seismometers typically have masses of 10 kg with periods of 10–30 s (Geotech S12 long-period seismometer) or 1 kg with periods of 1–3 s (Willmore Mk III short-period seismometer), and employ an electrodynamic transducer for sensing the position of the mass with near-critical damping.

A few miniature instruments have been developed, notably by Block and Moore (1970), Block and Dratler (1972) and by Jones (1967). Jones' instrument was a pendulum tiltmeter employing a mass of about 0·02 kg and a differential capacitance displacement transducer. It was operated open-loop with a natural period of about 0·3 s and had relatively high damping; while successful for recording very long-period (e.g. tidal) motions it was not designed for the range of interest in seismology and had a theoretical Brownian noise of about 5×10^{-9} m s^{-2}. The instruments developed by Block *et al* were miniature wideband seismometers in which a mass of 0·01–0·02 kg was supported by a quartz fibre in torsion with a natural period of about 1 s. A differential capacitance displacement transducer was used to sense the mass displacement. Although an electrostatic feedback system was proposed the instrument was operated open-loop with a fairly high Q (about 20) so that the transient response was not satisfactory and the instrument was not suitable for measurements at short periods without special filtering. The temperature coefficient for torsion of the quartz suspension resulted in a very high temperature stability requirement and the instrument was enclosed in two evacuated containers, its overall size being similar to that of conventional instruments.

A number of seismometers employing feedback have been described in the literature. Tucker (1958) applied force-feedback to a pendulum seismometer, whose natural period was about 1 s, by means of an inductive displacement transducer and electromagnetic force transducer. The loop gain used was only about 3 so that the response was not completely determined by feedback; the instrument was intended for the study of microseisms. Instruments described by de Bremaeker *et al* (1962) and by Sutton and Latham (1964) employed feedback at very long periods, but not in the seismic range. Block and Moore (1966) applied feedback to a conventional La Coste-Romberg instrument via a capacitive displacement transducer and electrostatic force transducer, but the forces available with this transducer are very small and the response in the range of interest was not controlled by feedback. As mentioned above, Dratler (1971) describes an electrostatic feedback system for a miniature seismometer, but the system was currently not used in practice. Systems have been described by Willmore (1959), Russell *et al* (1968) and Kolez-nokov *et al* (1975, private communication) in which short-period Willmore seismometers have been modified by force-feedback, using the existing electrodynamic transducer for sensing the mass motion. In these instruments feedback modifies the response and produces advantages in linearity and calibration, but the detectivity is the same as for an open-loop, short-period instrument. Melton (1976) has discussed force-feedback seismometers but the instruments described employ relatively large masses (about 0·4 kg).

It can thus be seen that a miniature wideband feedback instrument has not previously been developed. The miniature open-loop instruments described above either are long-period devices with inadequate detectivity in the seismic range or have large overall size. The feedback instruments described in the literature mostly have low loop gain in the range of interest or are not miniature devices. We have been unable to find a reference reporting the use of substantial feedback over the seismic range accurately defining the instrument response and controlling the transient behaviour, with its attendant advantages of linearity, dynamic range and calibration.

3 Design philosophy of miniature wideband seismometers

The signal levels required to be detected can be deduced from the spectrum of background seismic noise. This has been investigated by many workers, notably Brune and Oliver (1959), Fix (1972) and Savino *et al* (1972). Figure 2 is based on the work of Fix, and shows acceleration power densities at a very quiet site (Queen Creek). The minimum observed acceleration power occurs at a period of 30–50 s and has a value of 10^{-19} (m s^{-2})^2 Hz^{-1}.

To detect a signal of this magnitude, the instrumental noise, determined by Brownian motion of the mass and transducer/amplifier noise, must be sufficiently small. The Brownian motion of the mass sets the fundamental limit to detection of signals, and the required mass/damping can be deduced from equation (2). A mass of 0·04 kg with a Q factor of 50 has a Brownian noise-equivalent acceleration very close to this value, and is shown as a horizontal line in figure 2 together with the Brownian noise levels for other combinations of mass and damping.

The transducer/amplifier noise should be designed to be less than the Brownian noise, in order not to degrade the detection limit. A suitable transducer is the differential capacitive type, which is essentially noiseless. The electronic noise is largely determined by the following amplifier and can be represented by a series noise-equivalent generator R_{nv}. The noise-equivalent acceleration can be shown to be

$$(\ddot{x}_i)_{ne} = \frac{(s^2 + 2\xi\omega_0 s + \omega_0^2)}{r} (4\, R_{nv} k T \Delta f)^{1/2} \qquad (3)$$

$$= \frac{\omega_0^2}{r} (4\, R_{nv} k T \Delta f)^{1/2} \text{ for } \omega < \omega_0$$

where r is the responsivity of the transducer. The natural period T_0 ($=2\pi/\omega_0$) and responsivity r must be chosen such that $(\ddot{x}_i)^2_{ne} < 10^{-19}$(m s^{-2})^2 Hz^{-1} with the practical value of R_{nv}.

After the required minimum signal level has been chosen and the Brownian noise has been made of the same order as, and the electronic noise made less than this level, it is only necessary to make the noise from other sources (such as the phase-sensitive defector (PSD), filters, etc) negligible. This can be done by ensuring that there is sufficient gain prior to each element.

Finally, to achieve the desired wideband response of defined

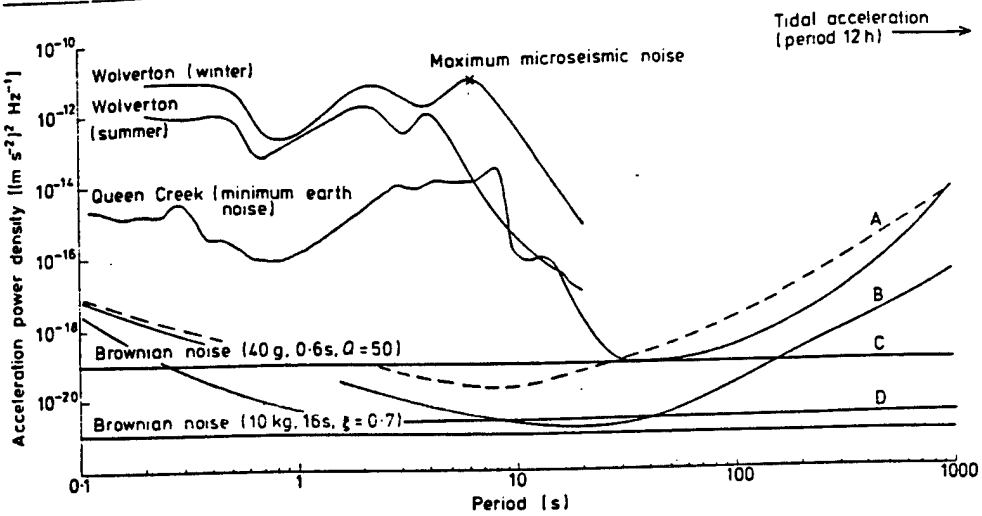


Figure 2 Spectrum of earth noise (after Fix). Acceleration power densities for Queen Creek (USA) and Wolverton (UK) are compared with the noise-equivalent accelerations of a standard seismometer (curves A and B) and with the Brownian and electronic noise levels of a miniature instrument (curves C and D). A, 10 kg mass with standard amplifier; B, 10 kg mass with special amplifier; C, Brownian noise of feedback seismometer; D, electronic noise of feedback seismometer.

magnitude and controlled damping, negative force-feedback is applied to maintain the mass position fixed with respect to the instrument frame.

4 Description of the instrument

4.1 *Mechanical system*

The basic requirements of the suspension system of the instrument are that the pendulum be constrained to move with a single degree of freedom and with a suitable fundamental period and Q factor. All other modes must be far removed from the frequencies of interest, and above the feedback loop cut-off frequency, and the long-term stability must be satisfactory.

It can be seen from equation (3) that the natural period T_0 must be large enough to make the amplifier noise negligible. A natural period of about 1 s was achieved by using an inverted pendulum supported by simple spring strips. These produce a smaller restoring torque than cross-spring pivots and can easily be made from a material of low temperature coefficient such as Ni-span D. Our earlier investigations of period-lengthening devices showed that these become very critical when used with small masses, and the arrangement used is much more compact and robust.

The transducer must have high responsivity and low noise, combined with adequate long-term stability. A differential capacitance transducer is excellent in these respects (Jones and Richards 1973), providing a higher responsivity than other available transducers such as linear variable differential transformers (LVDTs), and a high detectivity. The AC operation of such a transducer avoids the $1/f$ noise region, which is a major problem in the design of (DC) amplifiers for conventional seismometers.

The basic instrument is shown in figures 3 and 4, with the heater cylinder and outer jacket removed. The pendulum has a mass of about 0·04 kg and is supported by two spring strips (of Ni-span D) clamped at both ends. The main frame is machined from a single piece of brass and supports the outer plates of the differential capacitor, insulated from it by quartz spacers. The pendulum itself is the central plate of the differential capacitor, insulated by quartz spacers, the signal to the

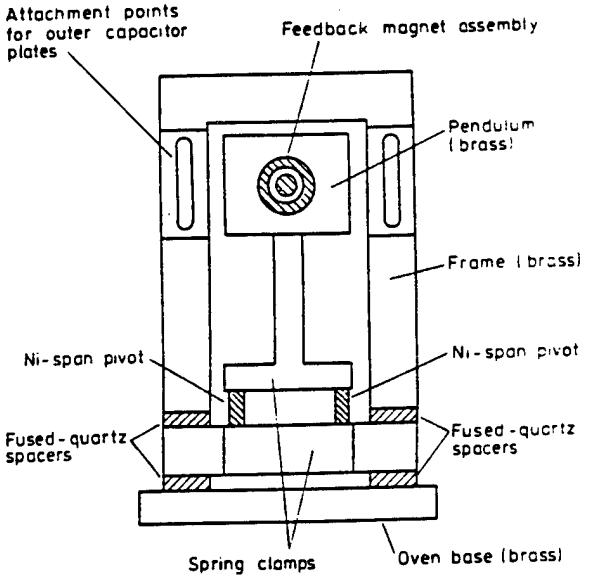


Figure 3 Diagram of seismometer with heater and cover removed.

preamplifier being taken via the tag shown. The components in the main frame assembly are kinematically mounted, permitting easy dismantling and assembly and improving stability. A high-stability magnet is attached to the centre of the pendulum bob, the coils of the force-feedback transducer being attached to the outer plates of the capacitor; this arrangement was found to be necessary to avoid a high-frequency instability in the closed-loop response when the feedback force did not coincide with the centre of mass. The thermal enclosure fits over the main-frame unit, with heater windings on the outside, and supports two electronic boxes containing the preamplifier/excitation and thermal control electronics, as shown in figure 5. The outer jacket fits over an O-ring seal and can be evacuated via a tap. It is attached to ar

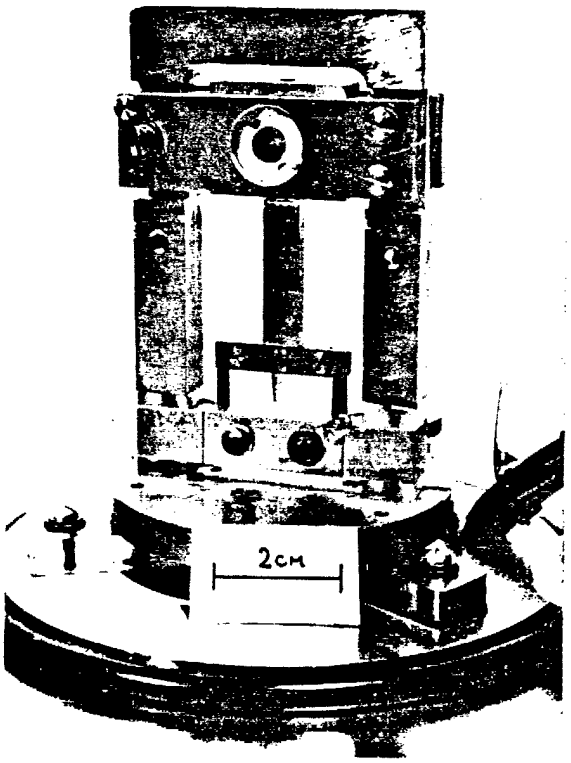
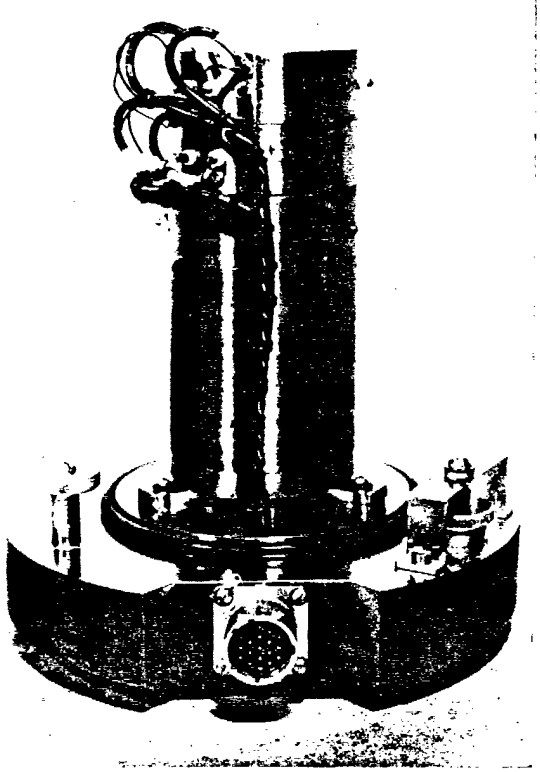


Figure 4 Photograph of seismometer with heater and cover removed.



5 Photograph of seismometer showing heater and electronics.

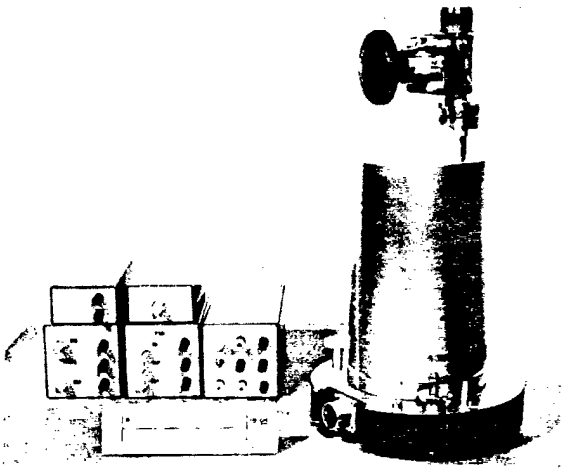


Figure 6 Photograph of complete instrument and electronics.

Invar base plate which contains a precision differential levelling arrangement and the socket for the electronics cable. The complete instrument and the associated electronics are shown in figure 6.

4.2 *Electronic system*

The essential requirement of the servo-system is that the response of the instrument be independent of the suspension and be defined only by the passive feedback elements over the whole range of operation from DC to 10 Hz. The loop gain should be about 100 over this range, and the response should preferably be flat to input acceleration or velocity and the damping near critical.

A block diagram of the complete system is shown in figure 7. The input ground acceleration $\ddot{x}_i$ produces a relative displacement x_r (between the mass and the frame) which is converted to a voltage v_r at the excitation frequency by the transducer and preamplifier and further amplified by the channel amplifier. This signal is demodulated by the PSD, and the feedback network drives the feedback coils to return the mass towards its original position.

The method outlined above (§1) can be implemented directly, but it is advantageous to introduce an integration into the loop to provide high gain at low frequencies. Mass displacements due to large tidal forces or thermal drift are thus reduced and the linearity of the transducer is improved. In addition, the design requirements of the PSD are eased, the necessary dynamic range is reduced and a greater $1/f$ noise level can be tolerated. If the time constant of the integration is made fairly large (e.g. 1 s) the behaviour of the system can be very similar to that of the simple proportional system, with a closed-loop resonance frequency of, say, 10 Hz. However, some difficulties were encountered with system stability in such an arrangement, due to poor recovery from overload and to stray mechanical resonances at high frequencies, and a modified arrangement was used in practice. The integration time constant was made 0·15 s producing a break in the response at 0·35 Hz; the response was then flat to acceleration at frequencies below 0·35 Hz and flat to velocity above this frequency. It is still defined by feedback, of course, and has the additional advantages that high-frequency ground motion does not produce an excessive output and that filtering to obtain the velocity response normally required is easier in some cases.

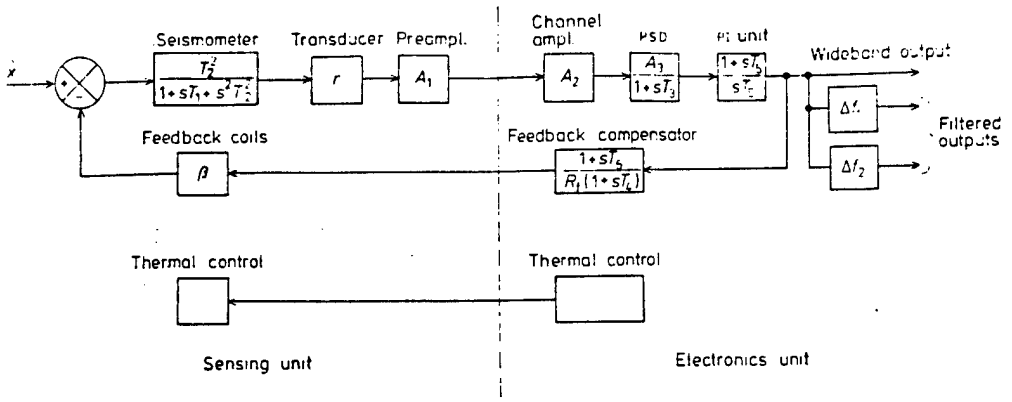


Figure 7 Block diagram of complete system. The input acceleration produces a relative displacement between sensing mass and instrument frame, which is detected by a displacement transducer, amplified and rectified, and fed back to maintain the mass stationary with respect to the frame.

4.2.1 *Capacitive transducer and preamplifier* The responsivity of a linear electrical displacement transducer is always of the form excitation voltage/range, and it is because the range can be made small that a high value of responsivity can be obtained. A responsivity of 10^4 V m^{-1} was obtained by applying 3 V RMS to the outer plates with a plate separation of 0·3 mm. A smaller separation leads to difficulty in levelling and increases the electrostatic forces on the mass, which must be kept small and constant. With the values used the electrostatic force has a value of 2×10^{-8} N when the centre plate is displaced by 10 μm (which just overloads the PSD on open loop).

The transformer arms and differential capacitor form a Blumlein bridge, of which the equivalent circuit is shown in figure 8. The series capacitance $C \approx 20$ pF and the stray capacitance $C_1 \approx 10$ pF. It can be seen that stray capacitances do not affect the balance point of the bridge. The resistor R_1 is mainly determined by the biasing resistors in the preamplifier, and has a value of about 10 MΩ.

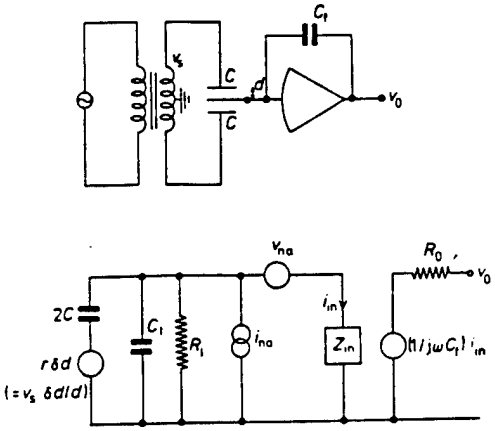


Figure 8 The Blumlein bridge and its equivalent circuit.

The charge amplifier configuration has the advantages that the signal level is not affected by changes in stray capacitance and that the output is independent of frequency over a wide range. The effective series noise-equivalent resistance referred to the signal source can be shown to be

$$R_n = R_{nv}\left[1+\left(\frac{C_1+C}{C}\right)+\frac{1}{\omega C R_1}\right]^2+\frac{1}{\omega^2 C^2}\left(\frac{1}{R_{ni}}+\frac{1}{R_1}\right)$$

where R_{nv} and R_{ni} are the noise-equivalent resistances of the input device, which was a low-noise FET. Using the values given above and assuming that $R_{nv} \approx 500\Omega$, $R_{ni} \approx 10$ MΩ, we find that the optimum noise performance occurs at a frequency of 100 kHz. The noise referred to the input is 5 nV RMS in a bandwidth of 1 Hz, corresponding to $R_n = 1{\cdot}5$ kΩ. This low value was possible because of the low value of C_1, due to the quartz spacers, and the high values of R_1 and R_{ni}. The value was confirmed experimentally. The voltage gain was 70.

The preamplifier was mounted close to the transducer to reduce cable capacitance. The channel amplifier A_2 is of conventional design, and was placed in the main electronics block, remote from the seismometer, together with the PSD feedback electronics, excitation oscillator and thermal control power amplifier.

4.2.2 *Drive oscillators* The stability of the excitation oscillator is important because of the very small displacements and forces involved. The least detectable acceleration of 3×10^{-10} ms^{-2} is equivalent to a force on the mass of about 10^{-11} N, and a high amplitude stability is therefore required to maintain the electrostatic forces sensibly constant. A Wien bridge oscillator was employed, in which the amplitude was controlled by a light-dependent resistor in the loop, the light level being derived by comparing the rectified output with a high-stability voltage reference. The amplitude stability was about 10 μV RMS over periods of up to 1 min and was found to be satisfactory.

4.2.3 *Phase-sensitive detector* The large dynamic range and low frequency of seismic signals make considerable demands on the properties of the PSD in terms of linearity, rejection of unwanted signals and noise level. The design employed was a complementary current-switching type similar to those developed by Faulkner and Grimbleby (1967), Danby (1968) and Grimbleby and Harding (1971). In order to achieve the necessary dynamic range and a good temperature coefficient the circuit employs a current mirror to produce a single-ended output (Buckner 1975). The output temperature coefficient was 0·001 % (equivalent to 1 ppm C^{-1}) and the noise about 10 μV RMS in a frequency band from 0·01 to 1 Hz.

COPYRIGHT
794323

4.2.4 *Proportional plus integral control unit* (PI) The PI unit comprises a high-quality operational amplifier providing an integration at frequencies up to 0·35 Hz and with a high-frequency gain of unity. Its noise and drift characteristics are important, since the wideband output voltage is taken directly from it, but are eased by the prior gain.

4.2.5 *Feedback compensator and feedback coils* The compensator is a passive, phase-lead circuit providing velocity-dependent feedback to stabilise the response. At medium frequencies the response is determined by a series resistor R_f in parallel with a capacitor C_f.

The feedback actuator comprises a high-stability magnet attached to the mass and small coils wound on formers attached to the centre of the outer capacitor plates. The acceleration coefficient was determined by applying a known tilt via a calibrated tilt table and measuring the current needed to reduce the output to zero (on open loop). The value of 1·41 m s^{-2} A^{-1} was constant over the full $\pm$10 V range of the output and no departure from linearity could be measured. Since mass movement on closed loop is negligible, linearity is required only between force and current. The maximum acceleration produced before overload at 10 V with R_f= 15 kΩ is approximately 10^{-3} m s^{-2}.

The stability of the magnet was very satisfactory, and no significant drift attributable to this was detected in six months' continuous operation of the instrument. The evacuated jacket used for environmental control was of mild steel and provided very effective magnetic shielding.

4.2.6 *Environmental control* The basic pendulum is enclosed in a temperature-controlled oven, which also encloses the preamplifier, transducer excitation transformer and temperature-sensing circuit: the whole assembly is enclosed in an evacuated case at a pressure of about 10 Pa. Evacuation was necessary to eliminate atmospheric and other similar effects, and to achieve a suitably high Q factor for the suspension.

Although the instrument was designed to be mechanically as symmetrical as possible to reduce the effects of thermal gradients, its measured temperature coefficient was 10^{-6} g °C^{-1}. A stability of 10^{-4} °C would be required to achieve a DC resolution of 10^{-10} g (the requirement in Block and Dratler's instrument (1972) was a stability of 10^{-6} °C). The thermal control system employs a type YSI 4001 thermistor in an AC bridge operated at 1 kHz, and a PSD similar to that described above feeds a modulator and AC power amplifier. The heater coil consists of bifilar-wound copper wire on a groove in the outside of the brass heater cylinder; AC power was necessary to avoid magnetic effects. The parameters of the control system were adjusted experimentally to give a satisfactory response to a step rise in temperature.

4.2.7 *Output filters* The wideband output from the instrument is proportional to ground acceleration from DC to 0·35 Hz and proportional to ground velocity from 0·35 to 100 Hz. Outputs from seismometers are usually required in the form of various agreed 'standard responses' and the wideband output was therefore fed to a filter unit outside the loop to achieve whatever overall response was required. The system design ensured that noise and drift due to the filters were negligible.

5 System response, calibration and noise level

With reference to figure 7, the acceleration $\ddot{x}_r$ produced by the feedback coil on open loop in response to an input acceleration $\ddot{x}_i$ is given by

$$\frac{\ddot{x}_r}{\ddot{x}_i} = \frac{K(1+sT_5)(1+sT_6)}{s(1+sT_1+s^2T^2_2)(1+sT_3)(1+sT_4)}$$

where

$$K = \frac{A_1A_2A_3r\beta}{\omega_0^2T_5R_f}, \qquad T_1 = \frac{2\xi}{\omega_0}, \qquad T_2^2 = \frac{1}{\omega_0^2};$$

r, A_1, A_2 and A_3 are the gains of the transducer, preamplifier, channel amplifier and PSD respectively, ω_0^2 is the natural angular frequency on open loop, β is the feedback force constant and R_f the series feedback resistor. T_1 and T_2 are time constants associated with the suspension, T_5 is the integration time constant, T_6 and T_4 refer to the phase-lead circuit and T_3 refers to the PSD. Time constants associated with the transducer, amplifier and feedback coil are less than 1 ms and are omitted. β is independent of frequency over the seismic range.

The open-loop and closed-loop responses are shown in figure 9, with A_1=70, A_2=5, A_3=5, $r=10^4$ V m^{-1}, β=1·4 m s^{-2} A^{-1}, ω_0=10 rad s^{-1}, R_f=15 kΩ, C_f=30 μF, T_1=0·025 s, T_2=0·1 s, T_3=0·001 s, T_4=0·001 s, T_5=0·15 s, T_6=0·45 s.

The open-loop gain has a value of over 40 dB over most of the range of interest (DC to 10 Hz) so that the response is accurately defined by feedback. The closed-loop response is flat to acceleration (of value $R_f/\beta = 10^4$ V (m s^{-2})$^{-1}$) from DC to 0·35 Hz, and is flat to velocity (of value $1/\beta C_6 = 2{\cdot}4\times10^5$ V(m s^{-1})$^{-1}$) from 0·35 Hz to the unity open-loop-gain frequency of 100 Hz.

The dynamic response of the instrument was measured by applying square or sinusoidal waveforms to one of the feedback coils, equivalent to an input acceleration, and the curve obtained agreed very closely with that of figure 9. The step response had a rise time of 0·1 s and an overshoot of about 20%, corresponding to ξ=0·7. The DC responsivity was measured using a calibrated tilt-table and agreed with the value above.

The theoretical noise-equivalent acceleration can be deduced from equations (2) and (3). For a mass of 0·04 kg and a Q factor of 50, equation (2) gives $(\ddot{x}_i)_{ne} \approx 3\times10^{-10}$ m s^{-2} Hz^{-1}

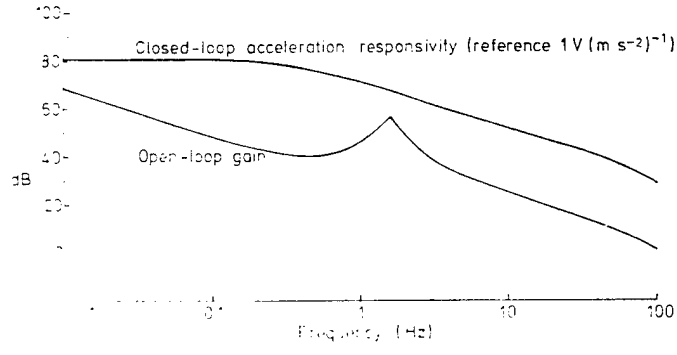


Figure 9 Open- and closed-loop responses of the feedback seismometer. The open-loop gain is greater than 40 dB over most of the seismic range (0·01 to 10 Hz). The closed-loop response is flat to acceleration from DC to 0·35 Hz (responsivity 10^4 V (m s^{-2})$^{-1}$) and flat to velocity from 0·35 to 100 Hz (responsivity $2{\cdot}4\times10^5$ V (m s^{-1})$^{-1}$).

for the Brownian contribution. The transducer/amplifier contribution with $R_n = 1{\cdot}5\ \text{k}\Omega$ and a natural period of 0·6 s gives $(\ddot{x}_1)_{ne} = 5 \times 10^{-11}\ \text{m s}^{-2}\ \text{Hz}^{-1}$ for $\omega < \omega_0$, which can be ignored. The other electronic sources of noise (PSD, filters, etc) have sufficient prior gain to make their contributions negligible. The theoretical noise-equivalent acceleration of the instrument is plotted in figure 2 (lines C and D).

The dynamic range of a wideband seismometer is particularly important, since seismic signals have a very large dynamic range. According to figure 2, the peak microseismic noise occurs at a period of about 8 s and has an acceleration amplitude of about $3 \times 10^{-6}\ \text{m s}^{-2}$. This is an average value (Wolverton in summer) and short-term values may be an order of magnitude greater. The 12 h tidal component has a similar value. The minimum earth noise (and the designed detection limit of the instrument) has an acceleration amplitude of $3 \times 10^{-10}\ \text{m s}^{-2}$. A range of about 80 dB is therefore required, excluding any additional range to accommodate excessive microseismic activity or seismic events. The maximum feedback acceleration is $10^{-3}\ \text{m s}^{-2}$, and the effective dynamic range of the instrument is therefore about 130 dB, allowing about 50 dB for events. In practice, however, the recording systems employed have a range of only 60–80 dB, so that several systems of different gain, bandwidth, etc, would be required to record the output of the instrument satisfactorily.

6 Testing and operation

The instrument has been operated over a period of several months in the AWRE vault at Wolverton, and compared with a Geotech S12 horizontal-component long-period seismometer. This instrument has a mass of 10 kg suspended with near-critical damping with a natural period of about 20 s; it employs an electrodynamic transducer and low-noise amplifier (type 610) and has overall dimensions of 70 cm × 30 cm × 30 cm.

The temperature variations in the vault are about 3°C per day but it was found unnecessary to use the thermal control system. This was switched off while investigating the source of a lack of coherence at long periods, and was not used again although the source was located elsewhere. The feedback instrument did not show an appreciable long-term drift and the results suggest that a period of six months without adjustment is feasible. The drifts observed were, however, about an order of magnitude greater than the earth tides, which we did not attempt to observe. The Q factor of the instrument was about 50 immediately after evacuation, and slowly fell to a value of about 15 after about four weeks.

The comparisons were carried out using various standard responses, obtained by filtering the wideband output of the feedback instrument. The responses used were a short-period response derived from a long-period instrument (LPSP), a broadband response (BB), and a long-period narrowband response (LPNB). The outputs were telemetered to Blacknest and recorded together on standard 'helicorders' at suitable magnifications.

Good coherence was obtained with the LPSP and BB responses, as expected because of the relatively high seismic noise in these bands. There was some excess high-frequency noise in the feedback instrument, probably due to its wider bandwidth. The outputs from the two instruments initially showed rather low coherence when using the LPNB response even though their responses had been closely matched. This was thought to be due to thermal fluctuations in the base plate of the feedback instrument. The improved Invar base plate described above was constructed and led to better coherence. Figure 10 shows an event recorded at low magnification and figure 11 shows typical long-period noise at high magnifica-

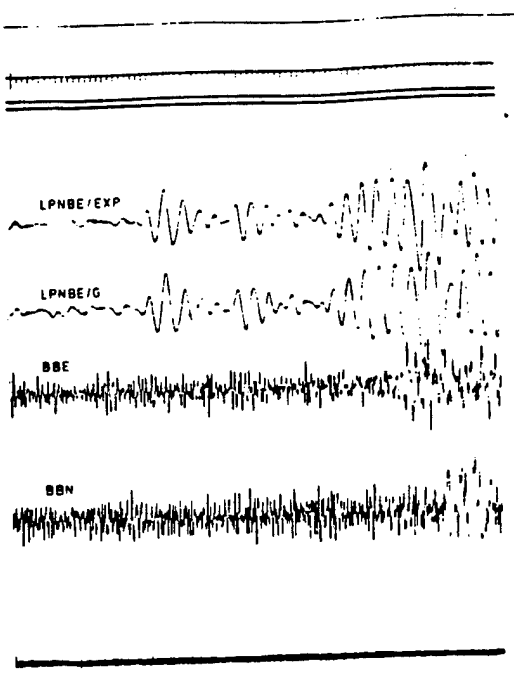


Figure 10 Comparison of seismometers (small event). A small event is shown at a magnification of 8200 at period 20 s using a long-period narrow-band (LPNB) response. LPNBE/EXP is from the feedback seismometer in the east–west orientation and LPNBE/G from the standard Geotech instrument. Two broadband records are also shown. (Event recorded at 03.30.00, 1 October 1975; scale, 8 s/division.)

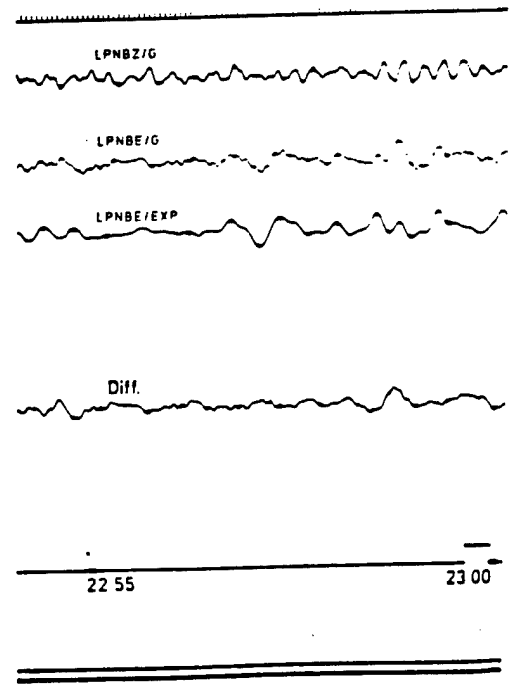


Figure 11 Comparison of seismometers (long-period noise). Long-period noise is shown at a magnification of 200 000 at 20 s using the LPNB response. LPNBZ/G is from a vertical-component Geotech instrument, and LPNBE/G and LPNBE/EXP are from the horizontal-component Geotech and feedback instruments. (Recorded at 22.55.00, 30 September 1975; scale, 4 s/division.)

COPIES 79433

tion. In the latter figure both recorded traces correspond to about 10 nm RMS at a period of 20 s (acceleration 10^{-9} m s^{-2}) in the LPNB bandwidth of about 0·04 Hz. The theoretical Brownian noise of the feedback instrument (with $Q=15$) in this bandwidth is 10^{-10} m s^{-2} and the theoretical electronic noise in the Geotech instrument has a similar value. Unfortunately it was not possible to compare two feedback instruments with one another (since only one was constructed) and two similar Geotech instruments were not available at the time.

The difference between the Geotech and feedback seismometer traces is mostly small with an RMS value of about 2 nm, corresponding to an RMS acceleration of 2×10^{-10} m s^{-2} at 20 s, which is close to the value expected theoretically from the Brownian noise value. However, occasional long-period fluctuations occur, with a magnitude of about 10 nm, probably due to mass movements caused by thermal effects or pressure changes. It is difficult to decide from the records whether the effect, which is very common in long-period instruments, occurs only in the feedback seismometer or the conventional seismometer or both, but it is more likely to be mainly due to the former.

The main practical problems in miniature instruments arise from long-term creep, thermal and pressure effects in the mechanical system, and air movements. As explained above, long-term creep did not prove to be a serious problem, and although thermal variations were observed their period was outside the seismic range of interest. (To observe longer-period seismic signals the thermal control system would have been required.) Pressure effects can be very serious, since the mass movements to be detected in a miniature instrument are smaller than in a conventional instrument in the ratio of the squares of the periods ($(0{\cdot}6)^2:(20)^2$). Similarly, unwanted forces due to air currents are equivalent to accelerations in the ratio of the masses (0·04 kg:10 kg). An evacuated pressure chamber is thus essential, though evacuation is required in any case to obtain a high Q factor. It is likely that the long-period fluctuations observed were due to thermal effects in the base plate of the instrument (the effect was much reduced by using the Invar base plate) and to pressure fluctuations leading to distortions of the base plate or instrument frame. In Block and Dratler's instrument the temperature was controlled to 10^{-6} °C and the pendulum was enclosed in a highly evacuated container and by an evacuated outer cover. However, it appears that the symmetrical design of the present instrument and the use of a rigid evacuated jacket and special base plate have enabled the above problems to be reduced to tolerable levels.

The comparison has successfully demonstrated that a miniature wideband feedback instrument can compare closely with a conventional open-loop instrument of much greater size. The increased electronic complexity is more than compensated by the reduced mechanical complexity, size and weight, and by the ease of operation and calibration. The cost, on a production basis, would be considerably less.

7 Conclusions

It has been demonstrated that it is feasible to build a truly miniature wideband seismometer with an output linearly proportional to ground motion and well defined by substantial negative feedback over the whole of the seismic range, and with a noise level comparable with much larger conventional instruments. The development depends strongly for its success on the use of a capacitive displacement transducer of high responsivity and low noise, and on the application of negative feedback to maintain the mass fixed with respect to its supports giving the attendant advantages of controlled wideband response, linearity and dynamic range.

A difficulty in designing a miniature seismometer is that, since a long natural period cannot easily be used, the displacements to be detected are very small and may become comparable with unwanted movements due to creep, air currents, etc. These problems were overcome by a carefully designed suspension system in which effects of temperature gradients were minimised. Although an evacuated cover and a thermal control system were used, the requirements of these systems were not stringent and the overall size is adequate for borehole applications.

It is, of course, easier to design a horizontal-component instrument than a vertical-component instrument, since the steady acceleration of gravity must be cancelled in the latter. However, the same principles can still be applied and an apparently satisfactory vertical-component instrument of similar overall dimensions is at present being tested.

Acknowledgments

The authors wish to thank Professor Fellgett of the Cybernetics Department, University of Reading, who initiated this research and who first suggested that a combination of low damping and feedback could give a high detectivity even with low mass. The work was supported initially by a research grant from NERC and later by a research contract from AWRE. This contract arose through the initiative of Dr Thirlaway of the AWRE Seismological Centre at Blacknest, and the authors are much indebted to him for his continued help and encouragement and for the provision of testing facilities at Blacknest and Wolverton.

References

Block B and Dratler J 1972 *J. Geophys. Res.* **77** 3678–89

Block B and Moore R D 1966 *J. Geophys. Res.* **71** 4361–75

Block B and Moore R D 1970 *J. Geophys. Res.* **75** 1493–505

de Bremaeker J C, Donoho P and Michel J G 1962 *Bull. Seismol. Soc. Am.* **52** 661–72

Brune J N and Oliver J 1959 *Bull. Seismol. Soc. Am.* **49** 349–53

Buckner I W 1975 *PhD Thesis* University of Reading

Danby P G G 1968 *Elect. Engr* **40** 668–9

Dratler J 1971 *Doctoral Thesis* University of California

Faulkner E A and Grimbleby J B 1967 *Elect. Engr* **39** 565–7

Fix J E 1972 *Bull. Seismol. Soc. Am.* **62** 1753–60

Grimbleby J B and Harding D W 1971 *J. Phys. E: Sci. Instrum.* **4** 941–4

Jones R V 1967 *Phys. Bull.* **18** 325–36

Jones R V and Richards J C S 1973 *J. Phys. E: Sci. Instrum.* **6** 589–600

Melton B S 1976 *Rev. Geophys. Space Phys.* **14** 93–116

Russell R D, Meldrum R D and Jersen O G 1968 *Bull. Seismol. Soc. Am.* **58** 1621–30

Savino J, McCamy K and Hade G 1972 *Bull. Seismol. Soc. Am.* **62** 141–76

Sutton G H and Latham G V 1964 *J. Geophys. Res.* **69** 3865–83

Tucker M J 1958 *J. Sci. Instrum.* **35** 167–71

Usher M J 1973 *J. Phys. E: Sci. Instrum.* **6** 501–7

Willmore P L 1959 *Bull. Seismol. Soc. Am.* **49** 99–114

Unconventional reciprocity calibration of transducers

Isadore Rudnick

Physics Department, University of California, Los Angeles, California 90024
(Received 6 January 1978)

The procedure for the reciprocity calibration of transducers in unconventional acoustic geometries is described.

PACS numbers: 43.88.Ar, 43.30.Yj

INTRODUCTION

It is customary when thinking about the reciprocity calibration of transducers to have in mind the acoustic geometry of either the free field or a pressure chamber. The theory and technique of obtaining calibrations for such cases have received close attention and are very well developed. However, there are situations in the experimental laboratory where the acoustic geometry is radically different and where the constraints of the experiments require that the transducers be calibrated *in situ*.

Perhaps a few examples drawn from our own experience will make the problem clear. We have a program of investigating the superfluid state of helium in which acoustic techniques are heavily used. Our usual transducer is electrostatic and consists of an insulated button electrode which may be part of the wall of a wave guide or resonantor, and a plastic diaphragm, metallized on one side, which sits on the high points of the button surface. Liquid helium fills the space (perhaps 1–10 μm in thickness) between the diaphragm and electrode where they do not touch. They are used both as speakers and microphones and have a high mechanical impedance because of the thinness of the liquid helium layer. Calibrations must be repeated each time the apparatus is cooled to liquid helium temperatures since there is no assurance that the transducer sensitivities remain unchanged in a thermal cycle.

Figure 1 shows a rigid rectangular waveguide with a nonreflecting termination. Three identically made transducers are shown. The waveguide is many wavelengths long and is operated in its plane-wave mode. There is negligible change in amplitude of the progressive plane wave as it propagates toward the termination. Our problem is to calibrate all three transducers.

In Fig. 2 the nonreflecting termination has been replaced by a rigid reflector. In the most common version of this geometry the rigid terminations are two end transducers themselves.

In another example shown in Fig. 3 two identical volumes are connected by a neck resulting in a double cavity Helmholtz Resonator. Three transducers like those in Figs. 1 and 2 are shown.

Since, in our examples, all three transducers are reversible each of them can be calibrated as a microphone and a speaker. However, the procedure requires only one reversible transducer (T), a microphone (M), a speaker (S), and results in a calibration of all three.

I. RATIONALE OF RECIPROCITY CALIBRATION OF TRANSDUCERS

A very brief outline of the basis for the calibration procedure is in order. Consider the linear passive four-pole electrical system shown in Fig. 4. We write

$$V_1 = ai_1 + bi_2, \qquad (1)$$
$$V_2 = ci_1 + di_2.$$

One or both of the terminal pairs can be replaced by mechanical or acoustic connections in which case (voltage, current) is replaced by (force, velocity) or (pressure, volume velocity). The system, whether it is electrical, mechanical, acoustic, electromechanical, or electroacoustic is said to obey the Reciprocity Theorem if[1,2]

$$b = \pm c. \qquad (2)$$

Consider two electrical transducers 1 and 2 in Fig. 5 connected by an acoustic link (this may be the waveguide or one of the resonators of Figs. 1–3), then choosing the positive sign in Eq. (2) we can write

$$V_1 = ai_1 + bi_2,$$
$$V_2 = bi_1 + ci_2. \qquad (3)$$

Following MacLean[3] we let

M_o = microphone sensitivity in open circuit volts divided by the pressure at the microphone,

M_s = microphone sensitivity in short circuit amps divided by the pressure at the microphone,

S_o = speaker output in pressure produced at the microphone divided by amps into the speaker, and

S_s = speaker output in pressure produced at the microphone divided by volts across the speaker.

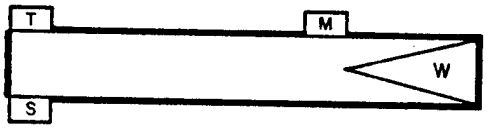


FIG. 1. Rigid wall rectangular wave guide with nonreflecting termination W at one end and a rigid termination at the other. (T), (S), and (M) are identically made electrostatic transducers whose sensitivities are not identical. The cross-sectional dimensions of the waveguide and transducer dimensions are very much less than λ (the acoustic wavelength), and the waveguide is long compared to λ. Transducers (T) and (S) are right at the rigid termination.

0001-4966/78/6306-1923$00.80
© 1978 Acoustical Society of America

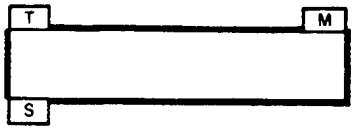


FIG. 2. Same as Fig. 1 except that the nonreflecting termination is replaced by a rigid termination. All transducers are right at the rigid terminations.

Now if we drive transducer 1 as a speaker and get the response of transducer 2 in both open circuit volts and short circuit amps and then make the same measurements but reverse the roles of the transducers we find[3] that

$$\frac{S_{o1}}{M_{o1}} = \frac{S_{o2}}{M_{o2}} = \frac{S_{s1}}{M_{s1}} = \frac{S_{s2}}{M_{s2}} = Z . \tag{4}$$

We have yet to determine Z which clearly is related to the acoustic geometry. But before doing this we point out that, when Z is known, the reciprocity calibration of identical transducers immediately follows since, if transducers 1 and 2 are identical, then

$$i_1 S_o M_o = V_2$$

and

$$M_o = (V_2/i_1 Z)^{1/2} \times 10^{-7/2}, \qquad S_o = (V_2 Z/i_1)^{1/2} \times 10^{-7/2} . \tag{5}$$

(The factor of $10^{-7/2}$ is necessary if electrical quantities are in volts and amps and mechanical quantities are in cgs.)

We now determine Z. We choose as our reversible transducer an "ideal microphone"—it is small compared to the wavelength of sound and is so noncompliant that its introduction at a point in the sound field never alters the sound pressure at that point. By the same token, when used as a speaker, its volume velocity is independent of the acoustic load. In Fig. 5 both transducers are identical and of this type. Thus when transducer 1 is used as a speaker driven by a voltage V_1 and a current i_1, a volume velocity U_1 is produced *at the speaker* and

$$V_1 = ai_1 + bU_1, \qquad 0 = bi_1 + cU_1 . \tag{6}$$

Zero appears at the left of the second equation in (6) because there is no impressed pressure and the transducer does not feel the pressure it, itself, generates.

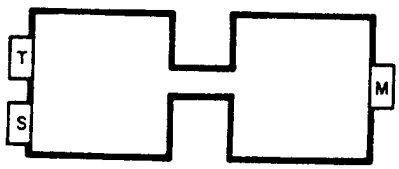


FIG. 3. A double cavity Helmholtz Resonator with three transducers, (T), (M), and (S). The volume of each cavity is V. The length of the neck is l and its cross-sectional area is A. Its angular resonant frequency ω is given by $\omega = C(2A/lV)^{1/2}$, where C is the sound velocity.

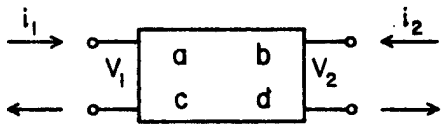


FIG. 4. A linear passive four pole network.

For the transducer (2) being used as a short circuited microphone we can write

$$0 = ai_2 + bU_2, \qquad p_2 = bi_2 + cU_2, \tag{7}$$

where p_2 is the pressure and U_2 is the volume velocity *at the microphone*. Now $S_s = p_2/V_1$, $M_s = i_2/p_2$ and $p_2/i_2 = V_1/U_1 = b - ac/b$.

Thus

$$Z = \frac{p_2}{U_1} . \tag{8}$$

We see that Z is just the transfer impedance relating the acoustic pressure at the microphone to the volume velocity at the speaker.[4]

II. THE CALIBRATION

We now describe the procedure for calibrating transducer T (and subsequently S and M) in the configurations of Figs. 1–3. We send a current i_T through T and measure the open circuit voltage V_{M_1} at M.

If M and T were identical transducers, M_o of T would be given by $(V_{M_1}/i_T 1/Z)^{1/2} 10^{-7/2}$. But they're not identical and we need to know what voltage T would have read if it were at the position of M. We get this by arranging to have a sound field generated by S which produces identical pressures at M and T.[5] Then if we drive S with an arbitrary current i_s open circuit voltages V_{M_2} and V_T will be read at M and T and the microphone and speaker sensitivity of T is given, respectively, by

$$M_o = \left(\frac{1}{Z}\,\frac{V_T}{i_T}\,\frac{V_{M_1}}{V_{M_2}}\right)^{1/2} \times 10^{-7/2}, \qquad S_o = \left(Z\,\frac{V_T}{i_T}\,\frac{V_{M_1}}{V_{M_2}}\right)^{1/2} \times 10^{-7/2}, \tag{10}$$

where M_o for transducer M is the product of V_{M_2}/V_T and the value of M_o above, and S_o for transducer S is the product of i_T/i_S and the value of S_o above. Using M as a speaker and S as a microphone will yield their sensitivities in these roles and the calibrations are complete.

In Fig. 1 the acoustic pressures at M and T are necessarily the same when S is driven at an arbitrary frequency. In Fig. 2 the pressures will be equal when the resonator is driven by S at any of its resonant fre-

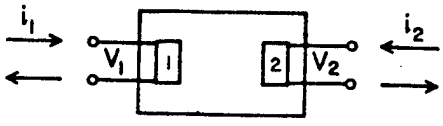


FIG. 5. Transducers 1 and 2 are coupled by an acoustic field.

quencies if the Q's are very high. In Fig. 3 the pressures will be equal when the resonator is driven by A at its resonant frequency if the Q is very high.

The transfer impedances to be used in Eq. (10) for the three cases are

$$\text{Fig. 1: } Z = \rho C/A, \tag{11}$$

$$\text{Fig. 2: } Z = 2\rho C Q_n/An\pi, \tag{12}$$

where Q_n is the quality factor of the nth resonant harmonic:

$$\text{Fig. 3: } Z = \rho C^2 Q/2V\omega, \tag{13}$$

where $\omega = C(2A/lV)^{1/2}$ is the resonant frequency of the resonator and Q the quality factor of the mode.

The speaker S is only needed to generate a sound field by which the relative microphone sensitivity of M and T is determined. In the experimental arrangement of Figs. 2 and 3 the speaker S can be dispensed with in the following way. T is used as a source to establish a natural resonance of the system and then it is switched to its microphone mode. A comparison of its output with that of M, as the resonance decays, provides the relative sensitivity of the two.

ACKNOWLEDGMENT

This research was supported by ONR Contract N00014-75-C-0246 and NSF Contract DMR 76-22306.

[1]L. Foldy and H. Primakoff, J. Acoust. Soc. Am. 17, 109–120 (1945), 1950–1958 (1947).
[2]E. M. McMillan, J. Acoust. Soc. Am. 18, 344–347 (1946).
[3]W. R. MacLean, J. Acoust. Soc. Am. 12, 140–146 (1940).
[4]Textbook treatments of reciprocity calibration of transducers would be easier to understand if the transducer which is an "ideal microphone," namely, the small, noncompliant transducer, is explicitly introduced. This seems to be essential to obtain the second equation of (6) and the result, Eq. (8). A minor mystery is why the fact that the ratio in Eq. (4) is the transfer impedance, relating the pressure at a microphone to the volume velocity at a speaker, is not emphasized.
[5]It is unnecessary that the pressures be identical. It is only necessary that their ratio is known.

JULY 1, 1929 PHYSICAL REVIEW VOLUME 34

A PROBLEM IN BROWNIAN MOTION

By G. E. Uhlenbeck and S. Goudsmit
Department of Physics, University of Michigan

(Received April 26, 1929)

Abstract

Gerlach investigated the rotatorial Brownian motion of a small mirror suspended on a fine wire. It follows from the theorem of equipartition that the average square deviation of the mirror will depend on the temperature alone of the surrounding gas. Gerlach verified this for a large range of pressures (1 to 10^{-4} atm). The analogy which we found that exists between this problem and the well-known treatment of the shot effect by Schottky enables us to give a more detailed theory of this phenomenon. If the displacement, registered during a time, long compared with the characteristic period of the mirror, is developed into a Fourier series, we find the square of the amplitude of each Fourier component to be a function of the pressure and molecular weight of the surrounding gas as well as of its temperature, (formula 18). The sum of the squares, however, is a function of the temperature alone (proved in section 4). This explains why the curves registered by Gerlach at different pressures, though all giving the same mean square deviation, are quite different in appearance. To get the fluctuating torque on the mirror, the expression:

$$\overline{\delta_\mu{}^2} = \frac{16}{\pi} \cdot \frac{1}{n} \cdot \frac{1}{\bar{c}\,\Delta I \Delta o}$$

is obtained for the fluctuation in time of the pressure of a gas on the wall (section 5). In this n represents the number of molecules per cc, $\bar{c}$ is the mean velocity and Δo is the surface of the wall.

I. Introduction

INTERESTING experiments on Brownian motion around a position of equilibrium have been performed by Zeeman and Houdyk[1] in Amsterdam and by Gerlach[2] in Tübingen. The former registered the motion of the loose end of a suspended wire, the latter photographed by means of reflected light the rotational Brownian movement of a little mirror fixed on a very fine wire. The first experiment is theoretically more complicated, because one has to consider the many natural frequencies of the observed body. In the experiment of Gerlach on the other hand the observed system has only *one* characteristic frequency. We will restrict ourselves therefore in the following treatment to the latter case.

In both cases one can immediately predict by means of the equipartition theorem what the average square of the deflection will be. This will depend on the properties of the observed system and on *the temperature only* of the surrounding gas, *not* for instance on its pressure or molecular weight. The experiments however give *more* than merely the average square deviation; the registered curves show to some extend at least the time-dependence of

[1] P. Zeeman and A. Houdyk, Proc. Acad. Amsterdam, 28, 52 (1925).
[2] W. Gerlach, Naturwiss. 15, 15 (1927).

the irregular Brownian motion. As Professor Gerlach kindly communicated to us, the general appearance of these curves is quite different at different pressures of the surrounding gas, though the average square deviation remains the same for any given temperature. The problem is therefore to give a more detailed theory of these curves.

It has occurred to us that this problem may be treated in a manner quite analogous to the method employed by Schottky,[3] to describe the well-known shot-effect. In the experiments of Hull and Williams,[4] the fluctuating voltage in the shot-circuit is coupled inductively with the amplification-circuit, which possesses only one characteristic frequency. In the experiments of Gerlach the fluctuating moment of momentum around the mirror-axis of the gas-molecules, is coupled, by means of collisions, with the "amplifying" mirror, which has also only one characteristic frequency.

The analogy is complete only when *the surrounding gas is much rarefied*, because only then are the moments of momentum given by the gas-molecules to the mirror in successive time-elements independent of each other.[5]

By applying the method of Schottky, we will show in Sections II and III that for this case the amplitudes of the Fourier components of the motion depend on the pressure and the molecular weight of the surrounding gas. This will explain the different forms of the observed curves under various circumstances.

II. The Fourier-Analysis of the Brownian Motion

The equation of motion of the mirror is given by:

$$I\ddot{\phi}+f\dot{\phi}+D\phi=M(t), \tag{1}$$

where: I is the moment of inertia around the mirror-axis; ϕ is the angle of deflection; f the friction-coefficient; D the directional force; and $M(t)$ the fluctuating torque, caused by the collisions of the gas molecules. When we introduce the frequency in 2π sec:

$$\omega=(D/I)^{1/2} \tag{2}$$

and the angular acceleration:

$$T(t)=M(t)/I \tag{3}$$

and put:

$$r=f/I \tag{4}$$

Eq. (1) becomes:

$$\ddot{\phi}+r\dot{\phi}+\omega^2\phi=T(t). \tag{5}$$

For our further purposes it is essential to give now a more detailed discussion of the meaning of $M(t)$ or $T(t)$.

[3] W. Schottky, Ann. der Phys. 57, 541 (1918); 68, 157 (1922). Comp. also: J. Tinbergen, Physica, 5, 361 (1925).

[4] A. W. Hull and N. H. Williams, Phys. Rev. 25, 147 (1925). Comp. also N. H. Williams and H. B. Vincent, Phys. Rev. 28, 1250 (1926).

[5] For higher pressures, the problem becomes analogous to the problem of the shot effect for high current densities. Because of the space charge the numbers of electrons hitting the anode in successive time elements are then not more independent of each other, and the fluctuations decrease. Comp. N. H. Williams and H. B. Vincent, ref.[4] p. 1262 and N. H. Williams and W. S. Huxford, Phys. Rev. 33, 773 (1929).

The actual microscopic $M^*(t)$ consists of a large number of sharp peaks, each corresponding to the impulse moment transferred to the mirror by the collision of one (or a few) molecules. We introduce now a "physically infinitely small" time-element Δt,[6] very small compared with the characteristic period $2\pi/\omega$ of the mirror but within which many collisions occur.[7] Our function $M(t)$ consists of the averages of all $M^*(t)$ values included in each time-element Δt. The actual value of $M(t_i)$ for the time-element Δt_i is of course unknown *a priori*, but we can tell some of its properties:

(*a*) $M(t_i)$ will have equal chance of bei ig positive or negative, so that the average over all the time-elements is zero.

(*b*) In our case, when the surrounding gas is rarefied, the $M(t_i)$ in the time-element Δt_i will be independent of the value $M(t_j)$ in the time-element Δt_j.

(*c*) In Section V we will show that the mean of $M^2(t_i)$ over all the time-elements is given by:

$$\overline{M^2(t_i)} = 4m\bar{c}pI/\rho\Delta t \tag{6}$$

where: m is the mass of the gas molecules; $\bar{c}$ their mean velocity; p the pressure of the surrounding gas, and ρ the mass of the mirror per square cm.

Let τ be the time of observation, which must be long compared with the period of the mirror, so that:

$$\Delta t \ll 2\pi/\omega \ll \tau. \tag{7}$$

Develop $T(t)$ within the interval $(0,\tau)$ in a Fourier series:

$$T(t) = \sum_{k=0}^{\infty} (A_k \cos \omega_k t + B_k \sin \omega_k t) \tag{8}$$

where:

$$\left.\begin{aligned} \omega_k &= 2\pi k/\tau \\ A_k &= \frac{2}{\tau}\int_0^{\tau} T(t)\cos\omega_k t \cdot dt \\ B_k &= \frac{2}{\tau}\int_0^{\tau} T(t)\sin\omega_k t \cdot dt \end{aligned}\right\} \tag{9}$$

We can now replace these integrals by the following sums:

$$\left.\begin{aligned} A_k &= \frac{2}{\tau}\sum_1^{s} T(t_i)\cos\omega_k t_i \cdot \Delta t_i \\ B_k &= \frac{2}{\tau}\sum_1^{s} T(t_i)\sin\omega_k t_i \cdot \Delta t_i \end{aligned}\right\} \tag{10}$$

where Δt_i are the successive, equal time-elements, and $T(t_i)$, $\cos \omega_k t_i$, $\sin \omega_k t_i$ are evaluated at a time t_i included within the i^{th} element Δt_i. Z

[6] We suppose them all equal.

[7] The introduction of such an "physically infinitely small" time element is characteristic in the kinetic theory of gases. Comp. e.g., P. and T. Ehrenfest, Enc. der Math. Wiss. Vol. IV, Art. 32, p. 39.

is the total number of time-elements. The motion of the mirror is then expressed by:

$$\phi(t)=\sum_k \phi_k(t)=\sum_k \frac{1}{(\omega^2-\omega_k^2)^2+r^2\omega_k^2}\left[\left\{A_k(\omega^2-\omega_k^2)-B_k r\omega_k\right\}\cos\omega_k t + \left\{A_k r\omega_k+B_k(\omega^2-\omega_k^2)\right\}\sin\omega_k t\right] \tag{11}$$

which is the solution of (5) under the conditions $\phi=\dot{\phi}=0$ for $t=0$. These conditions mean that we start our observations only when all external disturbances have been damped out, and the remaining motion is due only to collisions with the gas molecules, or in other words, we observe the Brownian motion around the position of equilibrium and not around an already existing vibration.

From this, we find for the time-average of ϕ_k^2:

$$\overline{\phi_k^2}=\frac{1}{2}\,\frac{A_k^2+B_k^2}{(\omega^2-\omega_k^2)^2+r^2\omega_k^2}. \tag{12}$$

In the case of very low pressure it is now possible to predict from the three properties (*a*), (*b*) and (*c*) of $M(t)$ (or $T(t)$) the value of A_k^2 and B_k^2. From (10) we have

$$A_k^2=\frac{4}{\tau^2}\sum_{i=1}^{Z}\sum_{j=1}^{Z}T(t_i)T(t_j)\cos\omega_k t_i\cos\omega_k t_j\cdot\Delta t_i\Delta t_j. \tag{13}$$

Consider first in this double sum the terms with $i\neq j$. As a consequence of property (*b*), in the case of low pressure $T(t_i)$ and $T(t_j)$ are completely independent; hence these terms will have equal chances of being positive or negative and for large Z their sum will vanish. Consider next the terms with $i=j$, which are all positive. Due to our choice of Δt_i, for all frequencies ω_k of the order of magnitude of ω (and only those give according to (12) an appreciable $\overline{\phi_k^2}$,) $2\pi/\omega_k$ will be very large with respect to Δt_i, so that $\cos^2\omega_k t$ changes very little over many time-elements Δt_i. We may therefore replace $T^2(t_i)$ by its average value and obtain:

$$A_k^2=\frac{4}{\tau^2}\overline{T^2(t)}\sum_{i=1}^{Z}\cos^2\omega_k t_i(\Delta t_i)^2. \tag{14}$$

Finally, replacing the sum once more by an integral, we have:

$$A_k^2=2/\tau\;\overline{T^2(t)}\cdot\Delta t \tag{15}$$

Obviously the value of B_k^2 is the same.

III. Discussion of the Final Formula

Formula (6), which will be developed in §5, together with equation (3) gives:

$$\overline{T^2(t)}=4m\bar{c}p/I\rho\Delta t. \tag{16}$$

In §6 we will prove, that for the case of low pressure, the friction-coefficient is given by:

$$f=2m\bar{c}p\rho I/kT. \tag{17}$$

Substituting (15), (16), (17) in (12), we get *the final formula:*

$$\overline{\phi_k^2}=\frac{\pi^{1/2}m^{1/2}(8kT)^{3/2}\cdot p}{\rho l\cdot\tau}\cdot\frac{1}{\pi kT(\omega^2-\omega_k^2)^2+32p^2\rho^2\omega_k^2} \tag{18}$$

in which the well-known relation is used that:

$$\bar{c}=(8kT/\pi m)^{1/2}.$$

From the analysis in Section II it is clear that this formula may be interpreted in the *two* following ways:

1. If we resolve into Fourier series *a great number of curves*, each observed over *a relatively short time* τ (which however must still fulfill the fundamental inequality (7)), the mean square of the amplitudes of the k^{th} components will be given by (18).

2. If we analyse *one curve*, observed over *a very long time* τ, then the square of the amplitudes of the k^{th} component will also be given by (18).

The formula (18) shows the noteworthy result, that $\overline{\phi_k^2}$ depends not only on the temperature, but also explicitly on *the pressure* and *the molecular weight* of the surrounding gas. As a test we must of course show, as we will do in §4, that *by summing over all values of k*, we obtain for the mean potential energy $\frac{1}{2}D\overline{\phi^2}$ the equipartition value $\frac{1}{2}kT$ which is *independent* of the pressure and the molecular weight of the surrounding gas.

The dependence on the pressure is rather complicated. For frequencies ω_k very near to ω, the $\overline{\phi_k^2}$ becomes inversely proportional to p, and for ω_k very large compared with ω, the $\overline{\phi_k^2}$ becomes almost directly proportional to p. These latter terms of course contribute very little to the total motion, the denominator being so large. When we plot therefore $\overline{\phi_k^2}$ against k, the resulting curve has a maximum in the neighborhood of ω, which rapidly becomes very sharp as the pressure decreases. The motion of the mirror than becomes more and more "monochromatic."

Appendix

IV. Proof of the Equipartition Theorem.

The average potential energy of the vibrating mirror is equal to:

$$\tfrac{1}{2}D\overline{\phi^2}=\tfrac{1}{2}D\sum_{k=0}^{\infty}\overline{\phi_k^2} \tag{19}$$

because the different Fourier components are independent of each other. Though we restricted ourselves to values of ω_k small with respect to $2\pi/\Delta t$, it is permissible to extend the summation to infinity, because the components with ω_k large compared to ω contribute very little to the sum.

Equation (15) shows, that A_k^2 and B_k^2 are independent of k; hence from (12) and (15):

$$\overline{\phi^2}=\sum_{k=0}^{\infty}\overline{\phi_k^2}=\frac{2}{\tau}\overline{T^2(t)}\cdot\Delta t\sum_{k=0}^{\infty}\frac{1}{(\omega^2-\omega_k^2)^2+r^2\omega_k^2}\cdot \tag{20}$$

The last sum we now replace by an integral, substituting:

$$x_k=\omega_k/\omega=2\pi k/\tau\omega$$

which gives:

$$\sum_{k=0}^{\infty}\frac{1}{(\omega^2-\omega_k^2)r^2+r^2\omega_k^2}\cong\frac{\tau}{2\pi\omega^3}\int_0^{\infty}\frac{dx}{(1-x^2)^2+(r^2/\omega^2)x^2}$$

with very good approximation. The value of the integral is $\pi\omega/2r$,[8] so that we obtain:

$$\overline{\phi^2} = \overline{T^2(t)} \cdot \Delta t/2\omega^2 r. \tag{21}$$

Introducing the relations (16) and (17), and substituting from (2) and (4) the values of ω^2 and r, we obtain immediately:

$$\tfrac{1}{2}D\overline{\phi^2} = \tfrac{1}{2}kT. \tag{22}$$

V. Proof of the Fluctuation Formula (6) or (16).

This relation follows from a consideration of *the fluctuations in time* of the pressure exerted upon a wall by a rarefied gas. Using Maxwell's distribution law one easily derives an expression for the probability that a molecule of a gas within a volume V at the temperature T gives to a portion Δo of the wall a momentum normal to the wall lying between G and $G+\Delta G$ during the time Δt. It is:[9]

$$W\Delta t\Delta o\Delta G = \frac{\pi}{2}\cdot\frac{1}{(2\pi mkT)^{3/2}}\cdot\frac{kT}{V}\cdot Gc^{-G^2/8mkT}\Delta t\Delta o\Delta G. \tag{23}$$

Let n_{ij} be the number of molecules, which in the time-element Δt_i give a momentum lying between G_j and $G_j+\Delta G_j$ to the portion Δo of the wall. Then the total momentum given to Δo during Δt_i becomes:

$$G(t_i) = \sum_{j=0}^{\infty} G_j n_{ij}.$$

Using the bar to denote the average over all time-elements, we find easily:

$$\overline{G^2(t_i)} - \overline{G(t_i)}^2 = \overline{(\Sigma_j G_j n_{ij})^2} - \overline{(\Sigma_j G_j n_{ij})}^2 = \Sigma_j G_j^2(\overline{n_{ij}^2} - (\bar{n}_{ij})^2) \tag{24}$$

because the cross-terms cancel, and the average over the time-elements Δt_i extends only over the n_{ij}. In general the fluctuation formula holds:

(Poisson Distribution)

$$\overline{n_{ij}^2} - (\bar{n}_{ij})^2 = \bar{n}_{ij} \tag{25}$$

and $\bar{n}_{ij}$ follows immediately from (23), after multiplying by the total number of molecules N. Substituting then (25) in (24), and replacing the sum by an integral, we obtain:

$$\overline{G^2(t_i)} - \overline{G(t_i)}^2 = \frac{\pi}{2}\cdot\frac{p}{(2\pi mkT)^{3/2}}\Delta o\Delta t\cdot\int_0^{\infty} G^3\exp[-G^2/8mkT]dG = 2m\bar{c}p\Delta o\Delta t \tag{26}$$

for:

$$p = NkT/V.$$

[8] See W. Schottky, Ann. der Phys. 68, 157 (1922).

[9] Integrating over G from 0 to ∞, we get for the probability that a molecule hits in the time Δt the surface element Δo:

$$w\Delta t\Delta o = \bar{c}\Delta t\Delta o/4V$$

corresponding to the well-known result for the mean number $\bar{n}$ striking 1 cm² of the wall per second:

$$\bar{n} = N\bar{c}/4V.$$

In the same way we find for the mean momentum given to the wall taken over all the molecules striking it:

$$\bar{G} = (2\pi mkT)^{1/2}$$

and for the mean square:

$$\overline{G^2} = 8mkT$$

This, divided by Δt^2, expresses the fluctuation in time of the pressure on the portion Δo of the wall.[10] From this it follows obviously that for a disc of surface o inside the gas, the mean of the square of *the force* $K(t_i)$ taken over all the time-elements Δt_i is given by:

$$\overline{K^2(t_i)} = 4m\bar{c}po/\Delta t \tag{27}$$

as for such a disc $\overline{K(t_i)} = 0$, and the fluctuations on the right and left side are independent of each other.

For a case like the experiment of Gerlach we must consider the moment of momentum around the mirror axis instead of the momentum. The analogous formula for *the torque* is then:

$$\overline{M^2(t_i)} = \frac{4m\bar{c}p}{\Delta t}\int\int x^2 do = \frac{4m\bar{c}pI}{\rho\Delta t}$$

where x denotes the distance from the axis. This is equation (6).

VI. Proof of Formula (17) for the Friction Coefficient.[11]

Consider a portion Δo of the mirror, which moves, say to the right in the direction of the x-axis with the velocity u. The number of molecules per second, which strike this from the left, and which lie within a certain velocity-range $d\xi d\eta d\zeta$ is:

$$dN_1 = \left(\frac{m}{2\pi kT}\right)^{3/2}\frac{N}{V}(\xi - u)e^{-mc^2/2kT}\Delta o\, d\xi d\eta d\zeta \tag{28}$$

where we have used Maxwell's distribution law, because *in our case of very low pressure*, the mean free path is large with respect to the dimensions of the mirror, so that the motion of the mirror *does not disturb the velocity distribution* of the molecules. If x is the distance from Δo to the axis of the mirror, then the moment of momentum imparted per second by these molecules is:

$$dM_1 = \left(\frac{m}{2\pi kT}\right)^{3/2}\cdot\frac{2Nm}{V}\cdot(\xi - u)^2\cdot x\cdot e^{-mc^2/2kT}\Delta o\, d\xi d\eta d\zeta. \tag{29}$$

Neglecting the term with u^2 and integrating over η and ζ from $-\infty$ to $+\infty$, and over ξ from 0 to $+\infty$, we get:

$$M_1 = px\Delta o - m\bar{c}Nux\Delta oV^{-1}. \tag{30}$$

In the same way, we find for the moment of momentum given to Δo per second by molecules striking from the right:

$$M_2 = -px\Delta o - m\bar{c}Nux\Delta oV^{-1} \tag{31}$$

so that the total moment of momentum given to Δo per second is:

$$M_1 + M_2 = -2m\bar{c}pux\Delta o/kT. \tag{32}$$

Now $u = x\dot{\phi}$, so that we find for the friction coefficient:

$$f = \frac{2m\bar{c}p}{kT}\int\int x^2 do = \frac{2m\bar{c}pI}{\rho kT}$$

which is formula (17).

[10] This can be written in the form:

$$\frac{\overline{p^2} - (\bar{p})^2}{(\bar{p})^2} = \overline{\delta_p^2} = \frac{16}{\pi}\cdot\frac{1}{n}\cdot\frac{1}{\bar{c}\Delta t\Delta o}$$

when n is the number of molecules per cc. It has then the same form as the expression for the fluctuation in pressure of a gas inside a volume element Δv (see R. Furth, Die Schwankungserscheinungen in der Physik, Vieweg, Braunschweig, 1920, p. 58):

$$\overline{\delta_p^2} = \frac{C_p}{C_v}\cdot\frac{1}{n}\cdot\frac{1}{\Delta v}$$

but it cannot be derived from it.

[11] Comp. H. A. Lorentz, Les theories statistiques en thermodynamique, Leipzig, 1916, p. 53.

LA-UR-96-1335

For Encyclopedia of Applied Physics; prepared 3/96

Thermoacoustic engines and refrigerators

Gregory W. Swift

Condensed Matter and Thermal Physics Group

Los Alamos National Laboratory, Los Alamos NM 87545

Abstract

We usually think of a sound wave as consisting of coupled pressure and displacement oscillations, but temperature oscillations also occur. Near a solid surface, the combination of these temperature oscillations with the pressure and displacement oscillations produces a rich variety of "thermoacoustic" effects. These effects can be harnessed to produce powerful heat engines, heat pumps, and refrigerators using intense sound waves in suitable geometries.

To introduce the subject of thermoacoustics, we discuss relevant length scales, and present ordinary thermoviscous dissipation of sound in quantitative detail. To provide intuitive appreciation of thermoacoustic processes, we describe the time-dependent motion, pressure, temperature, and density of typical parcels of gas. The emphasis is on standing-wave systems, but some discussion of Stirling engines and refrigerators as traveling-wave thermoacoustic systems is included. Current quantitative design and analysis methods are summarized, and relevant dimensionless groups are presented to enable analysis of systems operating beyond the range of current methods.

I. Introduction

We usually think of a sound wave in a fluid (gas or liquid) as consisting of coupled pressure and displacement oscillations, but temperature oscillations also occur. Near a solid surface, the combination of these temperature oscillations with the pressure and displacement oscillations produces a rich variety of "thermoacoustic" effects.

Although thermoacoustic effects are too small to be readily noticed in everyday life, they can be harnessed to produce powerful refrigerators (including heat pumps) and heat engines, using intense acoustic standing waves in suitable geometries. Whereas typical engines and refrigerators have crankshaft-coupled pistons or rotating turbines, standing-wave thermoacoustic engines and refrigerators have no moving parts, or at most a single flexing moving part (as in a loudspeaker) with no sliding seals. Such thermoacoustic systems appear attractive because of their elegance, reliability, and low cost, in spite of only modest efficiency.

While standing-wave thermoacoustic systems have developed only recently, Stirling engines and refrigerators have a long, rich history. Recently, new insights have resulted from applying a thermoacoustics viewpoint to Stirling systems, treating them as traveling-wave thermoacoustic systems. Although Stirling systems are usually regarded as a separate subject, we will include some discussion of thermoacoustic aspects of Stirling systems here.

To be consistent with the article "Heat engines and refrigerators" elsewhere in this encyclopedia, we will adopt the convention that "heat engines" (sometimes called prime movers) are thermodynamic devices that convert heat to work, and that "refrigerators" (including heat pumps) are devices that use work to transfer heat from low temperature to high temperature.

A. Simple examples

The Sondhauss tube, shown in Fig. 1(a), is the earliest reported thermoacoustic engine: Over 100 years ago, glassblowers noticed that when a small, hot glass bulb

was being blown on a cool glass tubular stem, the stem tip sometimes radiated sound; early investigators understood that the sound was generated by oscillatory thermal expansion and contraction of the air in the tube, which in turn was due to acoustic motion of the air toward and away from the hot end of the tube. The Sondhauss tube was a heat engine, converting heat into mechanical work in the form of sound. It utilized a predominantly standing acoustic wave, with the bulb and stem forming a resonator.

In the 1960's it was realized[CWS] that the performance of such devices could be greatly enhanced by inclusion of proper heat exchange components, especially what is now known as the "stack." The heat-driven thermoacoustic refrigerator shown in Fig. 1(b) illustrates these components, as well as the heat pumping/refrigerating function in thermoacoustics. The refrigerator's resonator consists of a cylindrical tube, closed at the top and opening at the bottom into a spherical bulb. Near the top is the heat-engine stack, which is an array of thin, uniformly spaced solid plates aligned parallel to the tube axis. Above the stack is the hot heat exchanger, consisting of well spaced conducting strips; below the stack, a similar set of strips forms a room-temperature heat exchanger, cooled by circulating water. This is the heat-engine assembly, producing acoustic work from heat: When enough heat is supplied to the hot heat exchanger by an external heat source, the gas oscillates spontaneously, maintaining a high-amplitude acoustic standing wave in the gas.

Just below the heat engine in Fig. 1b, another stack and pair of heat exchangers function as a refrigerator, which is driven by the acoustic work generated in the heat-engine stack. The refrigerator stack is identical to the engine stack, and its upper, room-temperature heat exchanger and lower, cold heat exchanger are made from copper strips. When the hot heat-exchanger temperature is high enough that the gas oscillates, the cold heat exchanger cools and absorbs an external heat load, as heat is pumped thermoacoustically from the cold heat exchanger to the room-temperature heat exchanger. Hence, the whole system functions as a refrigerator, with no moving parts, powered by heat delivered to the hot heat exchanger at high temperature.

Ceperley[Cep79,Cep85] has discussed thermoacoustic engines and refrigerators using traveling waves. One example of such a refrigerator is the orifice pulse-tube refrigerator[Rad90] shown in Fig. 1(c). In this refrigerator, an oscillating piston functions as a source of acoustic work at one end of the refrigerator, and an orifice and tank act as a sink of acoustic work at the other end. Between them, a regenerator, heat exchangers, and an open tube are filled with gas. The regenerator and adjacent heat exchangers function just like those of a Stirling-cycle refrigerator, and are the heart of this refrigerator. The gas is carried through four steps: compression, displacement downward, expansion, and displacement upward. The expansion of the gas at the cold end of the pulse tube cools the gas adiabatically, so that it can absorb heat in the cold heat exchanger when it is displaced upward. The oscillatory motion of the piston at the top, and the orifice and tank at the right, cause the gas to experience pressure oscillations and to move with the correct amplitude and phase relative to the pressure oscillations. The so-called "pulse tube" itself functions as a temperature buffer space between the cold heat exchanger and the warm orifice; its length is smaller than the gas displacement, so that the part of the gas which stays in the pulse tube throughout the cycle acts as a thermally insulating, compressible piston. With the hot heat exchangers anchored at room temperature, the cold heat exchanger of a pulse tube refrigerator reaches typically 50 K, and refrigeration is produced with a coefficient of performance of typically 1/4 of Carnot's.

B. Important length scales

In most thermoacoustic engines and refrigerators, the acoustic oscillation is essentially a pure tone, so that velocity, pressure, and temperature essentially vary sinusoidally with time, at the oscillation frequency f. The wavelength $\lambda = a/f$, where a is the sound speed, is an important length scale, especially in the direction of the gas displacement oscillations. For example, in the heat-driven refrigerator of Fig. 1b above, the distance from the hot end cap at the top of the tube to the

tube-bulb junction is approximately $\lambda/4$, so that the tube is essentially a quarter-wave acoustic resonator, with pressure maximum and velocity node at the top, and pressure node and velocity maximum at the entrance to the bulb. In thermoacoustic systems designed to date, 10 cm $\lesssim \lambda \lesssim$ 10 m. The lengths of stacks are much less than λ.

The amplitude x_1 of the gas displacement oscillations is a second important length scale. Typically, x_1 is everywhere much smaller than both the wavelength and the overall length of the system. Within the stack x_1 is typically smaller than the length of the stack, and in the heat exchangers x_1 is typically comparable to the length of the exchangers.

In the directions perpendicular to the displacement oscillations, the key length scale is the thermal penetration depth $\delta_\kappa = \sqrt{\kappa/\pi f \rho c_p}$, where κ is the thermal conductivity of the gas, ρ is its density, and c_p is its isobaric heat capacity per unit mass. The thermal penetration depth is approximately the distance that heat can diffuse through the gas in a time $1/\pi f$. Gas much farther than this from the nearest solid surface experiences adiabatic oscillations, and hence does not participate in thermoacoustic effects. Thus, the spacing in stacks and heat exchangers is of the order of δ_κ. In typical thermoacoustic systems, δ_κ is a fraction of a millimeter. In standing-wave thermoacoustic devices, a key geometrical requirement is that the smallest transverse dimension in the stack be roughly 1 to 3 times δ_κ; in traveling wave systems, the smallest transverse dimension must be much smaller than δ_κ.

Closely related to the thermal penetration depth is the viscous penetration depth $\delta_\nu = \sqrt{\mu/\pi f \rho} = \sqrt{\nu/\pi f}$, which is approximately the distance over which momentum can diffuse in a time $1/\pi f$. (Here μ is viscosity and ν is kinematic viscosity.) Thus, within δ_ν of solid surfaces, viscous shear causes gradients in the oscillatory velocity and displacement. The ratio $\delta_\nu/\delta_\kappa = \sqrt{\mu c_p/\kappa}$ is the square root of the Prandtl number, and is of order 1 for ideal gases. Hence, in general, viscous effects are important whenever thermoacoustic effects are important. In some of our heuristic discussion below, we will neglect viscosity; but it must be included when quantitatively accurate

results are desired.

II. Dissipative acoustic processes

To introduce the concepts and notation of thermoacoustics, we begin with a review of ordinary acoustic dissipation at solid boundaries, which arises from viscous and thermal contact. We consider sound propagating in the x direction in a duct with constant cross section. We adopt the usual complex notation for time-oscillatory quantities (pressure p, temperature T, velocity components u, v, and w, density ρ, entropy per unit mass s) :

$$p = p_m + \mathrm{Re}\left[p_1(x)e^{i\omega t}\right], \tag{1}$$

$$u = \mathrm{Re}\left[u_1(x,y,z)e^{i\omega t}\right], \tag{2}$$

$$T,\ \rho,\ s,\ \text{etc.} = \text{similar to } p, \tag{3}$$

$$v,\ w = \text{similar to } u, \tag{4}$$

$$\mu,\ \kappa,\ \text{etc.} = \text{constant}. \tag{5}$$

All the time dependence appears in the factor $e^{i\omega t}$, with $\omega = 2\pi f$. The mean values (subscript m) are real, but the small amplitudes (subscript 1) are in general complex to account for the time phasing of the oscillating quantities. We will assume that this expansion to lowest order in the acoustic amplitude suffices for all variables.

To develop quantitative understanding of viscous effects, we begin by finding the y and z dependence of the gas velocity, using the x-component of the momentum (Navier-Stokes) equation, for which the first-order approximation is

$$i\omega\rho_m u_1 = -dp_1/dx + \mu\nabla^2_{y,z}u_1\,. \tag{6}$$

With boundary condition $u_1 = 0$ at the solid surface, its solution is

$$u_1 = \frac{i}{\omega\rho}\left[1 - h_\nu(y,z)\right]\frac{dp_1}{dx}\,, \tag{7}$$

where $h_\nu(y,z)$ depends on the specific geometry under consideration. Below, we will need the spatial average (over y and z) of Eq. (7):

$$\langle u_1 \rangle = \frac{i}{\omega\rho}(1-f_\nu)\frac{dp_1}{dx}, \tag{8}$$

where f_ν is the spatial average of h_ν. Table 1 displays the functional forms[ABR91] of h_ν and f_ν, and Fig. 2 shows f_ν.

These functions describe how viscosity reduces the magnitude of the oscillatory velocity and shifts its phase. Gas which is much closer than δ_ν to the nearest solid surface is nearly at rest. Gas which is much farther than δ_ν from the nearest solid surface experiences essentially no viscous shear; it moves with a velocity that is independent of y and z. Gas which is roughly δ_ν from the nearest solid surface moves with a reduced velocity and significant phase shift. The acoustic power dissipation per unit volume, which is proportional to the square of $\nabla_{y,z}u_1$, is in general greatest at the solid surface where the velocity gradients are highest.

Similarly, to find the dependence of the oscillatory temperature on y and z, we begin with the general equation of heat transfer, for which the first-order approximation is

$$i\omega\rho_m c_p T_1 - i\omega T_m \beta p_1 = \kappa \nabla^2_{y,z} T_1, \tag{9}$$

where β is the thermal expansion coefficient. In general, the solid has sufficient heat capacity and thermal conductivity to enforce $T_1 = 0$ on the gas at the solid surface. Equation (9) is a first-order ordinary differential equation for $T_1(x,y)$, similar in form and boundary condition to Eq. (6) for $u_1(x,y)$. Exploiting the similarity, the solution can be written

$$T_1 = \frac{T_m\beta}{\rho_m c_p}[1-h_\kappa(y,z)]\,p_1, \tag{10}$$

and its spatial average can be written

$$\langle T_1 \rangle = \frac{T_m\beta}{\rho_m c_p}(1-f_\kappa)\,p_1, \tag{11}$$

where f_κ and h_κ are the same as f_ν and h_ν, but with δ_ν replaced by the thermal penetration depth δ_κ.

Similar to the effect of viscosity discussed above, thermal contact with the solid surface reduces the magnitude and shifts the phase of the oscillatory temperature. Gas which is much closer than δ_κ to the nearest solid surface is nearly isothermal. Gas which is much farther than δ_κ from the nearest solid surface is essentially adiabatic, experiencing adiabatic temperature oscillations $T_1 = (T_m\beta/\rho_m c_p)\, p_1$ in phase with the pressure oscillations. At approximately δ_κ from the nearest solid surface, these oscillations are reduced in magnitude and shifted in phase.

To find the time-averaged acoustic power $d\dot{w}$ dissipated in a length dx of the duct, we write

$$\frac{d\dot{w}}{dx} = -\int \frac{d\,(\overline{pu})}{dx}\, dy\, dz \tag{12}$$

where the overbar denotes time averaging and the integral is over the cross sectional area A of the duct. Rewriting Eq. (12) in complex notation and expanding the derivative gives

$$\frac{d\dot{w}_2}{dx} = -\frac{1}{2} A\ \mathrm{Re}\left(\langle \tilde{u}_1 \rangle \frac{dp_1}{dx} + \tilde{p}_1 \frac{d\,\langle u_1 \rangle}{dx} \right). \tag{13}$$

We can obtain dp_1/dx from Eq. (8) above. To find $d\,\langle u_1 \rangle /dx$, we will use our expression for $\langle T_1 \rangle$ in the continuity equation $\partial\rho/\partial t + \nabla \cdot (\rho \mathbf{v}) = 0$, which can be averaged with respect to y and z to obtain to first order

$$i\omega \langle \rho_1 \rangle + \rho_m\, d\,\langle u_1 \rangle /dx = 0. \tag{14}$$

Using $d\rho = -\rho\beta\, dT + (\gamma/a^2)\, dp$ (where γ is the ratio of isobaric to isochoric specific heats), we can express the spatially averaged density as $\langle \rho_1 \rangle = -\rho_m \beta \langle T_1 \rangle + (\gamma/a^2) p_1$ to lowest order. Substituting this into Eq. (14) and using Eq. (11) for $\langle T_1 \rangle$ yields

$$i\omega[1 + (\gamma - 1) f_\kappa]\, p_1 + \rho_m a^2\, d\,\langle u_1 \rangle /dx = 0 \tag{15}$$

as a first-order expression of the continuity equation which can be solved for $d\,\langle u_1 \rangle /dx$.

Finally, substituting Eqs. (15) and (8) into Eq. (13), we obtain

$$\frac{d\dot{w}_2}{dx} = \frac{A\omega}{2} \left[\frac{\rho_m\, |\langle u_1 \rangle|^2}{|1 - f_\nu|^2}\, \mathrm{Im}\,(-f_\nu) + \frac{(\gamma - 1)\, |p_1|^2}{\rho_m a^2}\, \mathrm{Im}\,(-f_\kappa) \right]. \tag{16}$$

The first term gives the viscous damping of sound and the second term gives the less intuitively obvious thermal damping. The work thus lost from the sound wave appears as heat, at the temperature T_m. In the boundary-layer approximation, in which all dimensions of the duct are much larger than the penetration depths, Eq. (16) reduces to

$$\frac{d\dot{w}_2}{dS} = \frac{1}{4}\rho_m \left|\langle u_1 \rangle\right|^2 \delta_\nu \omega + \frac{1}{4}\frac{|p_1|^2}{\rho_m a^2}(\gamma - 1)\,\delta_\kappa \omega, \tag{17}$$

where S is surface area of the duct. This expression is readily interpreted in terms of the acoustic energy density:

- The mean kinetic energy per unit volume $\rho\left|\langle u_1 \rangle\right|^2/4$ times δ_ν is roughly the kinetic energy in the gas within a viscous penetration depth of the solid surface, per unit area of the surface; the viscous dissipation term shows that this energy is dissipated by viscous shear at an average rate ω. Intuitively, we imagine layers of gas sliding relative to one another, with viscous "friction" between the layers turning mechanical energy into heat.

- Similarly, the thermal dissipation term shows that the mean adiabatic compressive energy density $|p_1|^2/4\rho_m a^2$ stored in unit volume of the gas is dissipated at an average rate $(\gamma - 1)\,\omega$ within a region of thickness δ_κ near the solid surface. The extra factor $(\gamma - 1)$ appears because this thermal dissipation is proportional to the difference between the isothermal and adiabatic compressibilities, which is proportional to $(\gamma - 1)$. Here, the intuitive picture is more subtle than for viscous dissipation. Gas immediately adjacent to the surface experiences isothermal density and pressure oscillations, which are reversible; gas far from the surface experiences adiabatic density and pressure oscillations, which are also reversible. We imagine the gas approximately δ_κ from the surface experiencing a complex cycle of density changes in response to the pressure oscillations: first an increase in density due to quasi-adiabatic compression by the sound wave, then a further increase in density as thermal contact with the surface removes heat from the gas, then a decrease in density due to quasi-adiabatic

expansion by the sound wave, and finally a further decrease in density as thermal contact with the surface delivers heat to the gas. Since this gas experiences thermal expansion at low pressure and thermal contraction at high pressure, it absorbs work from the sound wave. This gas, approximately δ_κ *from the surface*, is the most effective at absorbing work from the sound wave, whereas in the case of viscous dissipation the gas *at the surface* is most effective.

III. Lagrangian view of thermoacoustic processes

In the derivation of acoustic dissipation above, we adopted the usual Eulerian point of view of fluid mechanics, focusing attention on what happens at *a given point in space* as the gas moves by. Qualitative insight into thermoacoustic processes will be gained in this section by considering the alternative, Lagrangian point of view, focusing attention on the position, pressure, volume, and temperature of *a given parcel of gas* as it moves. It is simplest here to imagine square-wave motion and pressure in the gas, with pressure and displacement in phase for standing waves, and pressure and velocity in phase for traveling waves. However, real thermoacoustic systems operate sinusoidally, with a variety of phases between motion and pressure. Viscosity is ignored in this section, even though it has a significant, harmful effect on thermoacoustic efficiency. With these conceptual simplifications, the thermodynamic cycle for standing waves becomes identical to the Brayton cycle, with two reversible adiabats and two irreversible constant-pressure heat transfers, while the cycle for traveling waves becomes two isotherms and two reversible polytropic processes.

Time phasing is an important factor in the operation of traditional heat engines and refrigerators: Pistons, valves, and sparks must have the correct relative timing in order to produce the desired thermodynamic cycle in the working gas. Thermoacoustic systems often contain no obvious moving parts to perform these functions, yet the acoustic transport of heat and the generation (or absorption) of acoustic work point to some type of timed phasing of thermodynamic processes, here achieved in a

remarkably simple way.

One source of phasing in thermoacoustic systems is the presence of two thermodynamic media: gas and solid. As the gas oscillates along the solid at the acoustic frequency, it experiences oscillations in temperature. Part of the temperature oscillations comes from compression and expansion of the gas by sound pressure, and the rest is a consequence of the local temperature of the solid itself. In Section II above, the local solid temperature was constant. In this section, the temperature gradient along the stack ensures that each moving parcel of gas "sees" a *time-dependent* local solid temperature because of its motion along the temperature gradient.

In keeping with the square-wave assumption of this section, we need not retain complex notation to account for the phase of time-dependent variables. Hence, in this section only, all variables with subscript 1 are real and positive.

A. Standing-wave refrigerator

The standing-wave thermoacoustic refrigerator can be understood with reference to Fig. 3a, where the four steps of its cycle are displayed for one typical parcel of gas. In response to the standing wave, the gas parcel is pressurized and moves up along the adjacent solid in step 1, and is depressurized and moves down along the adjacent solid in step 3, absorbing and doing work dW''. In this picture, the pressure changes in steps 1 and 3 are adiabatic, so the gas parcel experiences adiabatic temperature changes proportional to the pressure changes. Steps 2 and 4 allow thermal equilibration between the gas parcel and the adjacent solid. In these two steps, net work $dW - dW'$ is absorbed from the sound wave by the gas parcel, as it thermally contracts at high pressure in step 2 and thermally expands at low pressure in step 4. If the sign of dT_m/dx is as shown (with higher T_m toward the nearest pressure antinode), the net thermal effect of the parcel of gas is to move heat $dQ \simeq dQ'$ a short distance along the solid, from slightly lower temperature to slightly higher temperature, during each cycle of the sound wave. Hence the system functions as a refrigerator, with acoustic

work being absorbed to pump heat from a lower temperature to a higher temperature.

If $dT_m/dx = 0$, the situation is identical to that described quantitatively in the section above (but with $\mu = 0$ here). In this case, $dW - dW'$ represents the thermal damping of sound, given by the second term in Eq. (16) or (17).

There is a so-called critical temperature gradient $\nabla T_{crit} \equiv (dT_m/dx)_{crit} = T_m\beta p_1/\rho_m c_p x_1$ at which no heat is transferred in steps 2 and 4. This occurs when the temperature rise due to adiabatic compression in step 1, $2T_m\beta p_1/\rho_m c_p$, exactly matches the rise in adjacent solid temperature that the gas sees because of its motion along the solid, $2x_1\nabla T_{crit}$. For all temperature gradients less than ∇T_{crit}, heat transfer between gas and solid has the signs shown in Fig. 3, with heat flow dQ from gas to solid in step 2, at the higher temperature and position, and heat flow dQ' from solid to gas in step 4, at the lower temperature and position. For such low temperature gradients it is also true that $dW > dW'$, so that the gas absorbs work from the standing wave. Hence, a standing-wave system acts as a refrigerator only if $dT_m/dx < \nabla T_{crit}$.

In Fig. 3a, we label the parcel's distance from the solid as $\sim \delta_\kappa$ because imperfect thermal contact between gas and solid is required for the cycle to operate as described. If thermal contact between gas and solid were perfect, the gas would arrive at its uppermost position at the end of step 1 already equilibrated at the local solid temperature, so that no heat would be transferred during step 2; and similarly for steps 3 and 4. At the other extreme, if thermal contact were nonexistent, no heat could be transferred in steps 2 and 4. Thus, intermediate, *imperfect* thermal contact between gas and solid is required for the cycle to operate as described. For the sinusoidal motion and pressure of real systems, this imperfect thermal contact is achieved for gas neither much closer nor much farther than δ_κ from the nearest solid. As suggested by the thermal dissipation term in Eq. (16), the quantitative measure of the correct geometry is $\mathrm{Im}(-f_\kappa)$. Thus, effective thermoacoustic systems contain stacks which fill the entire cross-section of the resonator, with spacings of a few thermal penetration depths, as shown in Fig. 1(b); and with $\mathrm{Im}(-f_\kappa)$as large as possible, suggesting (c.f. Fig. 2) that parallel-plates and pin arrays are the best stack

geometries.

B. Standing-wave engine

To understand the conversion of heat to acoustic work in a standing-wave thermoacoustic engine, consider the magnified view of part of the stack shown in Fig. 3b, which shows a typical parcel of gas at four instants of time during one cycle of the acoustic wave. The standing wave carries the parcel up and down, compressing and expanding it, with phasing such that it is at its most compressed state when at its upper position, and its most expanded state at its lower position. Here, the solid temperature gradient is greater than ∇T_{crit}, so the directions of heat transfer in steps 2 and 4 are reversed relative to the standing-wave refrigerator. Thus, the net thermal effect of the parcel of gas shown is to move a little heat down the solid, from higher to lower temperature; again, all parcels have a similar effect, so that overall a substantial amount of heat is carried down along the solid from the hot heat exchanger to the cold heat exchanger.

Since heat flows into the gas in step 2, and out of it in step 4, the gas expands at high pressure in step 2 and contracts at low pressure in step 4. Thus, the net effect in the thermoacoustic heat engine is the production of work, at the acoustic frequency, by this and all other gas parcels. One can imagine all the gas in the stack region of a thermoacoustic engine thermally expanding and contracting because of its displacement along the temperature gradient, with the correct phasing relative to the oscillatory standing-wave pressure to do net work. In the Sondhauss tube, for example, this net work maintains the standing wave against viscous and thermal dissipation in the resonator, and provides power to radiate into the room.

C. Traveling-wave systems

Ceperley[Cep79,Cep85] introduced the idea that Stirling engines and refrigerators, and related devices such as pulse-tube refrigerators, may be regarded as traveling-wave

thermoacoustic systems. This point of view is possible because the phasing between displacement and pressure in such engines and refrigerators is substantially that in traveling waves, with displacement and pressure approximately 90° out of phase. A second difference between standing- and traveling-wave systems is the typical spacing between solid elements: In traveling-wave systems, this distance is typically $\ll \delta_\kappa$, ensuring excellent thermal contact between gas and solid at all times.

These features are illustrated for the traveling-wave refrigerator in Fig. 4. Isothermal compression in step 2 transfers heat from the gas parcel to the solid at a higher temperature, and isothermal expansion at step 4 transfers heat from solid to gas at a lower temperature. The other two steps serve to preheat and precool the gas, reversibly storing and recovering heat in the solid. The net effect of one parcel over an entire cycle is, again, to move a little heat a short distance up the solid, up the temperature gradient. Excellent thermal contact between gas and solid is required in steps 1 and 3 for effective preheating and precooling respectively; in realistic, sinusoidal traveling-wave systems, this is achieved by maintaining transverse dimensions much smaller than a thermal penetration depth.

Traveling wave engines are also possible in principle; from the perspective of thermoacoustics, all Stirling engines can be considered as such in the low-amplitude limit.

IV. Components

The first efficient thermoacoustic refrigerator,[Hof88] designed, built, and studied by Tom Hofler, is shown in Fig. 5a. It illustrates several features of many of today's standing-wave thermoacoustic devices. Figure 5b shows some of the data obtained with this refrigerator, which had a cooling power of several watts with acoustic pressure amplitudes of 3% of the mean pressure.

A. Stacks and regenerators

The heart of a thermoacoustic system is its stack (or, in the case of a traveling-wave system, its regenerator). For standing-wave systems, Fig. 2 indicates that parallel plates and pin arrays are the best stacks known, because they have the largest values of $\mathrm{Im}(-f_\kappa)$. As yet there have been no experimental studies using pin arrays, so the parallel plates are a preferred geometry. Figure 2 suggests that the spacing should be about $2.2\delta_\kappa$, although viscous and system-wide effects often modify the optimum significantly.

For ease of construction and structural rigidity, parallel plate stacks have usually been spirally wound. The stack in Fig. 5 was a plastic spiral wound around a small plastic rod, with numerous tiny, parallel plastic rods serving as spacers between layers of the spiral. Hence the stack channels were not ideal parallel-plate; they were essentially rectangular channels with 13:1 aspect ratio, and with slight curvature. Spiral stacks have also been built with spacing determined by internal bumps throughout the stack or by radial ribs at the ends of the stack.

Typically stack materials are thin enough to block only 10-20% of the total cross-sectional area of the stack. Generally any stack material rigid enough to support its own weight and survive routine handling has sufficient heat capacity to enforce the $T_1 = 0$ boundary condition at the surface. (Liquid-metal thermoacoustics[MS88] is a rare exception.) Hence the key requirement on the stack material is a low thermal conductivity, lest heat be wastefully conducted along the stack from hot to cold. Near and below room temperature, plastics have generally been used; above room temperature, stainless steel, nickel alloys, and ceramics have been used.

Other stack geometries which have been tested include metal honeycomb with hexagonal cells and ceramic honeycomb with square cells (from automotive catalytic converters).

In traveling-wave systems, for which the "stack" is known as a regenerator, the requirement that the spacing be $\ll \delta_\kappa$ makes it difficult to construct parallel-plates

with uniform spacing. Hence, enhanced heat-transfer surfaces such as screen beds or intricately etched foils[MYS+96] are generally used. These geometries are not discussed here, but they can be treated by thermoacoustic methods.[WS94]

B. Resonators

A resonator is required for standing-wave thermoacoustic systems. It should satisfy many requirements, which we illustrate with Fig. 5. Most fundamentally, it should have the desired resonance frequency, and permit location of the stack at a location with the desired ratio of $|p_1|/|\langle u_1\rangle|$. The resonator of Fig. 5 is essentially a quarter-wavelength resonator (even though $p_1(x)$ is not exactly a cosine), with the pressure antinode at the driver piston and the pressure node near the sphere. The stack is located close to the pressure antinode, where $|p_1|$ is nearly equal to its value at the pressure antinode and $|\langle u_1\rangle|$ is significantly smaller than it is at the pressure node. (In the next section we will discuss this choice of location in terms of efficiency.)

A conceptually simpler quarter-wavelength resonator would comprise a uniform-diameter tube $\lambda/4$ long connected to an essentially infinite sphere. The resonator of Fig. 5 differs from this in important, advantageous ways. First, the tube consists of two sections of different diameters. The small-diameter section near the pressure node allows the tube to be significantly shorter than $\lambda/4$, reducing the overall size and weight. More subtly, this configuration also reduces the dissipation of acoustic power in the resonator, which is governed by Eq. (17). Basically, as the diameter of the lower section of the tube is reduced, its surface area decreases while $|\langle u_1\rangle|$ increases and $|p_1|$ remains about the same. For a modest reduction in diameter, as in Fig. 5, the net effect is a reduction in the total dissipation, by as much as a factor of 2.

The resonator of Fig. 5 also suppresses harmonic content in the standing wave. In this regard, a uniform-diameter resonator is extremely undesirable, because its higher resonant modes occur at frequencies that are exact integer multiples of the fundamental. In this situation, nonlinear effects drive resonances, and hence the har-

monic content of the wave is high. These nonlinear effects, such as the $(\mathbf{v} \cdot \nabla)\,\mathbf{v}$ term in the momentum equation (neglected elsewhere in this article), generate oscillations at integer multiples of the operating frequency. The resonance frequencies of the first few modes of the resonator of Fig. 5 are not integer multiples, and so higher-mode resonances are avoided.

Choice of sphere diameter in Fig. 5 was a compromise between size/weight and dissipation: A smaller sphere experiences larger pressure oscillations and requires a longer tube to maintain the desired resonance frequency. The tapered connection between the sphere and the small-diameter section reduces the Reynolds number at the sphere opening, in order to minimize dissipation due to turbulence.

Another typical standing-wave thermoacoustic geometry is essentially half-wavelength, with mirror-image stacks near the pressure antinodes at each end of the resonator. The central portion of such a resonator generally has a reduced diameter, to exploit the reduced size, dissipation, and harmonic content described above. Half-wavelength resonators are sometimes bent or coiled to reduce overall size.

With the requirement that $\delta_\kappa \gg$ spacing, traveling-wave systems typically operate at lower frequencies than standing wave systems, pushing conventional acoustic resonators to impractical lengths. Hence, traveling-wave systems often incorporate a mechanical element into the resonator, providing resonance with solid mass bouncing on a gas spring. Such "free-piston" Stirling machines are fully described elsewhere.[UB84]

C. Electroacoustic power transducers

To date, thermoacoustic systems have often used loudspeaker-like transducers, with permanent magnets and coils of wire in relative oscillatory motion. Piezoelectric and magnetohydrodynamic transducers are also candidates.

In Fig. 5, the loudspeaker-like driver was located at a pressure antinode of the standing wave, so that the acoustic power was delivered with high force and small

displacement, easing engineering difficulties associated with the flexing portion of the driver. This location also placed it at room temperature, where Joule heat generated in the driver could be removed most efficiently.

Commercial loudspeakers are often poorly matched to thermoacoustic applications, because they are designed to drive a comparatively low acoustic impedance with large bandwidth. For better impedance matching, the loudspeaker in Fig. 5 was modified, with the addition of an aluminum cone and associated flexing seal to exert the speaker's force over a small area. With this modification, it converted electric power to acoustic power with approximately 20% efficiency. Further modifications to commercial loudspeakers can achieve efficiencies of at least 50%. In the field of Stirling machines, linear motors and linear alternators of 90% efficiency have been demonstrated,[CY92] indicative of the performance to be expected when loudspeaker-like transducers are designed specifically for efficient conversion of power at a single frequency.

D. Heat exchangers

Heat exchangers are required in thermoacoustic engines and refrigerators to supply and extract heat at the two ends of the stack or regenerator. The length of each along the direction of the acoustic oscillation is of the order of the local displacement amplitude x_1; transverse dimensions are roughly of the order of those of the adjacent stack or regenerator. Equation (16) can be used to estimate the acoustic power dissipation in the heat exchangers, although entrance effects prevent quantitatively accurate predictions.

In Fig. 5 and other small thermoacoustic systems, the heat exchangers can simply consist of multiple copper strips spanning the resonator at the stack ends. Larger-diameter, high-power systems require integral pumped-fluid channels, heat pipes, or electric resistance heaters imbedded in the heat exchangers to deliver or extract heat uniformly from the stack ends. In this case, the typical geometry is like a car radiator,

with a set of high-conductivity fins extending from pipes; the assembly typically has 50-75% of the cross-sectional area open to gas flow.

V. Power and efficiency

A thermoacoustic engine converts heat into acoustic work. More specifically, it absorbs heat Q_H from a heat source at high temperature T_H, rejects heat Q_C to a heat sink at low temperature T_C, and produces work W. The first law of thermodynamics determines that $W + Q_C = Q_H$; the second law shows that the efficiency W/Q_H is bounded above by the Carnot efficiency $1 - T_C/T_H$.

Similarly, a thermoacoustic refrigerator uses acoustic work W to absorb heat Q_C from at a low temperature T_C, rejecting waste heat Q_H to a heat sink at higher temperature T_H. The first law of thermodynamics determines that $W + Q_C = Q_H$; the second law shows that the relevant efficiency, known as the coefficient of performance $COP = Q_C/W$, is bounded above by $COP_{\text{Carnot}} = T_C/(T_H - T_C)$.

The power (here, heat per unit time) of standing-wave thermoacoustic devices can be estimated roughly by $p_m a A \, |p_1/p_m|^2 \, /20$, with A the cross sectional area of the stack. Thus, high amplitude (as a fraction of mean pressure), high mean pressure, and high sound speed increase power density. Heat flux densities of the order of 1 MW/m^2 are easily achievable in high-pressure helium gas, which has the highest sound speed of the inert gases, so it is often used, typically at a pressure above 10 bar. Helium's high thermal conductivity, the largest of the inert gases', is an additional advantage, allowing spacings within the stack and heat exchangers to be conveniently large. In today's thermoacoustic systems, $|p_1/p_m|$ ranges from 0.03 to 0.10, chosen as a compromise between the high power density achieved at high amplitude and the high confidence in the quantitative accuracy of calculations (described in the next Section) for systems operating at low amplitude.

Calculations show that standing-wave thermoacoustic systems have second-law efficiencies (i.e., efficiencies normalized by the Carnot efficiency) as high as 0.5, con-

sidering effects in the stack only. Including resonator, heat exchanger, and transducer losses reduces this efficiency to between 0.15 and 0.35 for large-cross-section systems, for which resonator losses are not a significant fraction of the total power. Small-diameter systems can have even lower efficiencies, due to resonator losses, auxiliary equipment (such as circulating liquid loops for heat transfer to air streams) and heat losses through insulation to ambient temperature. Traveling-wave thermoacoustic systems can have higher efficiencies.

To date, the most efficient laboratory standing-wave engine and refrigerator have had second-law efficiencies of 0.2. This is impressive for devices with no moving parts, and is comparable to the efficiencies of small, inexpensive commercial equipment. It falls far short of the efficiencies of well engineered, expensive steam turbines or large-scale vapor-compression refrigeration equipment, for which over 80% of Carnot has been achieved. If future inventions and improvements to basic understanding can improve the efficiency or raise the power density of thermoacoustic engines and refrigerators, without sacrificing their simplicity, they will find more widespread use.

The efficiency of standing-wave thermoacoustic devices falls below Carnot's efficiency because of five major sources of irreversibility, which we call "inherent" loss, viscous loss, conduction loss, auxiliary losses, and transduction loss:

- First, inherent loss arises from the heat transfer to and from each parcel of gas in the stack as described in Figs. 3 and 4: Whenever heat dq is transferred across a nonzero temperature difference δT, the entropy of the universe increases[Bej82] by $dq\,\delta T/T^2$. This irreversibility is unavoidable in the standing-wave thermoacoustic process, relying as it does on *imperfect* thermal contact for the correct phasing between pressure oscillations and thermal expansion/contraction.

- Second, viscous loss in the stack occurs because work must be expended overcoming viscous shear forces as the gas oscillates between the plates. The viscous penetration depth is nearly as large as the thermal penetration depth, so most of the gas in the stack experiences significant viscous shear (with the parcels in

Figs. 3 and 4 actually experiencing significant shape distortions).

- Third, simple conduction of heat from the hot heat exchanger to the cold heat exchanger through the stack material and the gas reduces system efficiency.

- Fourth, the first three losses occur in auxiliary parts of a thermoacoustic system: Viscous and inherent losses in the heat exchangers, and conduction loss in the portion of the resonator case surrounding the stack, are usually the most important auxiliary losses in large systems, while viscous and inherent losses on other surface area in the resonator are important in small systems.

- Fifth, with rare exceptions such as shown in Fig. 1b, electric power—not acoustic power—is the usual interface to external systems, so electroacoustic transducers are needed. In principle, these can be 100% efficient; in practice, 90% appears possible, while 50% has been achieved to date in laboratory thermoacoustic systems.

For many designs with high power density, the first four sources of irreversibility contribute roughly equally to the inefficiency of thermoacoustic devices. Some reasons for this balance are easy to understand. First, thermoacoustic systems have inherent trade-offs between power density and efficiency, due to the inherent losses in the stack. In the inviscid picture, at the critical temperature gradient there are no inherent losses, but there is no work done or dissipated, and no heat transported; the difference between dT_m/dx and ∇T_{crit} controls both power and inherent loss. Thus, for a given temperature difference across which the engine or refrigerator must operate, the choice of the length of the stack is essentially a compromise between power density and efficiency. Often $dT_m/dx \sim 2\nabla T_{crit}$ for engines and $\sim \nabla T_{crit}/2$ for refrigerators. Second, stack location in the standing wave controls the trade-off between viscous losses, which are highest when the stack is near the pressure node, and conduction losses, which are highest near the pressure antinode where ∇T_{crit} is steepest. Usually this trade-off leads to a stack position much closer to the pressure antinode than to

the node; the proportions shown in Fig. 5 are typical.

One way to increase efficiency is to use a mixture of helium and a heavier inert gas as the working substance. The Prandtl number $\sigma = \mu c_p/\kappa = \delta_\nu^2/\delta_\kappa^2$ is a dimensionless measure of the ratio of viscous to thermal effects in fluids; lower Prandtl numbers give higher efficiencies in thermoacoustics. Kinetic theory predicts $\sigma = 2/3$ for hard-sphere monatomic gases, and indeed real monatomic gases have values very close to this (e.g., $\sigma = 0.68$ for helium at room temperature). Fortunately, mixtures of a heavy and a light monatomic gas have Prandtl numbers significantly lower than 2/3. The most efficient thermoacoustic refrigerator mentioned above[GPG94] used 89% helium, 11% xenon, with $\sigma = 0.27$, to achieve its highest efficiency of 20% of Carnot (cf. 17% of Carnot in Fig. 5, for a similar apparatus optimized for pure helium gas). However, with a sound speed less than half that of pure helium, the power density using this gas mixture was reduced.

VI. Methods of computation

Although attempts at theoretical understanding of these phenomena began with Lord Rayleigh 120 years ago, a quantitatively accurate theory of thermoacoustics was not developed until the 1970s, through the efforts of Nicholas Rott and coworkers.[Rot80] It is remarkable that such qualitatively simple classical phenomena went without a quantitatively accurate explanation until late in this century.

A. Rott's approach

Quantitative thermoacoustics analysis is based on Rott's wave equation and energy-flux equation. These equations result from the momentum, continuity, and energy equations in the acoustic approximation. The results are valid for arbitrary phasing between p_1 and $\langle u_1 \rangle$, in particular for both standing-wave and traveling-wave systems.

As with simpler one-dimensional acoustic wave equations, Rott's wave equation may be written as two coupled first-order differential equations in the oscillatory pressure $p_1(x)$ and the spatially averaged oscillatory velocity $\langle u_1(x) \rangle$. Rott's energy-flux equation may be regarded as a first order differential equation in mean temperature $T_m(x)$, coupled to the pressure and velocity equations. In this form, displayed in Table 2, Rott's equations are suitable for numerical integration along x.

The equations are based on the assumption that the relevant variables can be written

$$p = p_m + \mathrm{Re}\left[p_1(x)e^{i\omega t}\right], \tag{18}$$

$$u = \mathrm{Re}\left[u_1(x,y,z)e^{i\omega t}\right], \tag{19}$$

$$v,\, w = \text{similar to } u, \tag{20}$$

$$T = T_m(x) + \mathrm{Re}\left[T_1(x,y,z)e^{i\omega t}\right], \tag{21}$$

$$\rho,\, s = \text{similar to } T, \tag{22}$$

$$\mu = \mu(x), \tag{23}$$

$$\kappa,\ a,\ c_p,\ \text{etc.} = \text{similar to } \mu \tag{24}$$

We are most interested in T_m, p_1, and $\langle u_1 \rangle$.

For clarity, we consider here only the case of large solid heat capacity ($\epsilon_s = 0$ in Swift 1988), so that the temperature of the solid material in the stack is simply $T_m(x)$, independent of time, y, and z.

The expression for dp_1/dx shown in Table 2 is identical to Eq. (8) derived in Section II; the presence of nonzero dT_m/dx has no effect on the first-order momentum equation. In effect, we regard this approximation to the momentum equation as the origin of pressure gradient in thermoacoustics: The moving gas causes the pressure gradient. If $f_\nu = 0$, the pressure gradient is entirely "inertial", but when $f_\nu \neq 0$ the presence of viscosity and the stationary boundaries adds a resistive component to the pressure gradient and also effectively changes the magnitude of the gas' inertial contribution to dp_1/dx.

In both the continuity and energy equations an expression for the spatial average over y and z of the oscillatory temperature, $\langle T_1 \rangle$, is required. As in Section II, it is derived from the general equation of heat transfer, for which the first-order approximation is

$$\rho_m c_p \left(i\omega T_1 + u_1 \, dT_m/dx\right) - i\omega T_m \beta p_1 = \kappa \nabla_{y,z}^2 T_1 \tag{25}$$

when $dT_m/dx \neq 0$. Following the same procedure as in Section II, we find

$$T_1 = \frac{T_m \beta}{\rho_m c_p} (1 - h_\kappa) \, p_1 + \frac{i}{\omega (1 - f_\nu)} \frac{dT_m}{dx} \left(1 - \frac{h_\kappa - \sigma h_\nu}{1 - \sigma}\right) \langle u_1 \rangle \tag{26}$$

and

$$\langle T_1 \rangle = \frac{T_m \beta}{\rho_m c_p} (1 - f_\kappa) \, p_1 + \frac{i}{\omega (1 - f_\nu)} \frac{dT_m}{dx} \left(1 - \frac{f_\kappa - \sigma f_\nu}{1 - \sigma}\right) \langle u_1 \rangle \tag{27}$$

[c.f. Eqs. (10) and (11)]. Using Eq. (27) in the first-order continuity equation yields

$$\frac{d \langle u_1 \rangle}{dx} = -\frac{i\omega}{\rho_m a^2} [1 + (\gamma - 1) f_\kappa] \, p_1 + \frac{\beta (f_\kappa - f_\nu)}{(1 - f_\nu)(1 - \sigma)} \frac{dT_m}{dx} \langle u_1 \rangle \, , \tag{28}$$

which is also shown in Table 2.

This expression finds easy physical interpretation as a first-order approximation to the continuity equation, expressing the fact that $d \langle u_1 \rangle / dx$ is nonzero if and only if there is a density oscillation in the gas between x and $x + dx$. The two terms in Eq. (28) show that such density change can be caused either by pressure or by velocity along the temperature gradient. Consider the pressure term first. If $f_\kappa = 0$, there is no thermal contact between gas and solid, so the density oscillations are adiabatic; in this case, ρa^2 is the correct compressibility. If $f_\kappa = 1$, the thermal contact between gas and solid is perfect, so the gas is anchored at the local solid temperature. In this case, $\rho a^2/\gamma$, the isothermal compressibility, is appropriate. For intermediate thermal contact, an effective compressibility $\rho_m a^2/[1 + (\gamma - 1) f_\kappa]$, intermediate in magnitude and with nontrivial phase, describes the density oscillations in response to pressure oscillations. Next consider the velocity term, in the easily interpreted inviscid limit, with $f_\nu = 0$ and $\sigma = 0$. If $f_\kappa = 0$, there is no thermal contact between gas and solid, the velocity term is zero, and the gas density does not respond to the

temperature gradient. In the other extreme, if $f_\kappa = 1$, the gas is always at the local solid temperature, so that as it flows toward higher T_m its density decreases and its velocity increases. In the more interesting, intermediate regime, oscillatory motion of the gas along the temperature gradient leads to complex density oscillations.

At this point, we could combine Eqs. (8) and (28), eliminating $\langle u_1 \rangle$, to obtain a second-order differential equation in p_1, Rott's wave equation:

$$[1 + (\gamma - 1) f_\kappa] p_1 + \frac{\rho_m a^2}{\omega^2} \frac{d}{dx} \left(\frac{1 - f_\nu}{\rho_m} \frac{dp_1}{dx} \right) - \beta \frac{a^2}{\omega^2} \frac{f_\kappa - f_\nu}{1 - \sigma} \frac{dT_m}{dx} \frac{dp_1}{dx} = 0 . \qquad (29)$$

However, for numerical computations, it is easiest to use Eqs. (8) and (28) separately.

Finally, the energy equation yields an expression for dT_m/dx. To lowest order, the total energy flux[Rot80,Swi88] is

$$\begin{aligned} \dot{H}_2 &= A_{\text{gas}} \langle \rho_m \overline{u_1 w_1} \rangle - (A_{\text{gas}} \kappa + A_{\text{solid}} \kappa_{\text{solid}}) \, dT_m/dx \qquad (30) \\ &= A_{\text{gas}} \rho_m c_p \langle \overline{T_1 u_1} \rangle + A_{\text{gas}} (1 - T_m \beta) \overline{p_1 \langle u_1 \rangle} - (A_{\text{gas}} \kappa + A_{\text{solid}} \kappa_{\text{solid}}) \, dT_m/dx , \end{aligned}$$

where w is the enthalpy per unit mass of the gas and the overbar denotes time average. Using Eqs. (7) and (27) for u_1 and T_1 respectively, and taking the spatial averages of the products as indicated in Eq. (30), we find

$$\begin{aligned} \dot{H}_2 &= A_{\text{gas}} \frac{T_m \beta}{2} \text{Re} \left[p_1 \langle \tilde{u}_1 \rangle \left(1 - \frac{\left(f_\kappa - \tilde{f}_\nu \right)}{(1 + \sigma) \left(1 - \tilde{f}_\nu \right)} \right) \right] \qquad (31) \\ &\quad + \frac{A_{\text{gas}} \rho_m c_p |\langle u_1 \rangle|^2}{2\omega(1 - \sigma^2) |1 - f_\nu|^2} \text{Im} \left(f_\kappa + \sigma \tilde{f}_\nu \right) \frac{dT_m}{dx} \\ &\quad + A_{\text{gas}} \frac{(1 - T_m \beta)}{2} \text{Re} \left[p_1 \langle \tilde{u}_1 \rangle \right] - (A_{\text{gas}} \kappa + A_{\text{solid}} \kappa_{\text{solid}}) \frac{dT_m}{dx} . \end{aligned}$$

The time-averaged energy flux $\dot{H}_2$ along the stack must be independent of x, because the case surrounding the stack is assumed to be rigid and thermally insulated and the system is in time-averaged steady state. With $\dot{H}_2$ constant, we solve Eq. (31) for dT_m/dx as shown in Table 2, using it to predict the evolution of T_m.

Because the first two equations in Table 2 are complex, these 3 coupled equations represent 5 real coupled equations in the 5 variables $\text{Re}[p_1]$, $\text{Im}[p_1]$, $\text{Re}[\langle u_1 \rangle]$, $\text{Im}[\langle u_1 \rangle]$,

and T_m, and as such are easily integrated numerically using, for example, a Runge-Kutta method. These results are the basis of all present quantitative analyses of standing-wave systems, and are equally valid for traveling-wave systems.

The curves in Fig. 5b, in reasonable agreement with the data, were calculated using publicly available software[WS94] based on this theory. The software combines the thermoacoustic computations with similar acoustic solutions for the other portions of the resonator, using continuity of T_m, p_1, and complex volumetric velocity $A\langle u_1\rangle$ at the junctions between ducts and thermoacoustic components. The calculations have no adjustable parameters; they simply use the geometry of the apparatus and the properties of helium gas.

B. Approximate methods

The thermoacoustic calculation method summarized in the previous subsection is suitable for numerical integration. In this subsection we present approximate methods, suitable for making estimates of standing-wave system performance on the blackboard, with a hand calculator, or in a spreadsheet. An assumption common to all such approximate methods[Swi88Atc92] is that the stack is short enough and the temperature spanned is small enough that all $x-$ and T_m-dependent variables can be regarded as constant, including dT_m/dx itself. For example, with this approximation the pressure drop across the stack would be estimated by $\Delta p_1 = -i\omega\rho_m \langle u_1\rangle \Delta x/(1-f_\nu)$, where Δx is the stack length, with ρ_m, $\langle u_1\rangle$, and f_ν evaluated at the stack midpoint. For further simplification, the boundary-layer limit can be used for f_ν and f_κ with little additional loss in accuracy.

An approximate expression for the work absorbed or produced in the stack can be obtained by following the same procedure as was used to derive Eq. (16), but including the $\langle u_1\rangle$ term in Eq. (28). The result is

$$\Delta\dot{w}_2 = \frac{A_{\rm gas}\omega}{2}\Delta x\left[\frac{\rho_m|\langle u_1\rangle|^2}{|1-f_\nu|^2}\,\mathrm{Im}\,(-f_\nu) + \frac{(\gamma-1)|p_1|^2}{\rho_m a^2}\,\mathrm{Im}\left(-f_\kappa + \frac{f_\kappa - f_\nu}{(1-\sigma)(1-f_\nu)}\frac{dT_m/dx}{\nabla T_{crit}}\right)\right] \tag{32}$$

where $\nabla T_{crit} = T_m \beta \omega p_1 / i \rho_m c_p \langle u_1 \rangle$. This equation shows that a modest value of dT_m/dx reduces the power dissipated in the stack, and that a large enough value of dT_m/dx changes the sign of $\Delta \dot{w}_2$, so that the stack produces work. For $\mu = 0$, the crossover would occur at $dT_m/dx = \nabla T_{crit}$.

Equations (31) and (32) can be further simplified by any of the following assumptions: the boundary-layer limit expressions for f_κ and f_ν, exact standing-wave phasing, or ideal-gas equation of state. With all those assumptions simultaneously, simple approximate expressions result:

$$\dot{H}_2 = \frac{A_{\text{gas}}}{4} \frac{\delta_\kappa}{r_h} \frac{|p_1| |\langle u_1 \rangle|}{(1+\sigma)\, e} \left(1 + \sqrt{\sigma} - \delta_\nu / r_h - \Gamma \frac{1 + \sqrt{\sigma} + \sigma}{1 + \sqrt{\sigma}} \right) - \left(A_{\text{gas}} \kappa + A_{\text{solid}} \kappa_{\text{solid}} \right) \frac{dT_m}{dx} \tag{33}$$

$$\Delta \dot{w}_2 = \frac{A_{\text{gas}} \Delta x}{4 r_h} \left[\frac{\rho_m \left| \langle u_1 \rangle \right|^2 \delta_\nu \omega}{e} + \frac{(\gamma - 1) \left| p_1 \right|^2 \delta_\kappa \omega}{\gamma p_m} \left(1 - \frac{\Gamma}{(1 + \sqrt{\sigma})\, e} \right) \right], \tag{34}$$

where

$$\Gamma = (dT_m/dx) \; / \nabla T_{crit}, \tag{35}$$

$$\nabla T_{crit} = \frac{\omega \left| p_1 \right|}{\rho_m c_p \left| \langle u_1 \rangle \right|}, \tag{36}$$

$$e = 1 - \delta_\nu / r_h + \delta_\nu^2 / 2 r_h^2. \tag{37}$$

These expressions are easily used in spreadsheets or for hand calculations.

C. Similitude

The most promising route to higher power densities is by increasing $|p_1/p_m|^2$, but doing so takes us farther from the range of applicability of Rott theory, which is based on small oscillatory amplitude. Rott's assumptions include: gas displacement amplitude much smaller than the length of the stack and other components, Reynolds number of the oscillations small enough to avoid turbulence, and $|p_1| \ll p_m$. Extension of thermoacoustics theory beyond these assumptions will be challenging. In the meantime, application of similitude to thermoacoustics[OS94] has shown how to organize and correlate experimental data outside the range of our current quantitative under-

standing, and allows meaningful experimental study of scale models of thermoacoustic apparatus reminiscent of wind-tunnel models of aircraft.

The similitude analysis shows that the dimensionless variable groups in Table 3 provide a *complete* description of the gas in a thermoacoustic system. (Additional variables can be added to include transducers or solid properties if needed.) The analysis is powerful because it is true even when the thermoacoustic system is operating beyond the assumptions inherent in Rott's method. For example, it "includes" turbulence, even though the Reynold's number is not explicitly present in the table.

Figure 6, from the similitude study,[OS94] illustrates both the power of similitude analysis and the amplitude dependence of the accuracy of our current understanding of thermoacoustic phenomena. The data shown as points were taken with the apparatus sketched in the inset, a thermoacoustic engine driving no load. The plots can be interpreted as displaying what heater power Q_H and hot temperature T_H were required at the hot heat exchanger to maintain steady oscillations at a given pressure amplitude p_1.

The lines are the results of calculations[WS94] based on Rott's theory, the geometry of the apparatus, and the properties of the gases. The calculations agree well with the measurements in the limit of small p_1, as expected from the assumptions in the theory. However, at $p_1/p_m \sim 0.1$ measurements deviate from calculations significantly. The deviations are not surprising, in that they are of the same order as p_1/p_m, but they are disturbing from a practical point of view, because both Q_H and T_H deviate from Rott theory in directions which decrease the efficiency.

Three monatomic gases were used in the measurements, with different average pressures selected to make δ_κ the same for all three cases, ensuring similarity. The crux of the similitude analysis is that the only relevant thermophysical length scales are λ, δ_κ, and δ_ν, so that if these are held fixed all other dimensionless groups will also stay fixed. The data for all three gases do indeed fall along the same curves, demonstrating similarity. Although the similitude analysis does not indicate the cause of the deviations of the measurements from Rott-style calculations, it does allow

elimination of some potential suspects, such as gravity-driven convection, resonator vibration, and stack flutter, which are not included in the similitude analysis and would not be expected to scale from gas to gas.

Similitude also permits the building of meaningful thermoacoustic scale models. For example, to model a full-size helium thermoacoustic system at mean pressure p_m, we can use a half-scale model with argon at $0.79p_m$, the pressure which ensures that δ_κ in the argon is exactly half its value in the full-size helium system. Experiments with the scale model can be easier than those with the full-size system because of its smaller size and because powers are reduced by $(p_m a A)_{\mathrm{Ar}} / (p_m a A)_{\mathrm{He}} = 1/16$.

VII. Glossary

Ed.: I hope the editors or reviewers will suggest what terms should be included in the glossary. I will then provide definitions.

VIII. Research Bibliography

[ABR91] W. P. Arnott, H. E. Bass, and R. Raspet. General formulation of thermoacoustics for stacks having arbitrarily shaped pore cross sections. *J. Acoust. Soc. Am.*, 90:3228, 1991.

[Atc92] A. A. Atchley. Standing wave analysis of a thermoacoustic prime mover below onset of self-oscillation. *J. Acoust. Soc. Am.*, 92:2907, 1992.

[Bej82] A. Bejan. *Entropy generation through heat and fluid flow.* Wiley, New York, 1982.

[Cep79] P. H. Ceperley. A pistonless stirling engine–the traveling wave heat engine. *J. Acoust. Soc. Am.*, 66:1508, 1979.

[Cep85] P. H. Ceperley. Gain and efficiency of a short traveling wave heat engine. *J. Acoust. Soc. Am.*, 77:1239, 1985.

[CWS] R. L. Carter, M. White, and A. M. Steele. Private communication of Atomics International Division of North American Aviation Inc., 1962. See also K. T. Feldman, "A study of heat generated pressure oscillations in a closed end pipe", Ph. D. dissertation, Mechanical Engineering Department, University of Missouri, 1966.

[CY92] J. A. Corey and G. A. Yarr. HOTS to WATTS: The LPSE linear alternator system re-invented. In *Proceedings of the 27th Intersociety Energy Conversion Engineering Conference*, page 5.289, 1992.

[GPG94] S. L. Garrett, D. K. Perkins, and A. Gopinath. Thermoacoustic refrigerator heat exchangers: Design, analysis, and fabrication. In G. F. Hewitt, editor, *Heat Transfer 1994: Proceedings of the 10th International Heat Transfer Conference. Vol. 4*, page 375, 1994. Brighton, UK.

[Hof88] T. J. Hofler. Concepts for thermoacoustic refrigeration and a practical device. In *Proceedings of the 5th International Cryocoolers Conference*, page 93, August 1988. Monterey, California; T. J. Hofler, "Thermoacoustic refrigerator design and performance", Ph.D. thesis, Physics department, University of California, San Diego, 1986.

[MS88] A. Migliori and G. W. Swift. Liquid sodium thermoacoustic engine. *Appl. Phys. Lett.*, 53:355, 1988.

[MYS+96] M. P. Mitchell, R. Yaron, S. Shokralla, J. Yuan, P. E. Bradley, and R. Radebaugh. Etched foil regenerator. *Adv. Cryogenic Eng.*, 41, 1996. Proceedings of 1995 Cryogenic Engineering Conference.

[OS94] J. R. Olson and G. W. Swift. Similitude in thermoacoustics. *J. Acoust. Soc. Am.*, 95:1405, 1994.

[Rad90] R. Radebaugh. A review of pulse tube refrigeration. *Adv. Cryogenic Eng.*, 35:1191, 1990.

[Rot80] N. Rott. Thermoacoustics. *Adv. Appl. Mech.*, 20:135, 1980. and references therein.

[Swi88] G. W. Swift. Thermoacoustic engines. *J. Acoust. Soc. Am.*, 84:1145, 1988.

[UB84] I. Urieli and D. M. Berchowitz. *Stirling cycle engine analysis.* Adam Hilger, Bristol UK, 1984.

[WS94] W. C. Ward and G. W. Swift. Design environment for low amplitude thermoacoustic engines. *J. Acoust. Soc. Am.*, 95:3671, 1994. Fully tested software and users guide available from Energy Science and Technology Software Center, US Department of Energy, Oak Ridge, Tennessee. For a beta-test version, contact ww@lanl.gov (Bill Ward) via Internet.

Table 1. Complex function h and its spatial average f for several geometries. In the second column, the origin is at the center of symmetry of the geometry, except in the boundary-layer limit where it is at the surface. In cases with a sort of circular symmetry, $r = \sqrt{y^2 + z^2}$. The hydraulic radius r_h is the ratio of gas volume to gas-solid contact area.

Geometry	h	f
Parallel plates, spacing $2y_0$; $r_h = y_0$	$\frac{\cosh[(1+i)y/\delta]}{\cosh[(1+i)y_0/\delta]}$	$\frac{\tanh[(1+i)y_0/\delta]}{(1+i)y_0/\delta}$
Circular pores, radius R; $r_h = R/2$	$\frac{J_0[(i-1)r/\delta]}{J_0[(i-1)R/\delta]}$	$\frac{2J_1[(i-1)R/\delta]}{J_0[(i-1)R/\delta]\,(i-1)R/\delta}$
Pin array (triangular lattice), pin radius r_i, centers' separation $\sqrt{2\pi}\,r_o / \sqrt[4]{3} \simeq 1.905 r_o$; $r_h = (r_o^2 - r_i^2)/2r_i$	$\simeq \frac{Y_1(z_o)J_1z - J_1(z_o)Y_1(z)}{Y_1(z_o)J_0(z_i) - J_1(z_o)Y_0(z_i)}$	$\simeq -\frac{2z_i}{z_o^2 - z_i^2} \frac{Y_1(z_o)J_1(z_i) - J_1(z_o)Y_1(z_i)}{Y_1(z_o)J_0(z_i) - J_1(z_o)Y_0(z_i)}$, where $z = (i-1)\,r/\delta$
Rectangular pores, $2a \times 2b$; $r_h = 2ab/(a+b)$	$1 - \frac{16}{\pi^2} \sum_{m,n\ odd} \frac{\sin(m\pi y/2a)\sin(n\pi z/2b)}{mnY_{mn}}$	$1 - \frac{64}{\pi^4} \sum_{m,n\ odd} \frac{1}{m^2 n^2 Y_{mn}}$, where $Y_{mn} = 1 - i\frac{\pi^2\delta^2 b^2}{8a^2}\,(m^2b^2 + n^2a^2)$
Boundary-layer limit	$e^{-(1+i)y/\delta}$	$\frac{(1-i)\delta}{2r_h}$

Table 2. Expressions used for numerical integrations in thermoacoustics. Compare to Eqs. (8) and (15), which are appropriate when $dT_m/dx = 0$.

$$\frac{dp_1}{dx} = -\frac{i\omega\rho_m}{(1-f_\nu)}\langle u_1\rangle$$

$$\frac{d\langle u_1\rangle}{dx} = -\frac{i\omega}{\rho_m a^2}[1+(\gamma-1)f_\kappa]\,p_1 + \frac{\beta(f_\kappa-f_\nu)}{(1-f_\nu)(1-\sigma)}\frac{dT_m}{dx}\langle u_1\rangle\ ,$$

$$\frac{dT_m}{dx} = \left\{\frac{\dot{H}_2}{A_{\rm gas}} - \frac{1}{2}\mathrm{Re}\left[p_1\langle\widetilde{u_1}\rangle\left(1-\frac{T_m\beta\left(f_\kappa-\widetilde{f}_\nu\right)}{(1+\sigma)\left(1-\widetilde{f}_\nu\right)}\right)\right]\right\} \div$$
$$\left\{\frac{\rho_m c_p|\langle u_1\rangle|^2}{2\omega(1-\sigma^2)|1-f_\nu|^2}\mathrm{Im}\left(f_\kappa+\sigma\widetilde{f}_\nu\right) - \kappa - \frac{A_{\rm solid}}{A_{\rm gas}}\kappa_{\rm solid}\right\}.$$

Table 3. Dimensionless groups relevant to thermoacoustics. The reference location, indicated by the subscript ref, is usually taken to be the room-temperature end of the stack. Dimensions: Of the many dimensions x_j of the apparatus, three are singled out for special attention: the overall length L, the stack area A_{ref} at the reference location, and its hydraulic radius r_h at the reference location. Gas properties: b_μ and b_κ account for the temperature dependences of μ and κ.

Dimensions	$A_{ref}/L, \quad r_h/L, \quad x_j/L$
Gas properties	$\gamma, \quad \sigma_{ref}, \quad b_\mu, \quad b_\kappa$
Thermophysical lengths	$fL/a_{ref}, \quad \delta_\kappa/r_h$
Powers	$\text{Power}/p_m a A$
Spatially dependent variables	$T(\mathbf{x},t)/T_{ref}, \quad p(\mathbf{x},t)/p_m, \quad \mathbf{v}(\mathbf{x},t)/a_{ref}$

Figure captions:

Fig. 1. Examples of thermoacoustic engines and refrigerators. (a) Sondhauss tube, the earliest thermoacoustic engine. (b) A heat-driven refrigerator, comprising a thermoacoustic engine and a thermoacoustic refrigerator sharing the same standing wave. (c) Orifice pulse-tube refrigerator, a traveling-wave system.

Fig. 2. Spatial-average function f for several geometries. The rectangle here has 6:1 aspect ratio, and the pin array has $r_o/r_i = 6$. The boundary-layer limit is approached at large r_h for all geometries.

Fig. 3. (a) A typical gas parcel executing the four steps of the cycle of a standing-wave thermoacoustic refrigerator, assuming an inviscid gas and square-wave acoustic motion and pressure. In each step, the dashed and solid ellipses show respectively the initial and final positions and volumes of the parcel, while labels show the initial and final temperatures and pressures of that step; and broad arrows show directions of attendant heat flows (between parcel and solid) and work flows (into or out of the parcel). The solid temperature at the average location of the gas parcel is T_m. Averaged over one cycle, the parcel absorbs work $dW - dW'$ from the sound wave. In Figs. 3 and 4, we have dropped our usual complex notation, so that all variables here are positive real numbers. (b) A typical gas parcel executing the four steps of the cycle of a standing-wave thermoacoustic engine, assuming an inviscid gas and square-wave acoustic motion and pressure. Averaged over one cycle, the parcel delivers work $dW - dW'$ to the sound wave.

Fig. 4. A typical gas parcel executing the four steps of the cycle of a traveling-wave thermoacoustic refrigerator, assuming an inviscid gas and square-wave acoustic motion and pressure. Symbols have the same meaning as in Fig. 3.

Fig. 5. (a) An early thermoacoustic refrigerator, and (b) some of its measured and calculated performance, with 500 Hz pressure oscillations in 10 bar helium gas, and with $T_H = 300$ K. Data from Hofler, 1988: circles, $|p_1|/p_m = 0.015$; squares,

$|p_1|/p_m = 0.03$. Lines calculated using Ward and Swift, 1994. The gross cooling power Q_{TOT} includes the deliberately applied load plus some small parasitic loads such as heat leak from room temperature (see Hofler 1988 for details). $COP = Q_{TOT}/W$, with W the acoustic power delivered to the resonator.

Fig. 6. A recent example of thermoacoustic measurements and calculations. (a) Schematic of apparatus, a standing-wave engine driving no load. Small circles indicate locations of sensors. (b) Heater power Q_H and hot temperature T_H as functions of $|p_1|$, all normalized. Points are measurements; lines are calculations. Triangles and solid lines: helium, 0.96 MPa. Circles and dotted lines: neon, 0.70 MPa. Squares and dashed lines: argon, 0.36 MPa.

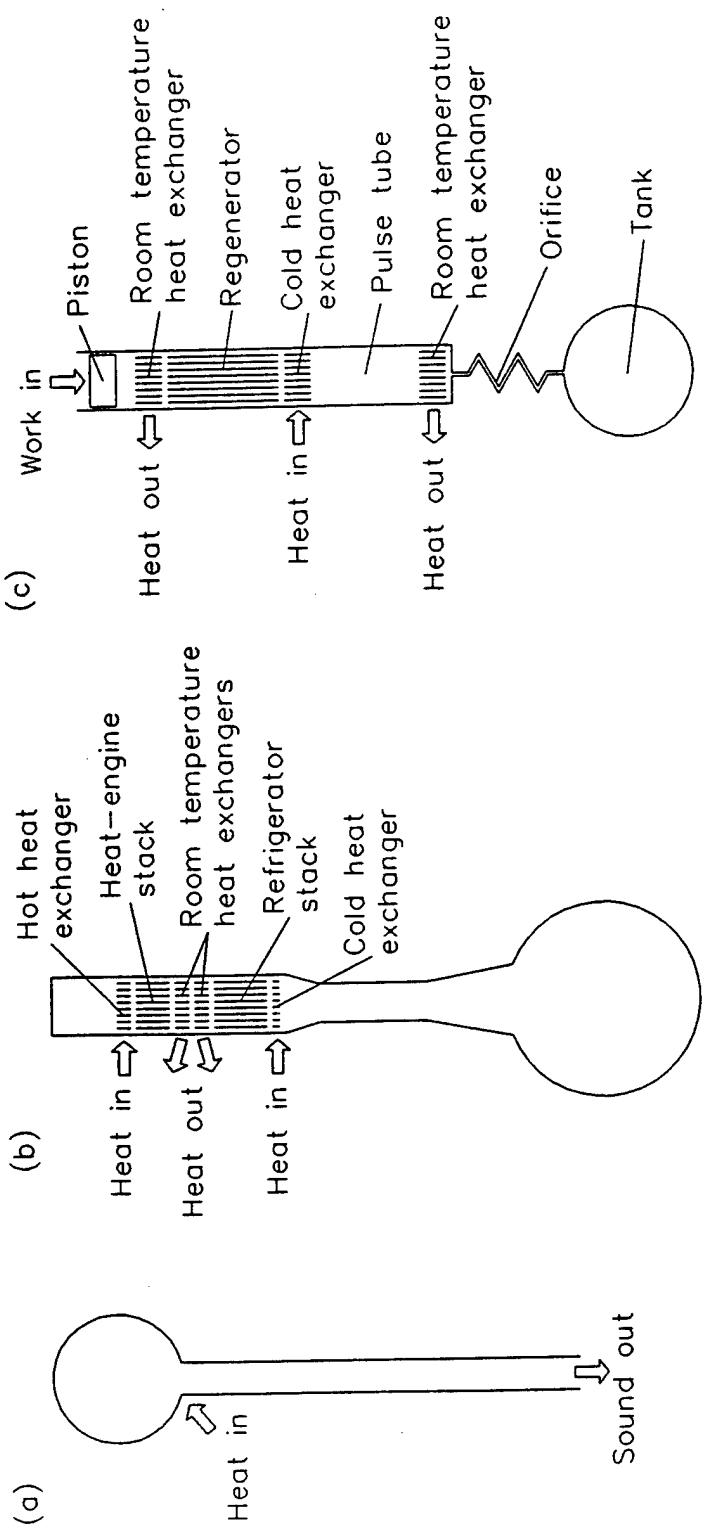

(a)
Heat in
Sound out
(b)
Hot heat exchanger
Heat-engine stack
Room temperature heat exchangers
Refrigerator stack
Cold heat exchanger
Heat in
Heat out
Heat in
(c)
Work in
Piston
Room temperature heat exchanger
Regenerator
Cold heat exchanger
Pulse tube
Room temperature heat exchanger
Orifice
Tank
Heat out
Heat in
Heat out

Fig. 1

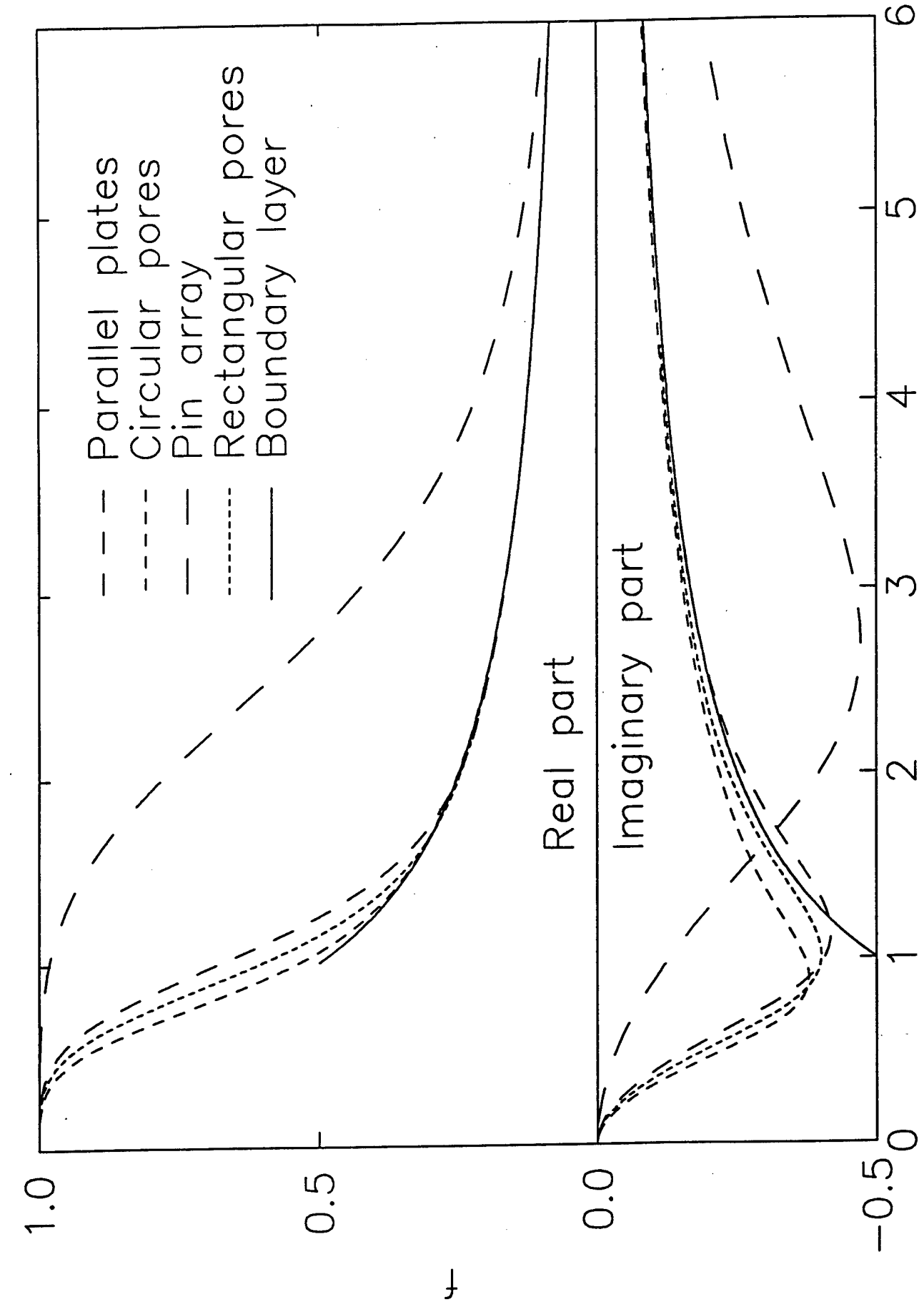
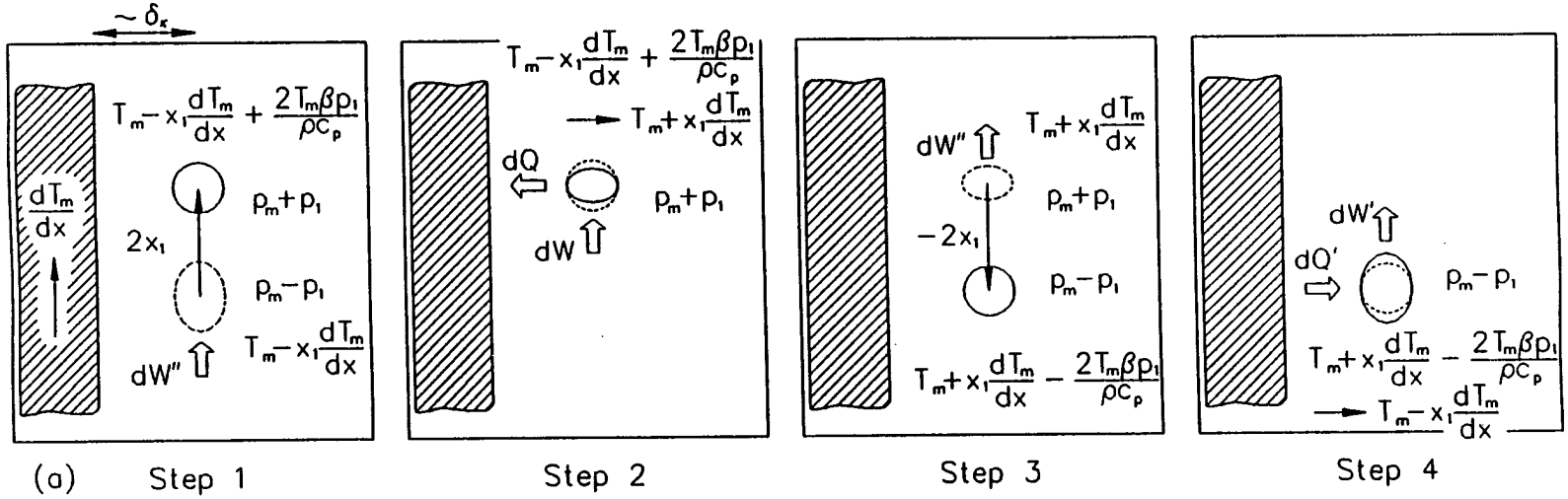

Parallel plates
Circular pores
Pin array
Rectangular pores
Boundary layer
Real part
Imaginary part
f
1.0
0.5
0.0
−0.5
0
1
2
3
4
5
6
r_h/δ

Fig 2


~ δ_x
$T_m - x_1\frac{dT_m}{dx} + \frac{2T_m\beta p_1}{\rho c_p}$
$\frac{dT_m}{dx}$
$2x_1$
$p_m + p_1$
$p_m - p_1$
dW''
$T_m - x_1\frac{dT_m}{dx}$
dQ
dW
$T_m + x_1\frac{dT_m}{dx}$
$-2x_1$
$T_m + x_1\frac{dT_m}{dx} - \frac{2T_m\beta p_1}{\rho c_p}$
dW'
dQ'
(a) Step 1
Step 2
Step 3
Step 4

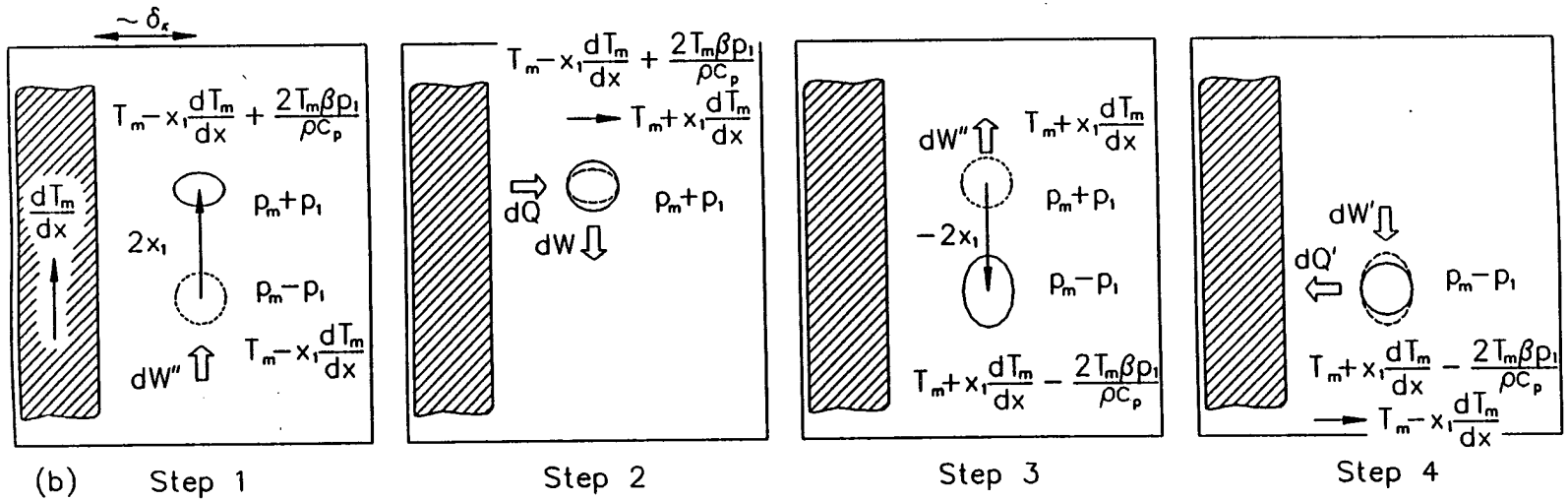

~ δ_x
$T_m - x_1\frac{dT_m}{dx} + \frac{2T_m\beta p_1}{\rho c_p}$
$\frac{dT_m}{dx}$
$2x_1$
$p_m + p_1$
$p_m - p_1$
dW''
$T_m - x_1\frac{dT_m}{dx}$
$T_m + x_1\frac{dT_m}{dx}$
dQ
dW
$-2x_1$
$T_m + x_1\frac{dT_m}{dx} - \frac{2T_m\beta p_1}{\rho c_p}$
dW'
dQ'
(b) Step 1
Step 2
Step 3
Step 4

Fig. 3

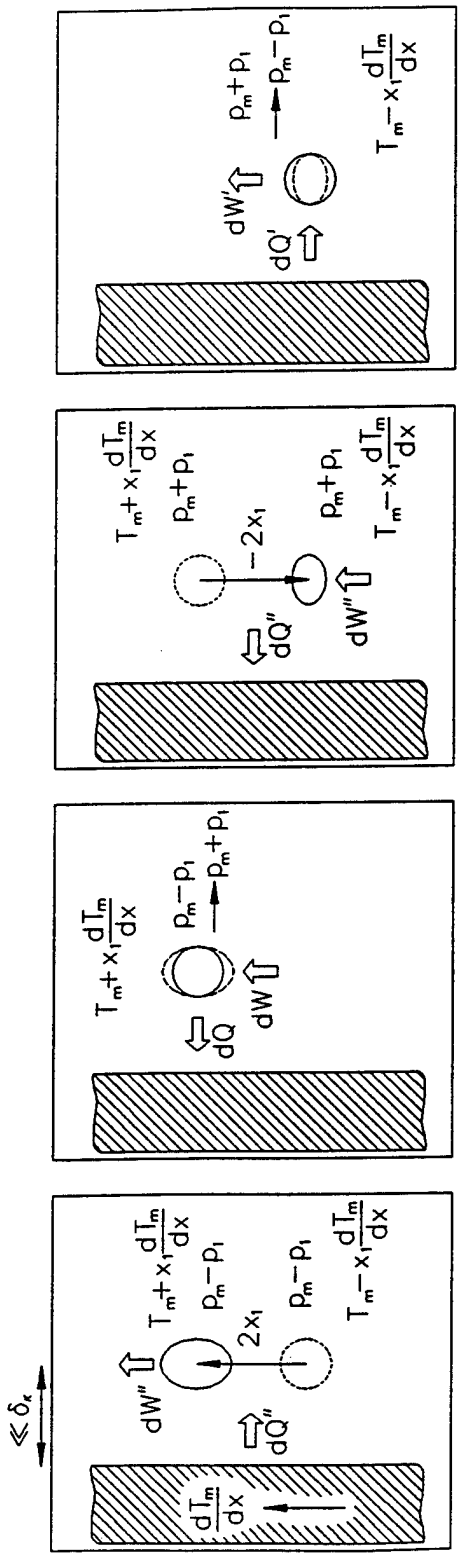
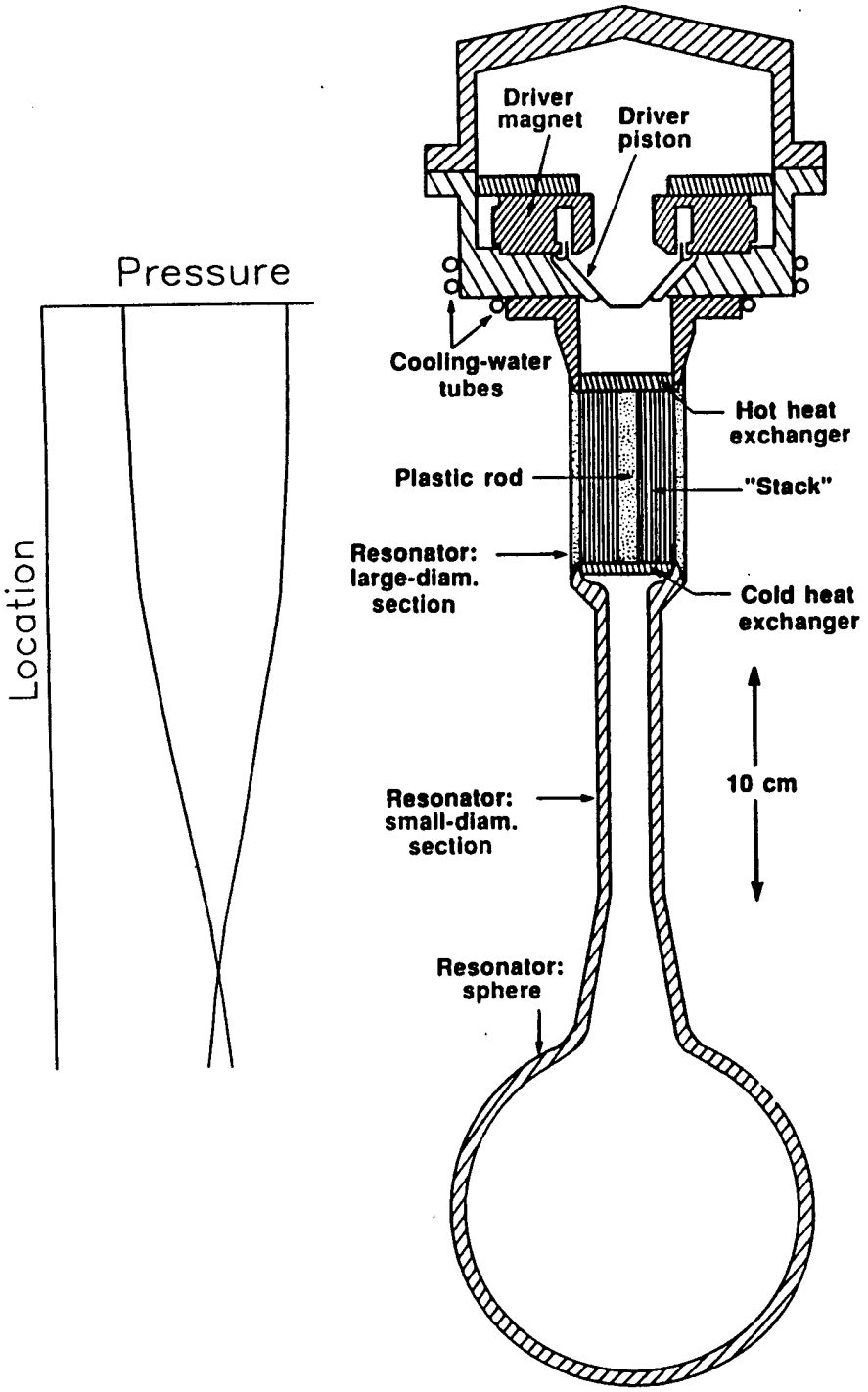
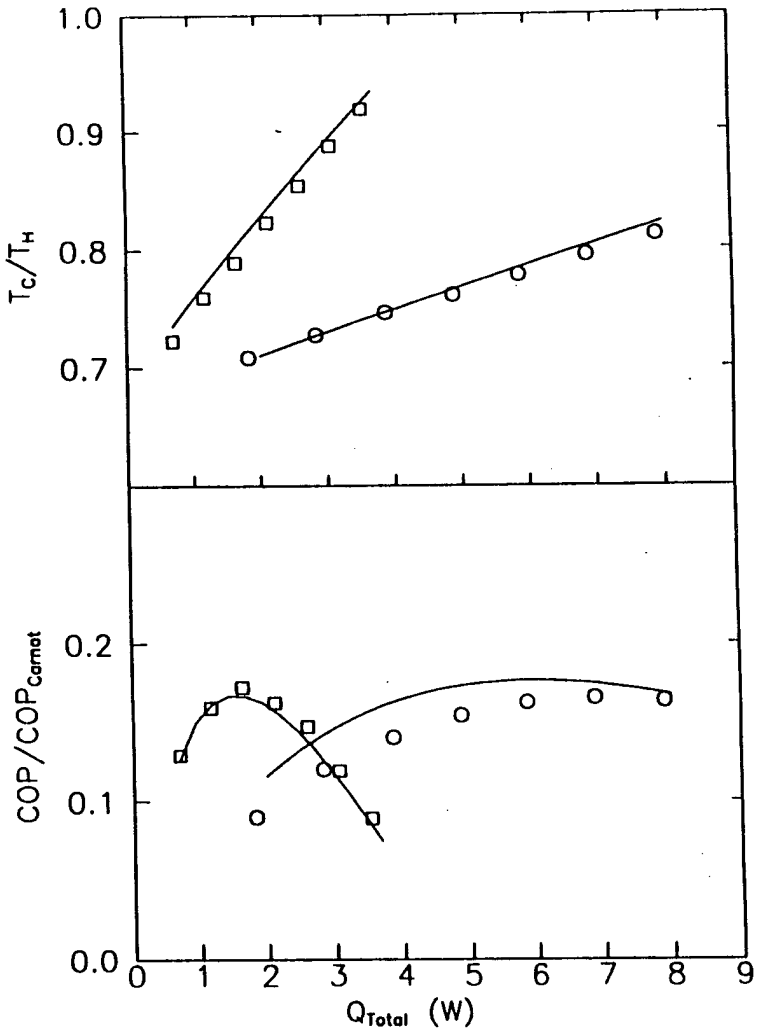
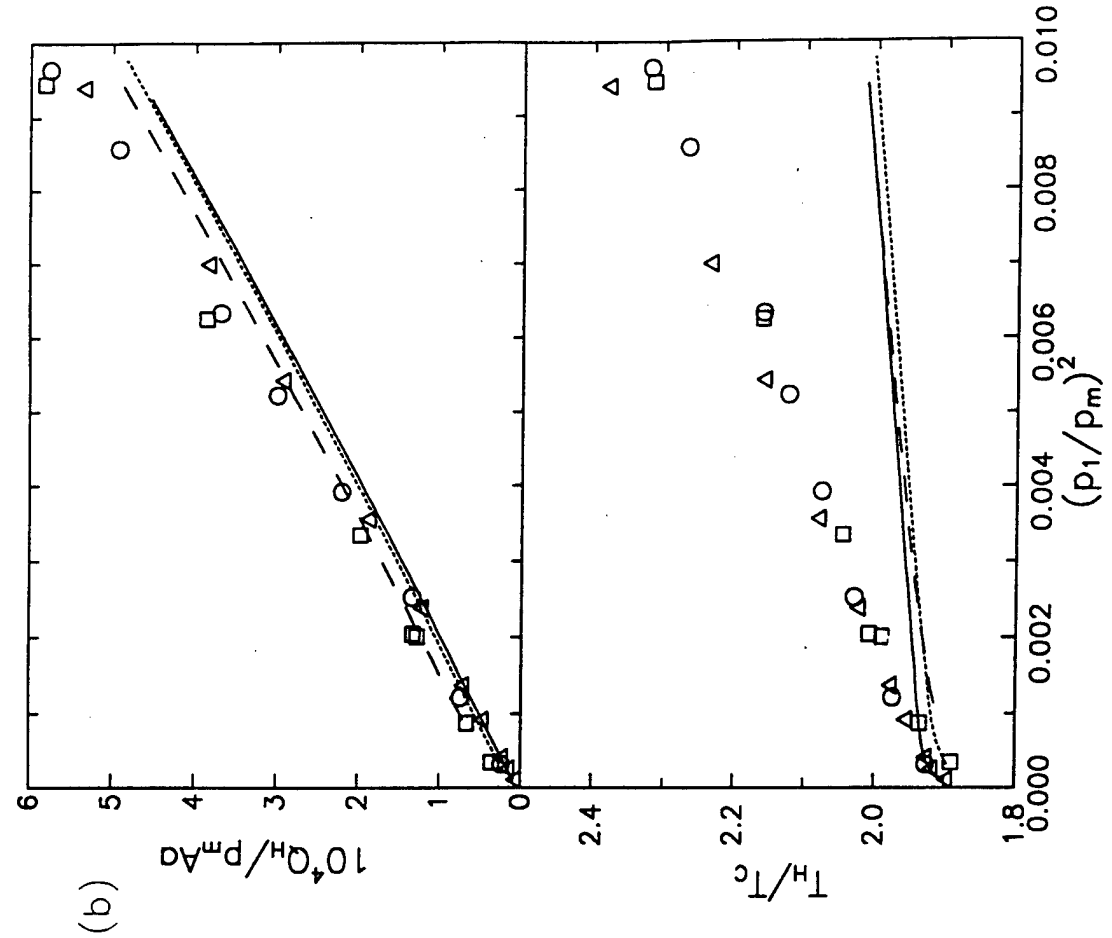
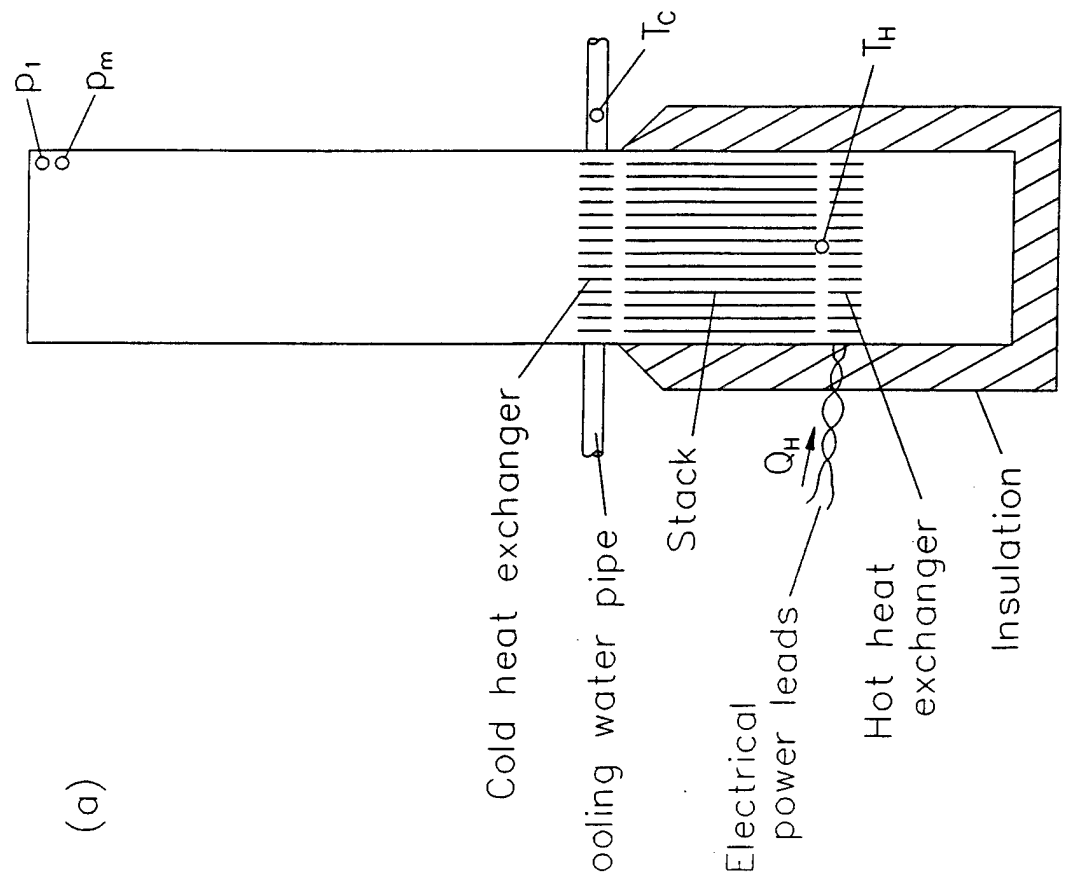

(b)
Step 1
Step 2
Step 3
Step 4
$\ll \delta_x$
$\frac{dT_m}{dx}$
dW''
dQ''
$T_m + x_1\frac{dT_m}{dx}$
$T_m - x_1\frac{dT_m}{dx}$
$p_m - p_1$
$p_m + p_1$
$2x_1$
dQ
dW
dW'
dQ'

Fig. 4


Pressure
Location
Driver magnet
Driver piston
Cooling-water tubes
Hot heat exchanger
Plastic rod
"Stack"
Resonator: large-diam. section
Cold heat exchanger
10 cm
Resonator: small-diam. section
Resonator: sphere

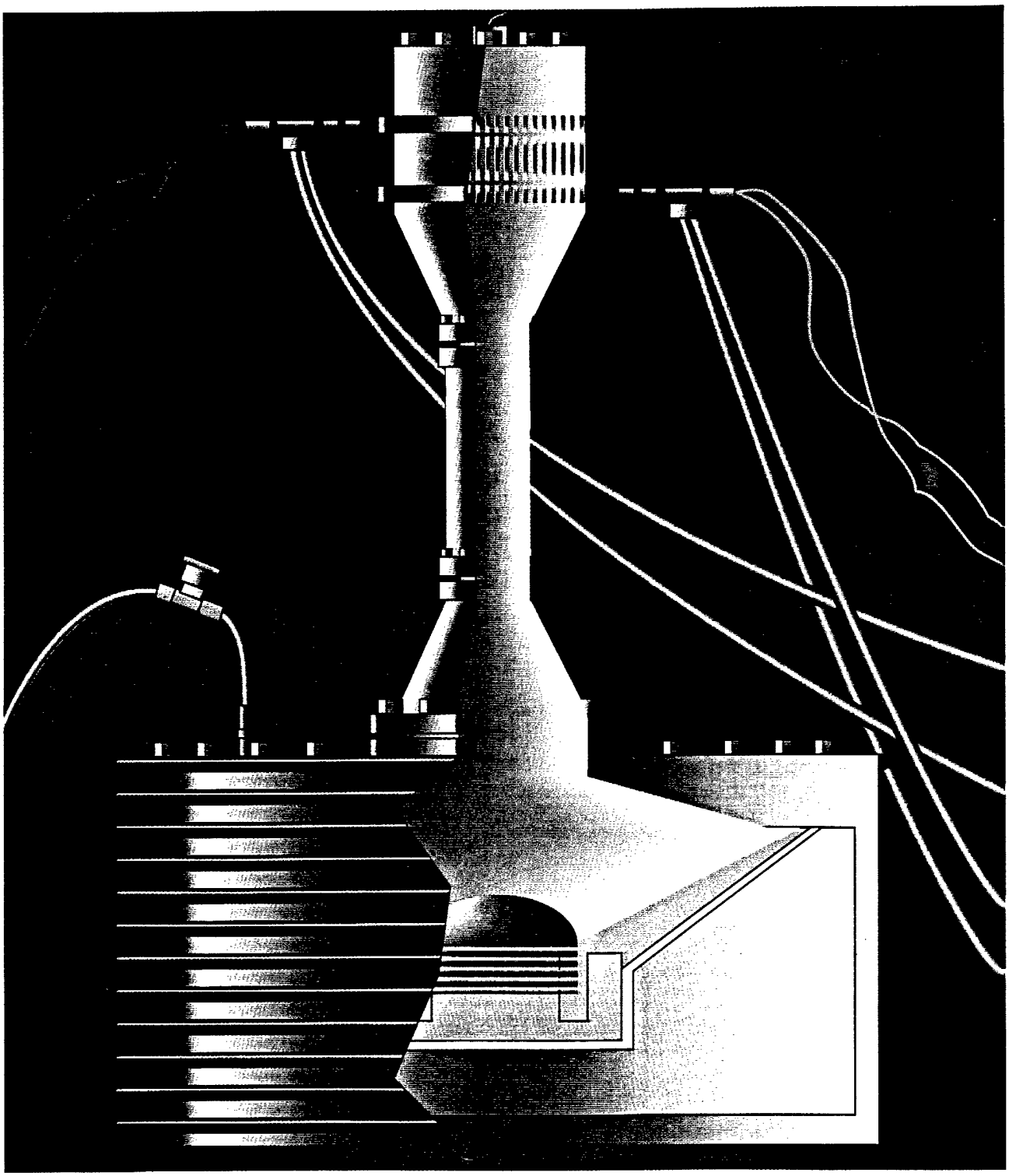

1.0
0.9
0.8
0.7
T_C/T_H
0.2
0.1
0.0
COP/COP_{Carnot}
0 1 2 3 4 5 6 7 8 9
Q_{Total} (W)


(b)
$10^4 Q_H/\rho_m A_D$
T_H/T_c
$(p_1/p_m)^2$
0.000
0.002
0.004
0.006
0.008
0.010
1.8
2.0
2.2
2.4
0
1
2
3
4
5
6


(a)
p_1
p_m
T_c
T_H
Q_H
Cold heat exchanger
Cooling water pipe
Stack
Electrical power leads
Hot heat exchanger
Insulation

PHYSICS TODAY

JULY 1995



THERMOACOUSTICS

THERMOACOUSTIC ENGINES AND REFRIGERATORS

On the heels of basic research, commercial developers are harnessing acoustic processes in gases to make reliable, inexpensive engines and cooling devices with no moving parts and a significant fraction of Carnot's efficiency.

Gregory W. Swift

We ordinarily think of a sound wave in a gas as consisting of coupled pressure and displacement oscillations. However, temperature oscillations always accompany the pressure changes. The combination of all these oscillations, and their interaction with solid boundaries, produces a rich variety of "thermoacoustic" effects. Although these effects as they occur in everyday life are too small to be noticed, one can harness extremely loud sound waves in acoustically sealed chambers to produce powerful heat engines, heat pumps and refrigerators. Whereas typical engines and refrigerators have crankshaft-coupled pistons or rotating turbines, thermoacoustic engines and refrigerators have at most a single flexing moving part (as in a loudspeaker) with no sliding seals. Thermoacoustic devices may be of practical use where simplicity, reliability or low cost is more important than the highest efficiency (although one cannot say much more about their cost-competitiveness at this early stage).

The basics: Thermoacoustic engines

A thermoacoustic engine converts some heat from a high-temperature heat source into acoustic power, rejecting waste heat to a low-temperature heat sink. The heat-driven electrical generator shown in figure 1 illustrates the basic principle of operation. The overall view, shown at the top of figure 1a, is reminiscent of the appearance of a heat engine in many introductory thermodynamics texts: The apparatus absorbs heat per unit time Q_h from a heat source at high temperature T_h, rejects heat per unit time Q_c to a heat sink at low temperature T_c and produces acoustic power W. The first law of thermodynamics tells us that $W + Q_c = Q_h$; the second law shows that the efficiency W/Q_h is bounded above by the Carnot efficiency $(T_h - T_c)/T_h$. (I will use Q and W for heat power and acoustic power, and q and w for the corresponding energies.)

One of the most important scales in a thermoacoustic device is the length of its resonator, which (together with the gas sound speed) determines the operating frequency, just as the length of an organ pipe determines its pitch. This length typically falls between 10 cm and 10 m. In figure 1a, with both ends of the resonator closed, the lowest resonant mode is that which fits a half-wavelength standing wave in the resonator, with displacement nodes and pressure antinodes at the ends, as shown in the lower portion of figure 1. The heat-exchange elements—a hot heat exchanger, a cold heat exchanger and a "stack" between them—are located where both oscillating pressure and oscillating gas displacement are nonzero. Each of the two heat exchangers is typically a set of copper fins, as open to gas flow as possible (like a car radiator), and is thermally anchored to its reservoir by metallic heat conduction or circulating fluids. The stack is simply a well-spaced stack of solid plates with high heat capacity, also quite open to gas flow, supporting the smooth temperature profile between the two heat exchangers as shown at the bottom of figure 1a. Most of the "parallel plate" stacks constructed so far have in fact been spirally wound, a low-cost configuration.

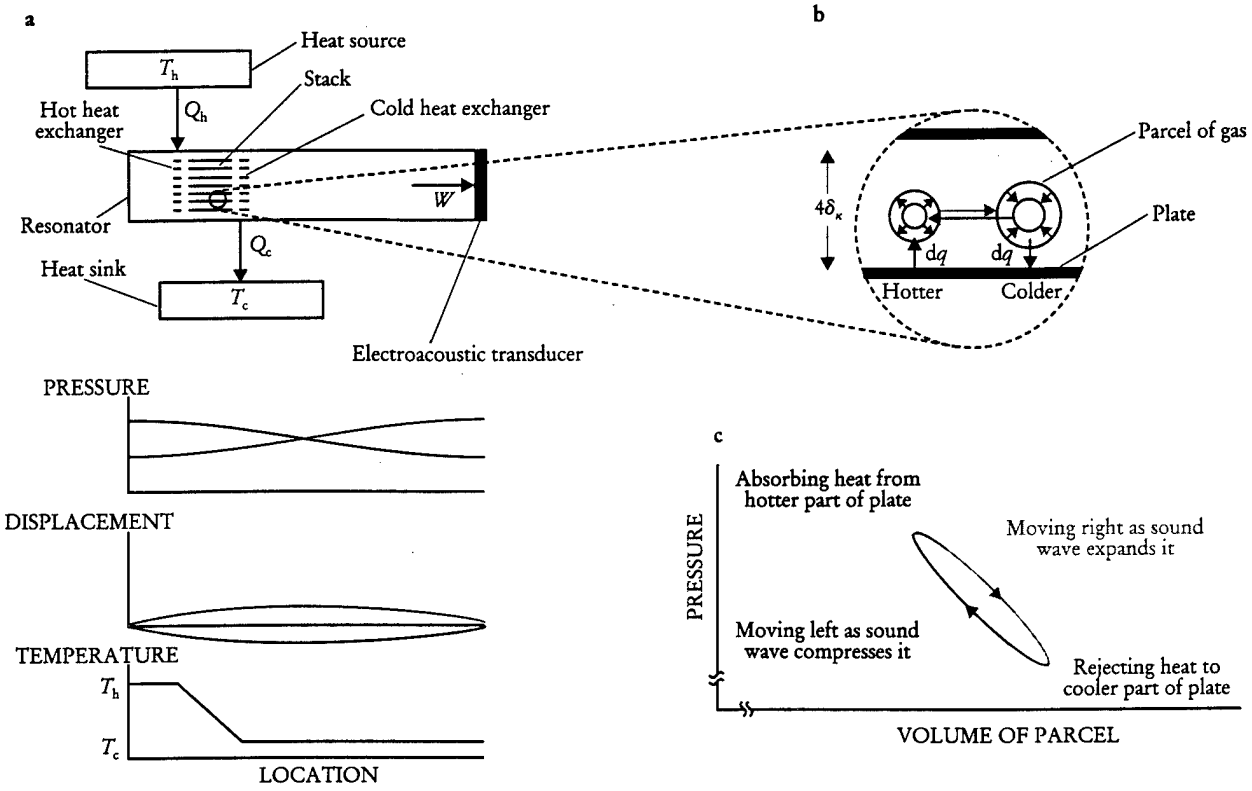
To understand in some detail the conversion of heat to acoustic power by this simple structure, consider the magnified view of part of the stack in figure 1b, which shows a typical parcel of gas at four instants of time during one cycle of the acoustic wave. The standing wave carries the parcel left and right, compressing and expanding it, with phasing such that it is at its most compressed state when at its farthest left position, and its most expanded state at its farthest right position. In typical thermoacoustic engines and refrigerators the amplitude of the pressure oscillation is 3–10% of the mean pressure, and the displacement amplitude is a similar percentage of the length of a plate in the stack.

The presence of an externally imposed temperature gradient in the stack adds a new feature to what would otherwise be a simple acoustic oscillation: oscillatory heat transfer between the parcel of gas and the stack. (To simplify this discussion, I will neglect the adiabatic temperature oscillations that accompany the pressure oscillations.) When the parcel is at its leftmost position, heat flows from the relatively hot stack plate to the parcel, expanding it; when the parcel is at its rightmost position, heat flows from it to the relatively cool stack plate, contracting the parcel. The parcel does net work on its surroundings, because the expansion takes place at the high-pressure phase of the cycle and the contraction at the low-pressure phase, as shown in figure 1c.

(Readers with Internet access are encouraged to view our computer animations of this process and of thermoacoustic refrigeration as described below. The thermoacoustics home page is at http://rott.esa.lanl.gov/; select "educational

GREGORY SWIFT *works in the condensed matter and thermal physics group at Los Alamos National Laboratory, in New Mexico.*

© 1995 American Institute of Physics



SIMPLE THERMOACOUSTIC ENGINE. **a:** Heat exchangers and a stack in a half-wavelength acoustic resonator convert some of the heat power Q_h from a thermal reservoir at temperature T_h into acoustic power W, rejecting waste heat power Q_c to another reservoir at T_c. The acoustic power is delivered to an electroacoustic transducer, which converts it to electricity. Plots below show gas pressure, gas displacement in the horizontal direction and average temperature as functions of location in the resonator. Pressure and displacement are each shown when the gas is at the leftmost extreme of its displacement (red), with density and pressure highest at the left end of the resonator and lowest at the right end, and 180° later in the cycle (blue). **b:** Magnified view of part of the stack shows a typical parcel of gas (greatly exaggerated in size) as it oscillates in position, pressure and temperature, exchanging heat dq with the nearby plates of the stack. Plates are separated by about four thermal penetration depths δ_κ. **c:** Pressure-volume (p–V) diagram for the parcel of gas shows how it does net work $\delta w = \oint p\, dV$ on its surroundings. **FIGURE 1**

demonstrations." For DOS-based computers, the executable file FANCY.EXE and text file FANCY.TXT can be downloaded.)

The net work that the parcel does on its surroundings is delivered in each cycle of the acoustic oscillation. The parcel and all others like it within the stack deliver acoustic power W to the standing wave; the standing wave delivers it in turn to the electroacoustic transducer. Each parcel absorbs a little heat from one location in the stack and deposits a little heat farther to the right, at a cooler location in the stack. With respect to heat, all the parcels act like members of a bucket brigade, with the overall effect being absorption of Q_h at the hot heat exchanger and rejection of Q_c at the cold heat exchanger.

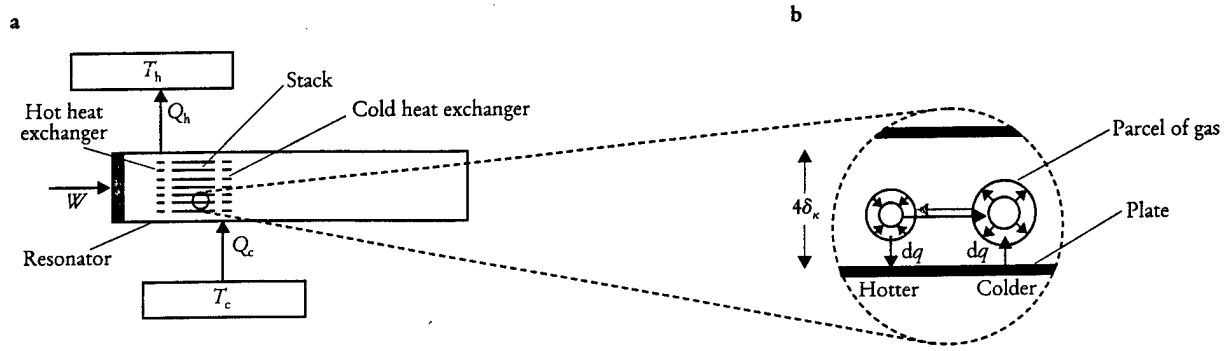
A second important scale in a thermoacoustic engine is the separation between plates in the stack, which determines the nature of thermal contact between the plate and the typical parcel of gas. Imperfect thermal contact is needed to accomplish the cycle shown in figure 1, because it is desirable to have good thermal contact when the parcel is stationary or nearly so, but poor thermal contact while it is moving. Detailed analysis shows that a spacing between plates of about four thermal penetration depths $\delta_\kappa = \sqrt{\kappa/\pi f \rho c_p}$ is best, where κ is the thermal conductivity of the gas, ρ is its density, c_p its isobaric specific heat per unit mass and f the frequency of the acoustic oscillation; δ_κ is roughly the distance heat can diffuse through the gas during a time $1/\pi f$. In today's thermoacoustic systems, δ_κ is typically a fraction of a millimeter.

Heat-driven acoustic oscillators such as that shown in figure 1 have been known for over a century. The earliest and simplest, known as the Sondhauss tube, was discovered accidentally by European glassblowers; a later example, the Taconis oscillator, is well known today to users of liquid helium. In each of these early thermoacoustic engines, the resonator itself was only several δ_κ in diameter, and its walls also served the functions of stack and heat exchangers. The use of an explicit stack (with multiple parallel passages) and heat exchangers in larger-diameter resonators to increase overall power began with Robert L. Carter in the 1960s.[1]

Although progress toward a theoretical understanding of these phenomena began with Lord Rayleigh 120 years ago, a quantitatively accurate theory of thermoacoustics was not developed until the 1970s, through the efforts of Nicholas Rott and coworkers.[2] This theory is based on a low-amplitude linearization of the Navier–Stokes, continuity and energy equations with sinusoidal oscillations of all variables. It is remarkable that such qualitatively simple classical phenomena went without a quantitatively accurate explanation until late in this century.

The basics: Thermoacoustic refrigerators

Another chapter in the development of thermoacoustics began at Los Alamos National Laboratory in the early



SIMPLE THERMOACOUSTIC REFRIGERATOR. **a:** Electroacoustic transducer at the left end delivers acoustic power W to the resonator, producing refrigeration Q_c at low temperature T_c and rejecting waste heat power Q_h to a heat sink at T_h. As in figure 1, this is a half-wavelength device with a pressure node at the midpoint of the resonator. The temperature gradient in the refrigerator's stack is much less steep than that in the stack for the engine shown in figure 1. **b:** Magnified view of part of the stack shows a typical parcel of gas as it moves heat dq up the temperature gradient. Here $\oint p\,dV < 0$, so the pressure–volume cycle analogous to figure 1c goes counterclockwise, and the parcel absorbs work from its surroundings. **FIGURE 2**

1980s with the invention of thermoacoustic refrigeration.[3] The basic principle of operation, illustrated in figure 2, is very similar to that of thermoacoustic engines, but here the temperature gradient in the stack is much lower. As the gas oscillates along the stack, it experiences changes in temperature. Much of the gas's temperature change comes from adiabatic compression and expansion of the gas by the acoustic pressure, and the rest is a consequence of heat transfer with the stack. At the leftmost position of the parcel of gas shown in figure 2b, it rejects heat to the stack, because its temperature was raised above the local stack temperature by adiabatic compression caused by the standing wave. Similarly, at its rightmost position, the parcel absorbs heat from the stack, because adiabatic expansion has brought its temperature below the local stack temperature. Thus the parcel of gas moves a little heat from right to left along the stack, up the temperature gradient, during each cycle of the acoustic wave.

All the other parcels in the stack behave similarly, so that the overall effect, again as in a bucket brigade, is the net transport of heat from the cold heat exchanger to the hot heat exchanger, with Q_c absorbed at T_c and Q_h rejected at T_h. The parcel absorbs acoustic work from the standing wave, because the thermal expansion of the parcel of gas occurs during the low-pressure phase of the acoustic wave and the thermal contraction during the high-pressure phase. The resulting acoustic power W absorbed by all the parcels in the stack can be supplied by a loudspeaker, a thermoacoustic engine or other means. The first law of thermodynamics once again determines that $W + Q_c = Q_h$; the second law shows that the relevant efficiency, known as the coefficient of performance, is bounded above by the Carnot coefficient $T_c/(T_h - T_c)$.

The steepness of the temperature gradient in the stack determines whether a thermoacoustic device is a refrigerator (which has work done on it) or an engine (which does work). In an engine, with a steep temperature gradient as shown in figure 1, the gas parcel finds itself cooler than the local stack temperature after its adiabatic compression during displacement to the left, so it absorbs heat from the stack at high pressure and expands. In contrast, in a refrigerator, with a shallow gradient, the gas parcel finds itself warmer than the local stack temperature after its adiabatic compression during displacement to the left, so it rejects heat to the higher-temperature part of the stack and contracts.

Figure 3a shows schematically the first efficient thermoacoustic refrigerator,[4] designed, built and studied by Tom Hofler. It illustrates several features of many of today's thermoacoustic devices. The resonator had a slightly complicated geometry, which maintained the desired frequency, pressure amplitude and displacement amplitude at the stack while reducing the total length to much less than half the wavelength. This geometry also reduced viscous and thermal losses on the resonator walls and suppressed the harmonic content so that the sound wave remained purely sinusoidal in time. The pressure antinode is at the driver piston, and the pressure node is at the widening neck near the sphere, so this is essentially a quarter-wavelength apparatus, even though the spatial dependence of the pressure is not exactly a cosine. High-pressure helium gas was used: High pressure increases the power per unit volume of apparatus, and helium, having the highest sound speed and thermal conductivity of the inert gases, further increases the power density and allows spacings within the stack and heat exchangers to be as large as possible, for ease of fabrication. The loudspeaker-like driver was located at a pressure antinode of the standing wave, so that the acoustic power was delivered with high force and small displacement, easing engineering difficulties associated with the flexing portion of the driver. This location also placed it next to the hot heat exchanger, where heat generated in the driver could be removed most efficiently.

Figure 3b shows some of the data obtained with this refrigerator, which reached a T_c of –70 °C and had a cooling power of several watts with acoustic pressure amplitudes of 3% of the mean pressure. The curves in the figure were calculated using publicly available software[5] based on the theory developed by Rott and are in reasonable agreement with the data. The calculations have no adjustable parameters; they simply use the geometry of the apparatus and the properties of helium gas.

Commercial developments

Attempts to develop practical devices based on thermoacoustics began just a few years ago, throughout the US and on four other continents. This surge of interest was due to the interaction of several factors: the new "tech transfer" emphasis at government laboratories; the engineering development of some thermoacoustic refrigerators at the Naval Postgraduate School in Monterey, California,

under the enthusiastic leadership of Steve Garrett; the crisis in the refrigeration industry caused by the destruction of stratospheric ozone by chlorofluorocarbons; and the marriage of thermoacoustic engines with orifice pulse-tube refrigerators (discussed below). To illustrate the breadth of applications under way, I have chosen four examples from among the corporate-sponsored thermoacoustics projects that I know of.

The thermoacoustic refrigerator shown in figure 4a is a prototype for a food refrigerator. Built at CSIR (formerly called the Council for Scientific and Industrial Research) in the Republic of South Africa with corporate support, it is a direct descendant of a thermoacoustic refrigerator[6] that was originally intended for preserving blood and urine samples on the space shuttle. It is a symmetrical, essentially half-wavelength device driven by modified loudspeakers on both ends, with two stacks, each with two heat exchangers. The pressure node is at the center of the bottom section. Use of two stacks maximizes cooling power for a given resonator size, all other things being equal. For compactness, the fiberglass resonator is formed in a "U" shape, with little effect on the acoustics. Cooling power (typically 100 W) and temperatures are appropriate for residential food refrigeration in most of the world. (American refrigerators are larger than most and hence require about twice as much cooling power.)

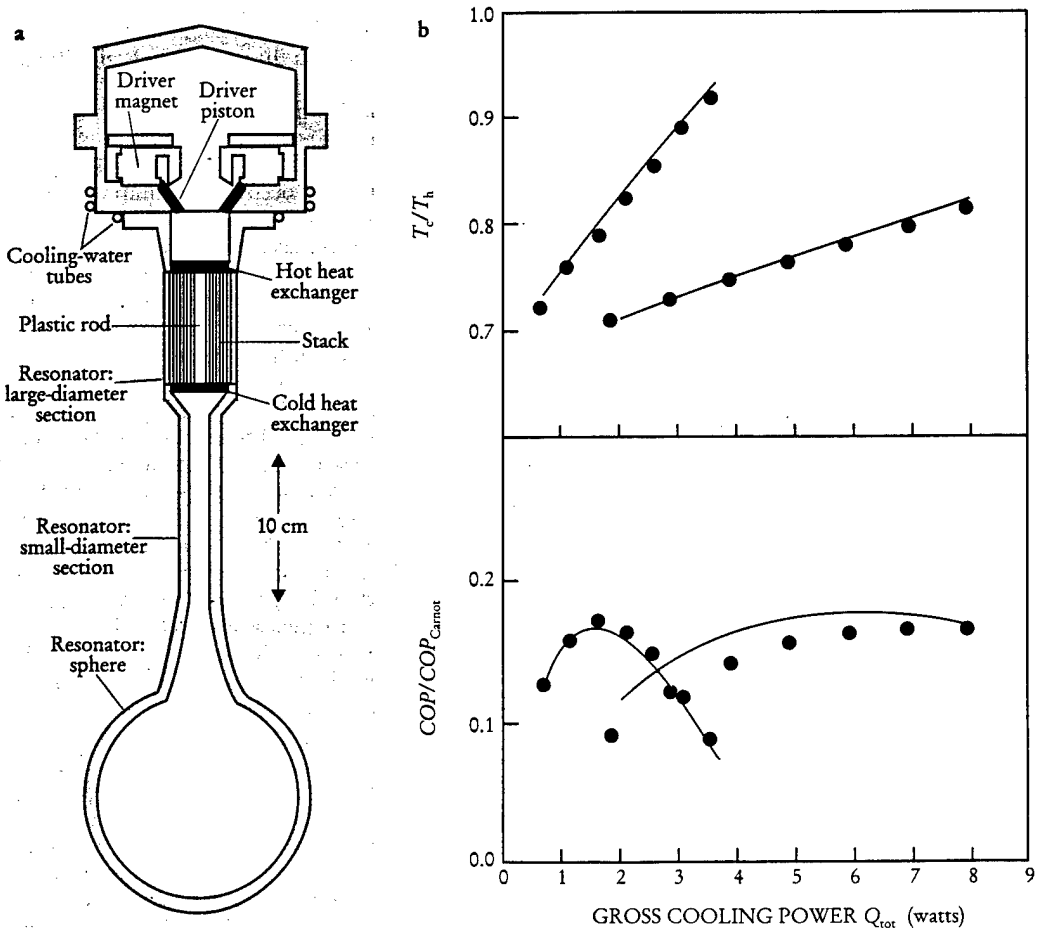
Scientists at Ford Motor Company built the thermoacoustic refrigerator depicted on the cover of this issue. The driver is at the displacement maximum of the quarter-wavelength standing wave (instead of at the pressure maximum as in the refrigerators of figures 3a and 4a). Thus in this refrigerator acoustic power is delivered with small pressure and large volumetric displacement, accomplished by using a large area in the driver. The driver's losses flow to the cold heat exchanger, but this is a minor problem if the driver is efficient and T_c is not too far below T_h. Water inlet and outlet tubes (gray in the diagram), essentially serving as the hot and cold thermal reservoirs, are clearly visible at the heat exchangers. This device operates at 10 bars with either helium driven at 430 Hz or a mixture of 80% helium—20% argon driven at 260 Hz.

At Tektronix Corporation in Beaverton, Oregon, researchers are developing a system for cooling electronics to cryogenic temperatures. A thermoacoustic engine drives an orifice pulse-tube refrigerator,[7] which is related to both Stirling and thermoacoustic refrigerators. The system thus constitutes a heat-driven cryogenic refrigerator having no moving parts; it has cooled to 150 K. It has a half-wavelength resonator with resistively heated thermoacoustic engines near both ends. In a sidebranch, 500 W of acoustic power from these engines is delivered to the pulse-tube refrigerator.

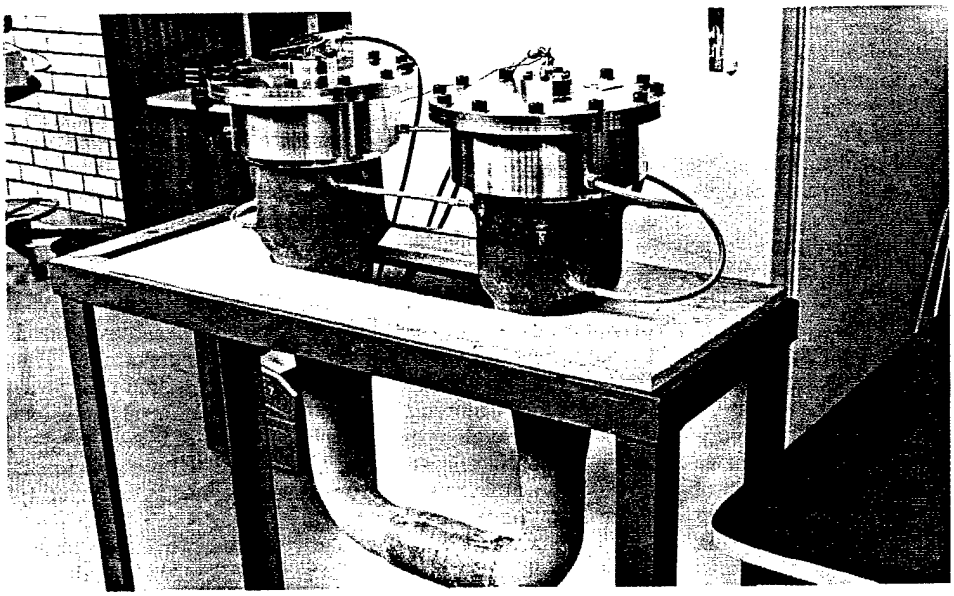
The largest thermoacoustic engine to date, under construction at Cryenco Inc in Denver, Colorado, will also drive a pulse-tube refrigerator. The heat source for the engine will be natural gas combustion. Intended for industrial and commercial liquefaction of natural gas it will measure 12 meters long, and will use two ½-meter-diameter spiral stacks (figure 4b) to produce 40 kW of acoustic power at 40 Hz in 30-bar helium gas. The device should be completed this year.

Power and efficiency

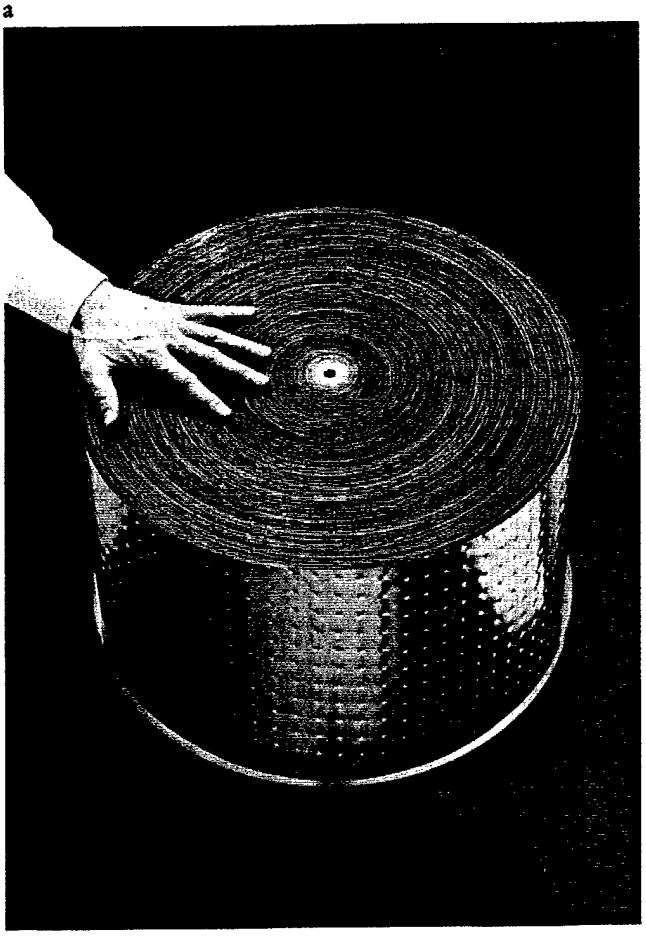
The power of thermoacoustic devices is roughly proportional to $p_{avg}Aa(p_{osc}/p_{avg})^2$, where p_{avg} is the average pressure, A the cross-sectional area of the stack, a the sound



FIRST EFFICIENT THERMOACOUSTIC REFRIGERATOR (a) and some of its performance parameters (b) as measured[4] (data points) and calculated[5] (curves) for operation with 500-Hz pressure oscillations in 10-bar helium gas, and with T_h = 300 K. Blue circles are data for 1.5% pressure oscillations; red circles, 3%. The gross cooling power Q_{tot} includes the deliberately applied load plus some small parasitic loads such as heat leak from room temperature. The coefficient of performance (*COP*) equals Q_{tot}/W, with W the acoustic power delivered to the resonator. **FIGURE 3**



a



b

TWO COMMERCIALLY INTERESTING thermoacoustic systems. **a:** Half-wavelength refrigerator with two stacks driven by two loudspeakers was built at CSIR in South Africa. It operates at 120 Hz with 15-bar neon. The heat exchangers are located where the water lines connect to the green resonator body. (Courtesy of Peter Bland, Quadrant.) **b:** One of the two spiral stacks for the largest thermoacoustic engine to date, being built by Cryenco Inc. (Courtesy of John Wollan, Cryenco.) FIGURE 4

speed of the gas and p_{osc} the amplitude of the oscillatory pressure. Helium (with high sound speed) is often used, typically at a pressure above 10 bars. In the examples cited in the previous section, p_{osc}/p_{avg} values range from 0.03 to 0.10, chosen as design compromises between the high power density achieved at high amplitude and the high confidence in the quantitative accuracy of Rott-based calculations at low amplitude.

The efficiency of thermoacoustic devices falls below Carnot's efficiency because of five major sources of irreversibility—"inherent," viscous, conduction, auxiliary and transduction losses:

▷ Inherent loss arises from the heat transfer to and from each parcel of gas in the stack as shown in figures 1 and 2: Whenever heat dq is transferred across a nonzero temperature difference δT, the entropy of the universe increases[8] by $dq\ \delta T/T^2$. This irreversibility is unavoidable in the thermoacoustic process, relying as it does on imperfect thermal contact for the correct phasing between pressure oscillations and thermal expansion–contraction.

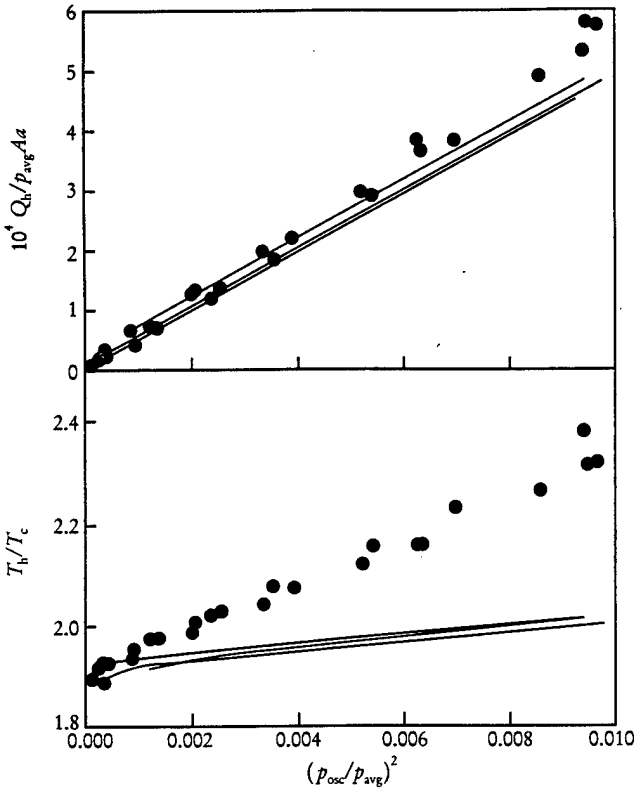
▷ Viscous loss in the stack occurs because work must be done to overcome viscous shear forces as the gas oscillates between the plates. The viscous penetration depth $\delta_\mu = \sqrt{\mu/\pi f\rho}$ (where μ is viscosity) is comparable to the thermal penetration depth, so most of the space between the plates experiences significant viscous shear (and the parcels in figures 1 and 2 actually experience significant shape distortions).

▷ Simple heat conduction from the hot heat exchanger to the cold one through the stack material and the gas is a further loss.

▷ These first three losses also occur in auxiliary parts of a thermoacoustic system: Viscous and inherent losses in the heat exchangers and conduction loss in the portion of the resonator case surrounding the stack are usually the most important auxiliary losses in large systems, while viscous and inherent losses on other surfaces in the resonator are important in small systems.

▷ Electroacoustic power transducers introduce additional loss. For the refrigerators shown on the cover and in figures 3a and 4a, the dominant transducer loss is Joule heating in the copper wires of the loudspeakers.

For many high-power-density designs, the first four sources of irreversibility contribute roughly equally to the inefficiency of thermoacoustic devices. About 40% of Carnot's efficiency is typical of the best current designs, using computer modeling based on Rott theory, for both engines and refrigerators; higher efficiencies are possible if power density is sacrificed, and lower efficiencies are the reality when electroacoustic transduction losses, losses in auxiliary equipment and deviations from Rott theory at high amplitudes are included. The most efficient engine built to date delivered acoustic power to its load at 23% of Carnot's efficiency (based on total heater power); the most efficient refrigerator[9] provided gross cooling power at 20% of Carnot (based on acoustic power delivered to the resonator). These efficiencies are impressive for devices with no moving parts, and they are comparable to the efficiencies of small, inexpensive commercial equipment. They



RECENT MEASUREMENTS AND CALCULATIONS for a thermoacoustic engine similar to that shown in figure 1, but with no transducer. The heater power Q_h, hot temperature T_h and oscillating pressure amplitude p_{osc} are all normalized, but note the different vertical scales. Circles are measurements from ref. 16; lines are calculations done using ref. 5. Red is for helium at 0.96 megapascals; green, neon at 0.70 MPa; blue, argon at 0.359 MPa. These pressures were chosen to make the thermal penetration depth δ_κ equal to 0.22 mm at the cold end (T_c) in all cases. FIGURE 5

fall far short of the efficiencies of well-engineered, expensive steam turbines or large-scale vapor-compression refrigeration equipment, for which over 80% of Carnot's efficiency has been achieved.

If future inventions and improvements to basic understanding can improve the efficiency or raise the power density of thermoacoustic engines and refrigerators without sacrificing their simplicity, they will find more widespread use. One way to increase efficiency, first demonstrated by Hofler,[4] is to use a mixture of helium and a heavier inert gas as the working substance. The Prandtl number $\sigma = \mu c_p/\kappa = \delta_\mu^2/\delta_\kappa^2$ is a dimensionless measure of the ratio of viscous to thermal effects in fluids; lower Prandtl numbers give higher efficiencies in thermoacoustics. Kinetic theory predicts $\sigma = \frac{2}{3}$ for hard-sphere monatomic gases, and indeed real monatomic gases have values very close to this. (For instance, helium at room temperature has $\sigma = 0.68$.) Fortunately mixtures of a heavy and a light monatomic gas have Prandtl numbers significantly lower than $\frac{2}{3}$. The thermoacoustic refrigerator intended for the space shuttle that was mentioned above[6] used 89% helium and 11% xenon, with $\sigma = 0.27$, to achieve its 20% efficiency, compared with 17% efficiency for the similar apparatus shown in figure 3 when optimized for pure helium gas. However, with a sound speed less than half that of pure helium, the gas mixture reduced the power density.

Beyond the basics

In the US our understanding of thermoacoustics is advancing beyond the foundations established by Rott, thanks to physicists at many universities and national laboratories.

Rott and his collaborators considered two geometries for thermoacoustic processes: parallel-plate channels (most commonly used, as we have discussed above) and circular channels. Wondering if some geometries might be better than others, W. Patrick Arnott, Henry Bass and Richard Raspet[10] at the University of Mississippi added rectangular and triangular channels, established a common formalism for all channel geometries and concluded that parallel-plate channels are the most efficient. The reason is subtle: Viscous losses occur mostly at and near channel walls, *within* a characteristic distance equal to the viscous penetration depth δ_μ, while the desirable thermoacoustic effects portrayed in figures 1 and 2 occur mostly away from walls, at a characteristic distance equal to the thermal penetration depth δ_κ *from* them. Thus for the usual case of $\delta_\mu \approx \delta_\kappa$, extremely concave channels (imagine triangles) squeeze the desirable effects into a small fraction of the channel cross-sectional area in the center, leaving a relatively large area near the perimeter causing viscous loss. Capitalizing on this analysis, Ulrich Müller has proposed that the "channels" formed by the space in a two-dimensional array of parallel wires[11] (aligned along the direction of acoustic oscillation and spaced by a few δ_κ) would give even higher efficiency than parallel plates. Tapered channels[12] and modifications of the phase between pressure and velocity[13] are also being studied for improving efficiency.

The most promising route to higher power densities is increasing $(p_{osc}/p_{avg})^2$, but doing so will take us further from the range of small oscillations on which Rott theory and its current extensions are fundamentally based. Rott's assumptions include the following:

▷ a gas displacement amplitude much smaller than the length of the stack and other components

▷ a Reynolds number of the oscillations small enough to avoid turbulence

▷ $p_{osc} \ll p_{avg}$.

Extensions beyond each of these limitations have begun.

At the Naval Postgraduate School, Anthony Atchley and his colleagues[14] have undertaken high-amplitude experiments on the simplest possible thermoacoustic device, a short stack with no heat exchangers in a loudspeaker-driven resonator. Their data extend into the large-gas-displacement regime, with amplitudes approaching the length of the stack.

A growing body of literature is establishing the characteristics of several regimes of turbulent oscillatory flow[15] at high Reynolds numbers, although as yet there are no fundamental studies of heat transfer under these conditions.

A similitude study has shown[16] how to organize and correlate experimental data in the high-amplitude range, allowing meaningful experimental studies of scale models of thermoacoustic devices reminiscent of wind-tunnel studies of model aircraft.

Researchers at several institutions, including Johns Hopkins University and Los Alamos and Livermore National Laboratories, are beginning numerical and analytical studies that seek to illuminate other features of high-amplitude thermoacoustics.



APD's 6.5 Kelvin Displex®

The Displex refrigerator, long regarded as the industry's 10 Kelvin standard, just got colder. This next generation of two-stage expanders is capable of 6.5 K operation with more than three times the refrigeration capacity at 10 K. APD is now the only company with a complete line of closed cycle 4 K, 6.5 K, 10 K and 35 K refrigerators to meet all your specific needs.

So...to get colder, call APD Cryogenics at 800-525-3072.

APD CRYOGENICS INC
A SUBSIDIARY OF INTERMAGNETICS GENERAL CORPORATION

1833 Vultee Street • Allentown, PA 18103 • (610) 791-6700 • FAX: (610) 791-0440

Circle number 14 on Reader Service Card

R&D

Our n&k Analyzer Got The Award*
You Get The Results

Semiconductor, Dielectric, and Metal Films

- Thickness
- *n* and *k* Spectra, 190 nm to 900 nm

Applications include characterization of a-C:H, a-Si:H, SiO_xN_y:H, SiO_2 / Poly-Si / SiO_2, ITO, SOI, Ti, TiN, and Ag. Substrate can be opaque or transparent, smooth or rough.

Not an Ellipsometer

n&k Technology, Inc.

3150 De La Cruz Blvd., Suite 105 · Santa Clara, CA 95054
Tel: (408) 982-0840 · Fax: (408) 982-0252

*Selected by R&D Magazine as one of the 100 most technologically significant products of the year.

Circle number 15 on Reader Service Card

Figure 5 illustrates the amplitude dependence of the accuracy of our current understanding of thermoacoustic phenomena. The data shown as points were taken from a no-load thermoacoustic engine.[16] The plots can be interpreted as displaying the required heater power Q_h and temperature T_h at the hot heat exchanger for maintaining steady oscillations at a given pressure amplitude p_{osc}. Three monatomic gases were used, with average pressures selected to make δ_κ the same for all three cases, ensuring similarity; the data for all three gases do indeed fall along the same curves. The lines are the results of calculations based on Rott's theory, the dimensions of the apparatus and the properties of the gases. The calculations agree well with the measurements in the limit of small p_{osc}, as expected from the assumptions in the theory. However, as p_{osc}/p_{avg} approaches 0.1, the measurements deviate significantly from calculations. The deviations are not surprising, in that they are of the same order as p_{osc}/p_{avg}, but they are disturbing from a practical point of view because both Q_h and T_h deviate in directions that decrease the efficiency.

The fundamentals of thermoacoustics at low amplitudes are reasonably well understood, and a few practical uses of thermoacoustics have been tentatively identified. Much study, engineering and especially invention remains to be done before these simple, elegant devices reach their full potential.

Most of the fundamental research on thermoacoustics in the US is supported by the Department of Energy and the Office of Naval Research. Most of the applied developments are supported privately but with important contributions from ARPA, DOE and the Navy. I am particularly grateful to DOE's Office of Basic Energy Sciences for its steady support of thermoacoustics research at Los Alamos. This article benefited from constructive criticism by Hank Bass, Steve Garrett and Tom Hofler.

References

1. K. T. Feldman, J. Sound Vib. **7**, 71 (1968).
2. N. Rott, Z. Angew. Math Phys. **20**, 230 (1969); **26**, 43 (1975). Reviewed by G. W. Swift, J. Acoust. Soc. Am. **84**, 1145 (1988).
3. J. C. Wheatley, T. J. Hofler, G. W. Swift, A. Migliori, J. Acoust. Soc. Am. **74**, 153 (1983).
4. T. J. Hofler, PhD dissertation, U. Calif., San Diego (1986). T. J. Hofler, in *Proc. 5th Int. Cryocoolers Conf.*, P. Lindquist, ed., Wright–Patterson Air Force Base, Ohio (1988), p. 93.
5. W. C. Ward, G. W. Swift, J. Acoust. Soc. Am. **95**, 3671 (1994). Fully tested software and users guide available from Energy Science and Technology Software Center, US Dept. of Energy, Oak Ridge, Tenn. For a beta-test version, contact ww@lanl.gov (Bill Ward) via Internet.
6. S. L. Garrett, D. K. Perkins, A. Gopinath, in *Heat Transfer 1994: Proc. 10th Int. Heat Transfer Conf.*, G. F. Hewitt, ed., Inst. Chem. Eng., Rugby, UK (1994), p. 375.
7. R. Radebaugh, Adv. Cryogenic Eng. **35**, 1191 (1990).
8. A. Bejan, *Entropy Generation Through Heat and Fluid Flow*, Wiley, New York (1982).
9. S. L. Garrett, J. A. Adeff, T. J. Hofler, J. Thermophys. Heat Transfer **7**, 595 (1993).
10. W. P. Arnott, H. E. Bass, R. Raspet, J. Acoust. Soc. Am. **90**, 3228 (1991).
11. U. A. Müller, US patent 4 625 517 (1986). G. W. Swift, R. M. Keolian, J. Acoust. Soc. Am. **94**, 941 (1993).
12. N. Rott, G. Zouzoulas, Z. Angew. Math. Phys. **27**, 197 (1976). U. A. Müller, PhD dissertation 7014, Eidgenössische Technische Hochschule, Zurich, Switzerland (1982).
13. R. Raspet, H. E. Bass, J. Kordomenos, J. Acoust. Soc. Am. **94**, 2232 (1993). P. H. Ceperley, J. Acoust. Soc. Am. **66**, 1508 (1979).
14. A. A. Atchley, T. J. Hofler, M. L. Muzzerall, M. D. Kite, C. Ao, J. Acoust. Soc. Am. **88**, 251 (1990).
15. R. Akhavan, R. D. Kamm, A. H. Shapiro, J. Fluid Mech. **225**, 395, 423 (1991).
16. J. R. Olson, G. W. Swift, J. Acoust. Soc. Am. **95**, 1405 (1994). ■

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	31 Jan 97	Final 01 Feb 96 - 31 Jan 97

4. TITLE AND SUBTITLE
Proceedings of the 1997 Physical Acoustics Summer School: Volume III: Background Materials

5. FUNDING NUMBERS
PE 61153N
G N00014-96-1-0033

6. AUTHOR(S)
Henry E. Bass

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Mississippi
NCPA
University, MS 38677

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
ONR 331
800 North Quincy Street
Arlington, VA 22217-5660

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release: Distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Volume III of the Proceedings of the 1997 Physical Acoustics Summer School contains abstracts and copies of publications which were useful to the graduate students who attended the summer school. These materials helped the students prepare for the lectures which can be found in transcription form in Volume I of these proceedings. Volume II of these proceedings contains copies of the transparencies used by the lecturers.

14. SUBJECT TERMS
Physical Acoustics, Molecular Acoustics, Acoustics Demonstrations, Resonant Ultrasound Spectroscopy, Sonoluminescence, Nonlinear and Atmospheric Acoust., Sensor Phys., Thermoacoustics

15. NUMBER OF PAGES
377

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	