PHYSICAL ACOUSTICS SUMMER SCHOOL

ASILOMAR CONFERENCE CENTER

VOLUME III: BACKGROUND MATERIALS

1998 PHYSICAL ACOUSTICS SUMMER SCHOOL

VOLUME III: BACKGROUND MATERIALS

19991227 101

TABLE OF CONTENTS

ACOUSTICS DEMONSTRATIONS

Robert M. Keolian

Penn State University Applied Research Laboratory
and Graduate Program in Acoustics

Some of the nonlinear properties of the harmonic oscillator, waves on the surface of water, and of high amplitude sound will be shown in a series of demonstrations:

- Bent tuning curves of a rubber band
- Parametrically driven pendulum
- Doubly bent tuning curves
- Parametric stabilization of an inverted pendulum
- Non-propagating hydrodynamic soliton
- Shock waves, N waves, and sound eating sound

- Acoustic Bernoulli effect
- Acoustic levitation
- Acoustic match blower
- Izzy Rudnick's lawn sprinkler
- Izzy Rudnick's tea party
- Acoustic log starter

REFERENCES

A.B. Pippard's book, *The Physics of Vibration*, (Cambridge University Press, Vol. 1, 1st edition, 1978 or the omnibus edition, 1989), chapters 9-12, is a good reference for nonlinear oscillators, parametric excitation, and other interesting effects.

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

The soliton is described in: J. Wu, R. Keolian, and I. Rudnick, "Observation of a Non-Propagating Hydrodynamic Soliton," *Phys. Rev. Lett.* **52**, 1421-1424 (1984) and in J. Wu, I. Rudnick, "Amplitude-Dependent Properties of a Hydrodynamic Soliton," *Phys. Rev. Lett.* **55**, 204 (1985).

ACOUSTIC RESONATORS AND THE PROPERTIES OF GASES

BY

MICHAEL R. MOLDOVER

Objectives

To identify most of the phenomena that *may* be encountered in gas-filled acoustic resonators operating in the linear regime ($P_{acoustic} \ll P_{ambient}$). To motivate the choices made when designing measurement systems and when estimating the properties of gases. To exploit the analogies and differences between sound waves and electromagnetic waves. Not covered: detailed calculations, derivations.

Resonators [1]

1. Ideal frequencies
2. Shape perturbations, degeneracies
3. Visco-thermal boundary layers
4. Bulk viscosity
5. Mechanical compliance, shell resonances, avoided crossings; why resonators are rarely used to measure the properties of liquids.
6. Inefficient transducers are useful; waveguides and diaphragms [2]
7. Mean free path effects: accommodation at the gas-solid boundary *et al.*
8. Standards Measurements: The gas constant R and the thermodynamic temperature T [1]

Properties of Gases

1. Thermodynamic speed of sound [3];
     A. speed of sound measurements for technical applications
     B. know your impurities
2. Connections between gas properties and intermolecular potentials [3,4]
     A. virial coefficients
     B. viscosity
     C. thermal conductivity
     D. bulk viscosity [5]
     E. helium: a standard for gas properties [6]
3. Mixtures; why I wonít measure the speed of sound of air
4. Acoustic measurements of transport properties: status report [7]

Primary Reference

Moldover, M.R., Mehl, J.B. and Greenspan, M., Gas-filled spherical resonators - Theory and experiment, J. Acoustical Soc. of America, Vol. 79, pp. 253-272 (1986)

Other References

[1] Moldover, M.R., Trusler, J.P.M, Edwards, T.J., Mehl, J.B., and Davis, R.S., Measurement of the universal gas constant R using a spherical acoustic resonator, J. of Res. of NBS, Vol. 93, No. 3, pp. 85-114 (Mar-Apr 1988).

[2] Gillis, K.A., Moldover, M.R., and Goodwin, A.R.H., Accurate acoustic measurements in gases under difficult conditions, Rev. Sci. Instruments, Vol. 62, No. 9, pp. 2213-2217 (1991).

[3] Gillis, K.A. and Moldover, M.R., Practical determination of gas densities from the speed of sound using square-well potentials, Int. J. Thermophysics, Vol. 17, No. 6, pp. 1305-1324 (1996)

[4] Joseph O. Hirschfelder, Charles F. Curtiss, and R. Byron Bird, "Molecular Theory of Gases and Liquids," John Wiley & Sons, Inc., New York (1954) or J. P. M. Trusler, Equation of state for gaseous propane determined from the speed of sound, Int. J. Thermophysics, Vol 18, pp. 635-654 or J. P. M. Trusler, W. A. Wakeham, and M. P. Zarari, Molecular Physics, Vol. 90, pp. 695-703 (1997).

[5] R. Holmes and W. Tempest, The Propagation of sound in monatomic gas mixtures, Proc. Phys. Soc. London, Vol. 75, pp 898-904 (1960); K. F. Hertzfeld, and T. A. Litovitz, "Absorption and dispersion of ultrasonic waves," Academic Press, New York (1959).

[6] Aziz, R.A., Janzen, A.R., and Moldover, M.R., *Ab initio* Calculations for helium: a standard for transport property measurements, Phys. Rev. Lett., Vol. 74, No. 9, pp. 1586-1589 (1995).

[7] Gillis, K.A., Mehl, J.B., and Moldover, M.R., Greenspan acoustic viscometer for gases, Rev. Sci. Instrum., Vol. 67, No. 5, pp. 1850-1857 (1996).

Chaos and Nonlinear Bubble Dynamics

Werner Lauterborn
Drittes Physikalisches Institut, Universität Göttingen, Bürgerstr. 42-44,
D-37073 Göttingen, Germany, Tel: ++551-39-7713, Fax: ++551-39-7720,
E-mail: lb@physik3.gwdg.de

Chaos has become a learned word in science as the deterministic companion of a stochastic process. Nonlinearity lies at its basis and a certain degree of complexity. To enter the realm of chaos a number of concepts have to be looked at for deeper insight: The state space, trajectories and attractors therein, attractor basins, bifurcations of different kinds, in particular the period-doubling bifurcation and its cascade, bifurcation diagrams, the different routes to chaos, Poincaré sections and Poincaré maps, the connection between maps and differential equations, phase or parameter space diagrams, embedding of data (time series) into higher dimensional spaces as the link between experiment and a theoretical description (Fig. 1), fractal dimensions for chaotic attractors (static description), Lyapunov exponents of attractors as quantitative measures of the sensitive dependence of the dynamics on slight deviations (dynamic description), and even more involved concepts.
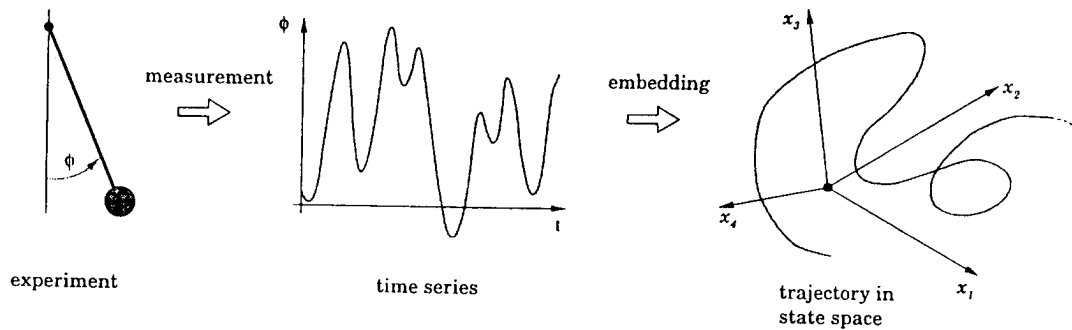

Fig. 1: The way from experiment to a trajectory in a state space by embedding.

From chaos theory a new method for characterizing systems has evolved as put forward graphically in Fig.2. Starting with some system to be investigated or monitored a measurement yields a time series, i. e. a series of samples typically equally spaced in time. These may be analyzed by the usual linear means as given by Fourier analysis and correlations. These are not discussed here. The new way of nonlinear analysis starts with the embedding as visualized in Fig.1 yielding a reconstructed state space. The embedding may be preceded by preprocessing steps (linear filtering for instance) or by tests on determinism or nonlinearity by constructing surrogate data from the original data set. Also after embedding the somehow scrambled new set may be processed, for instance to achieve noise reduction by nonlinear operations. However, the main aim of embedding is the characterization of the data in a way that surpasses the usual linear methods. This is done by determining the static and dynamic properties of the embedding set as they are given by fractal dimensions and Lyapunov exponents discussed in the lectures. A second way of characterization

may proceed via modelling for describing the system and for prediction of its future states on the basis of the embedding. The gained insight may also be used for controlling purposes. The final aim, however, is a complete description of the system in a way of diagnosis of its state at the time of measurement.
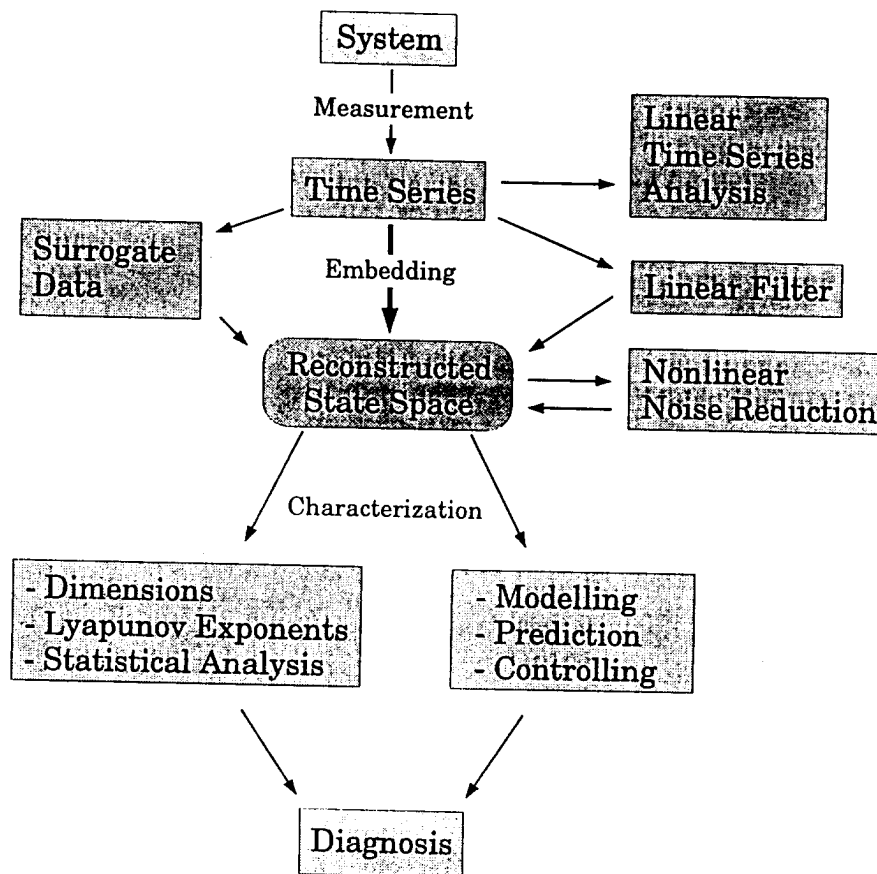

Fig. 2: Nonlinear time series analysis for system characterization via embedding.

Nonlinear oscillators are ideal objects where chaos shows up and where the concepts mentioned above yield considerable new insight as compared to a purely linear description, i. e. a description working with linearization only and linear transformations. In the context of acoustics the bubble in a sound field is a nonlinear oscillator, indeed a strongly nonlinear one at higher sound pressure amplitudes. It is even so peculiar as to emit light upon collapse and to damage any material when coming close to it. Bubble dynamics will be discussed both from the view of experiments and of theoretical descriptions and, of course, the view of chaos theory.

W. Lauterborn and U. Parlitz, Methods of chaos physics and their application to acoustics, *J. Acoust. Soc. Am.* **84** (1988) 1975-1993.
U. Parlitz, Englisch, V., Scheffczyk, C. and Lauterborn, W.: Bifurcation structure of bubble oscillators, *J. Acoust. Soc. Am.* **88** (1990) 1061-1077.
W. Lauterborn and J. Holzfuss, Acoustic chaos, *Int. J. Bifurcation and Chaos* **1** (1991) 13-26.
W. Lauterborn, Nonlinear dynamics in acoustics, *Acustica* **82** (1996) S-46 - S-55.

ACOUSTIC CHAOS

WERNER LAUTERBORN and JOACHIM HOLZFUSS
Institut für Angewandte Physik, Technische Hochschule Darmstadt
Schloßgartenstr. 7, D-6100 Darmstadt, F R Germany

The history of chaos pertaining to acoustics is briefly reviewed from the first period-doubling experiments reported by Faraday (vibrated liquid layer) and Melde (string oscillations) to today's investigations on chaos in thermoacoustics, musical instruments, the hearing process, and ultrasonics. A closer analysis is given of the sound produced in a liquid by standing acoustic waves of high intensity.

1. Introduction

Chaotic (i.e., irregular and unpredictable) motion seems to appear in any sufficiently complicated or complex dynamical system. Acoustics, that part of physics that describes (and makes use of) the vibrations of usually larger ensembles of molecules in gases, liquids and solids makes no exception. As a main necessary ingredient of chaotic dynamics is nonlinearity, acoustic chaos can only be found in the realm of nonlinear acoustics. Thus, acoustic chaos is closely related to nonlinear vibrations of and in gases, liquids and solids. It is the science of never-repeating sound waves. This property it shares with noise, a term having its origin in acoustics and formerly attributed to every sound signal (and meanwhile other, e.g., electrical, signals) with a broadband Fourier spectrum. But Fourier analysis is especially adapted to linear oscillatory systems. There it develops its full power. It is the standard interpretation of the lines in a Fourier spectrum that each line corresponds to a (linear) mode of vibration and a degree of freedom of the system. However, as examples from chaos physics show, a broadband spectrum can already be obtained with just three (nonlinear) degrees of freedom (i.e., three dependent variables). Chaos physics thus develops a totally new view of the noise problem. It is a deterministic view. But it is still an open question how far the new approach will reach in explaining still unsolved noise problems, e.g., the $1/f$-noise spectrum encountered so often. The detailed relationship between chaos and noise is still an area of active research. An example, where the properties of acoustic noise could be related to chaotic dynamics, is given below for the case of acoustic cavitation noise.

2. Historical notes

Acoustic chaos and the closely related science of chaotic vibrations are young disciplines [Lauterborn, 1989; Lauterborn & Parlitz, 1988; Parker & Chua, 1989; Thompson & Stewart, 1986; Moon, 1987]. Yet most experiments conducted today have ancient ancestors. Two classes of oscillatory systems are most susceptible to chaotic motion. These are periodically driven (passive) nonlinear systems, which often respond to an excitation with subharmonics, in particular in the form of period doubling, and self-excited systems, which develop sustained oscillations from seemingly constant exterior conditions. A special subclass of the second class forms self-excited systems when driven. The simplest model of that class is the driven van der Pol oscillator (see, e.g., Parlitz & Lauterborn [1987]). From real physical systems, the weather can be put into this category. It is periodically driven by the solar radiation with the low period of 24h. But already a constant heating leads to oscillations (Rayleigh–Benard convection), i.e., it is a self-excited system.

The first reported subharmonic oscillation of order one half ($f/2$, f= driving frequency) belongs to the first class and dates back to Faraday [1831]. Starting from the investigation of sound-emitting, vibrating

surfaces with the help of Chladni figures, Faraday also sprinkled water instead of sand onto his vibrating plates (vibrating at frequency f), and extended the work to complete layers of fluid to investigate the "beautifully crispated appearance" of the liquid layer. His aim was the "progress of acoustical philosophy". To be able to watch the motion of the fluid layer, he enlarged his vibrating plates to lower the frequency of oscillation, and ultimately came up with a board eighteen feet long, upon which a liquid layer of three quarters of an inch in depth and twenty-eight inches by twenty inches in extent could be vibrated vertically. Then, by ordinary inspection, he could observe that the heaps of liquid making up the crispations where oscillating in a sloshing motion to and fro between neighboring heaps of liquid. He states: "Each heap (identified by its locality) recurs or is re-formed in two complete vibrations of the sustaining surface" and adds in a footnote "A vibration is here considered as the motion of the plate, from the time that it leaves its extreme position until it returns to it, and not the time of its return to the intermediate position". This result was confirmed by Rayleigh [1883a].

Today, the powerful methods of nonlinear dynamics and computerized experimental instrumentation are applied to this problem, and the nonlinear and chaotic oscillations are investigated in some detail [Keolian et al. 1981; Miles, 1984; Ciliberto & Gollub, 1985; Meron & Procaccia, 1986; Gu & Sethna, 1987]. Subharmonic oscillations as low as $f/35$ ($f =$ driving frequency), quasiperiodic and chaotic oscillations have been observed in an essentially one-dimensional variant of the Faraday experiment [Keolian et al., 1981]. Period doubling, quasiperiodicity and chaos have been observed for the original two-dimensional case with a cylindrical fluid layer of 1 cm in depth and radius 6.35 cm [Ciliberto & Gollub, 1985]. Mode competition is singled out as the mechanism underlying the complex dynamics. This experiment may naturally occur on the curved surface of bubbles set into motion in a sound field (see Sec. 3.4 below).

A second topic, where subharmonics were observed early, is the Melde experiment [Rayleigh, 1883b, 1887]. In this experiment, the tension of a string is modulated periodically by fastening it to the prong of a tuning fork. Under suitable conditions, the string will vibrate at half the driving frequency. Rayleigh developed a theory for parametrically driven systems of this kind. String vibrations play an important role in musical instruments, and the investigation of nonlinear string vibrations form a large, separate part in nonlinear dynamics (see Tufillaro [1989] to enter the field).

Before going ahead with periodically driven systems, which, when nonlinear, almost always seem to develop chaos in some parameter region, we turn to the second class, that of self-sustained vibrations. Among these, we pick the thermoacoustic oscillations which come in two main varieties: Sondhauss oscillations [Sondhauss, 1850; Feldmann, 1968a] and Rijke oscillations [Rijke, 1859; Feldmann, 1968b]. The Sondhauss oscillation occurs when the closed end of a gas-filled pipe is heated (externally or internally), or, conversely, when the open end is cooled (then called Taconis oscillations [Taconis et al., 1949]). The Rijke oscillation occurs when an internal grid located in the lower half of a vertical pipe is heated. Both ends must be open to allow for a flow of gas (self-generated or enforced) through the pipe. Sound is also produced when the grid, this time being located in the upper half of the vertical pipe, is cooled [Riess, 1859].

Taconis oscillations with steep temperature gradients and large temperature difference show complex behavior [Yazaki et al., 1987]. The oscillation may period double, develop quasiperiodic oscillations through the appearance of a second incommensurate frequency and also chaotic oscillations. This behavior is attributed to mode competition similar to the Faraday experiment of Ciliberto and Gollub [1985]. When Taconis oscillations are confronted with periodic, externally imposed acoustic oscillations, the whole set of nonlinear dynamical phenomena seems to occur [Yazaki et al., 1990], as encountered, when self-excited systems are additionally driven (see, e.g., the driven van der Pol oscillator [Parlitz & Lauterborn, 1987]). These acoustic oscillations are not just of scientific interest, but have a potential as "natural engines" [Wheatley & Cox, 1985].

Coming back to driven systems of the passive (i.e., not self-excited) type, period doubling has been observed in connection with loudspeakers. Conical loudspeakers, when strongly driven, start to period double [Pedersen, 1935]. Subharmonics have later been observed when a liquid is irradiated with a pure tone of high intensity [Esche, 1952]. This work, where frequencies $f/2$, $f/4$, and $f/3$ of the driving acoustic sound wave of frequency f were found in the output sound from the liquid (Fig. 1), stimulated the discussion as to the physical mechanism producing these subharmonic frequencies. This topic will be investigated in more detail below.

Nonlinear dynamics also appears in musical instruments, where oscillations are produced to generate sound. These systems are therefore a priori apt to generate chaotic sound waves. But there are not many investigations in this area. A first account can be
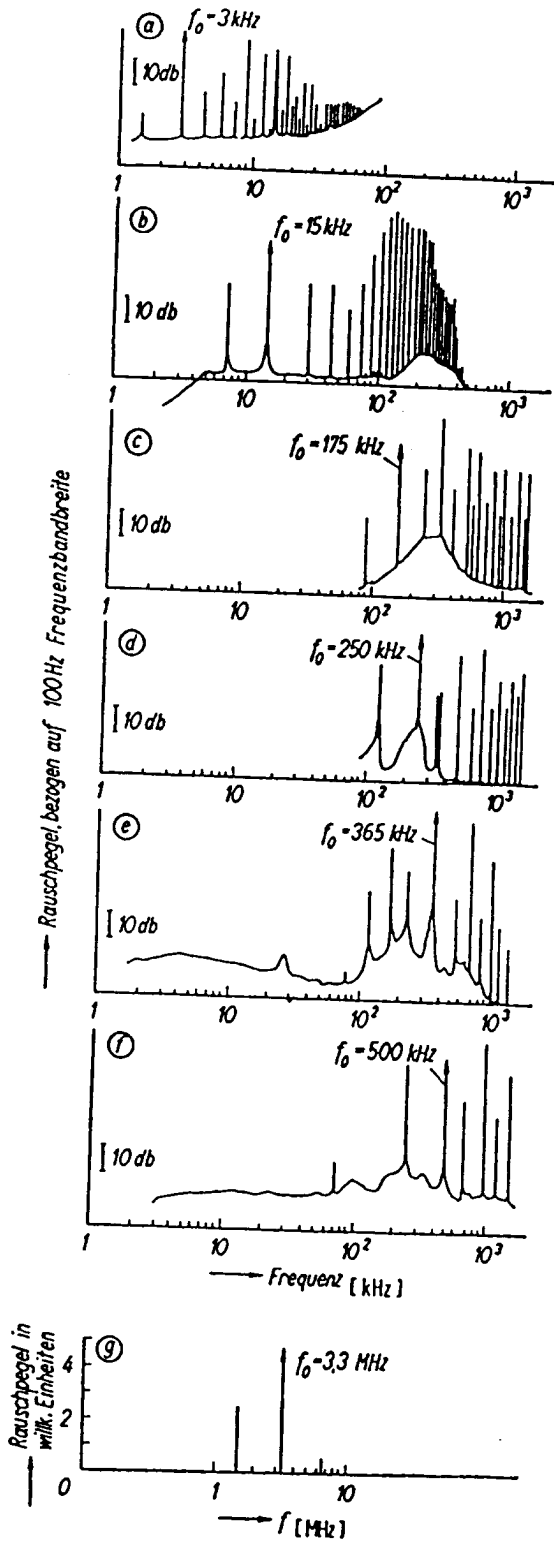
Fig. 1. The historical spectra of acoustic cavitation noise reported by Esche in 1952.

instruments seem capable of producing chaotic sound when properly excited: clarinet-like systems [Maganza *et al.*, 1986], woodwind instruments (recorder) [Gibiat, 1988], and gongs [Legge & Fletcher, 1989].

The human hearing process is intrinsically nonlinear. To account for the astonishing power of the ear, active filtering, i.e., amplifiers, are assumed to be at work (see, e.g., Bialek & Wit [1984]). Such systems may get unstable and start to oscillate. Indeed, it has been found that the ear may also emit sound, either spontaneously ("objective tinnitus" [Kemp, 1979]) or (click) evoked ("Kemp-echo" [Kemp, 1978]). These otoacoustic emissions had been predicted by Gold as early as in 1948 [Gold, 1948] and have been an area of active research in the past ten years. The interaction of spontaneous emissions with external sounds has been studied [Rabinowitz & Widin, 1984] and rather sophisticated models have been proposed to account for evoked (i.e., stimulated) emissions [Strube, 1989]. As the problem belongs to the type of driven self-oscillatory systems, chaotic dynamics can be expected to be present in the appropriate parameter ranges.

An interesting area is the area of opto- and photoacoustics, i.e., the area of the interaction of light with matter, yielding acoustic waves. The area seems not yet influenced by the nonlinear dynamics virus (see the recent review on sound generation by optical radiation [Sigrist, 1986]), but it is easy to envisage what may happen when modulated laser light is used to drive acoustic resonant systems or flow systems with instabilities.

The large field of noise in combustion processes is omitted here, as well as the area of plasma physics, where, e.g., ion acoustic waves occur and there is a sheer endless list of instabilities. They need a special treatment.

Surely, the list showing where acoustic chaos can be found is not complete, and we apologize for all topics and works which are omitted. But we are eager to learn of any new development in the area of acoustic chaos.

3. Acoustic Chaos in Ultrasonic Cavitation

3.1. Historical notes

When a liquid is irradiated with sound of high intensity, a hissing noise is heard. This phenomenon was the starting point of Esche's work [Esche, 1952]. The noise emission presents an interesting physical problem, as a sound wave of a single frequency is transformed into a broadband sound spectrum. At the time of Esche, nothing was known of deterministic chaos, and no physical mechanism could be thought of to explain this fact

found in a review on the oscillations of musical instruments [McIntyre *et al.*, 1983]. Several musical

besides the statistical rupture of the liquid. Esche's work showed that the phenomenon of broadband noise-emission is preceded by strong subharmonic spectral lines at 1/2, 1/4 and also 1/3 the driving frequency and their harmonics. The explanation of these lines presented even more difficulties, and was sought in the nonlinear dynamics of the bubbles known to be generated in the process. For the subharmonic of order 1/2, it was suggested that bubbles driven at twice their resonance frequency should be responsible for the observed frequency halving [Güth, 1956]. This could be confirmed by numerical calculations [Eller & Flynn, 1969; Lauterborn, 1969/70], and it was even speculated that there might be a connection between subharmonics and noise [Lauterborn, 1969/70]. However, serious difficulties arise with this explanation, because bubbles of the necessary size (double resonance size) are not stable in a standing acoustic field as they are driven away from the pressure antinodes. Later, it was found that bubbles driven at frequencies lower than their resonance frequency could also oscillate subharmonically in regions of ultraharmonic resonances, and also that there are regions in parameter space of driven bubbles where periodic solutions seem to be lacking [Lauterborn, 1976].

Experiments done to improve the material available from Esche's work led to the observation of a period-doubling sequence to broadband noise when the sound intensity of the driving sound field was raised [Lauterborn & Cramer, 1981], as well as other sequences hard to classify (see Lauterborn [1986]). In these experiments, water had been used as a liquid. But almost the same phenomena can also be observed in liquid helium [Smith *et al.*, 1982]. Period-doubling sequences (already to be detected in Lauterborn [1976]) are also obtained numerically for (purely radially) oscillating bubbles [Lauterborn & Suchla, 1984; Parlitz *et al.*, 1990]. Experimental observation of the liquid by high-speed holographic cinematography simultaneous with the acoustic measurements revealed that the complete bubble field is undergoing period doubling to chaos [Lauterborn & Koch, 1987]. Specific methods of nonlinear dynamics have been applied to the acoustic time series, which surprisingly gave evidence of a low-dimensional strange attractor through phase space analysis in conjunction with dimension estimation [Lauterborn & Holzfuss, 1986] and the determination of Lyapunov exponents [Holzfuss & Lauterborn, 1989].

In the sequel, the paper reports on advances in ultrasonic acoustic chaos. Phase space representations of acoustic signals and a color spectral bifurcation diagram of measured data, which exemplifies the successive period doubling of the sound output from the liquid, are given.
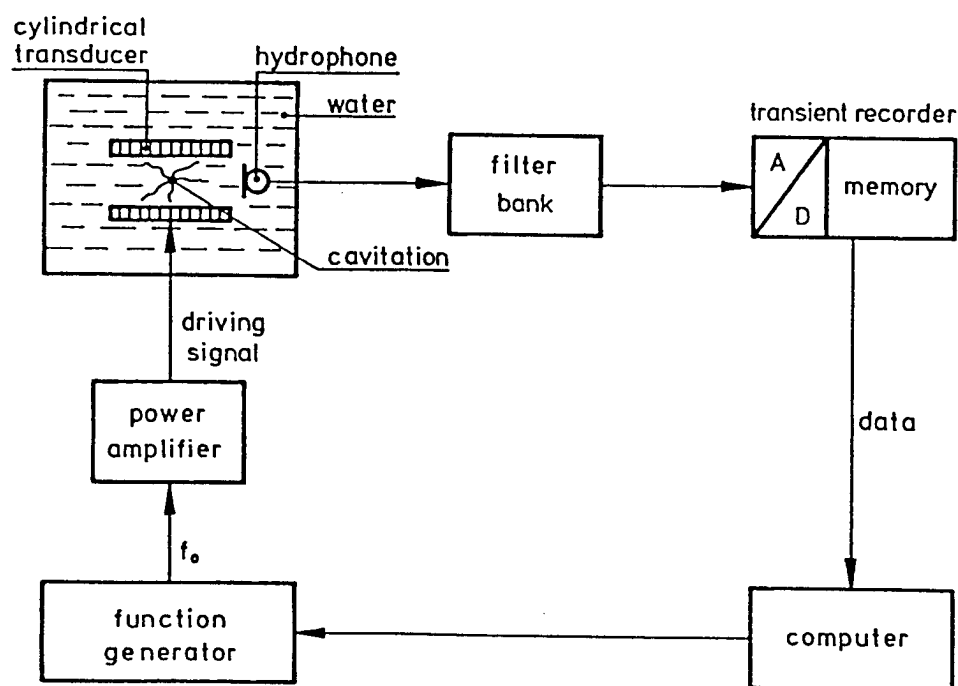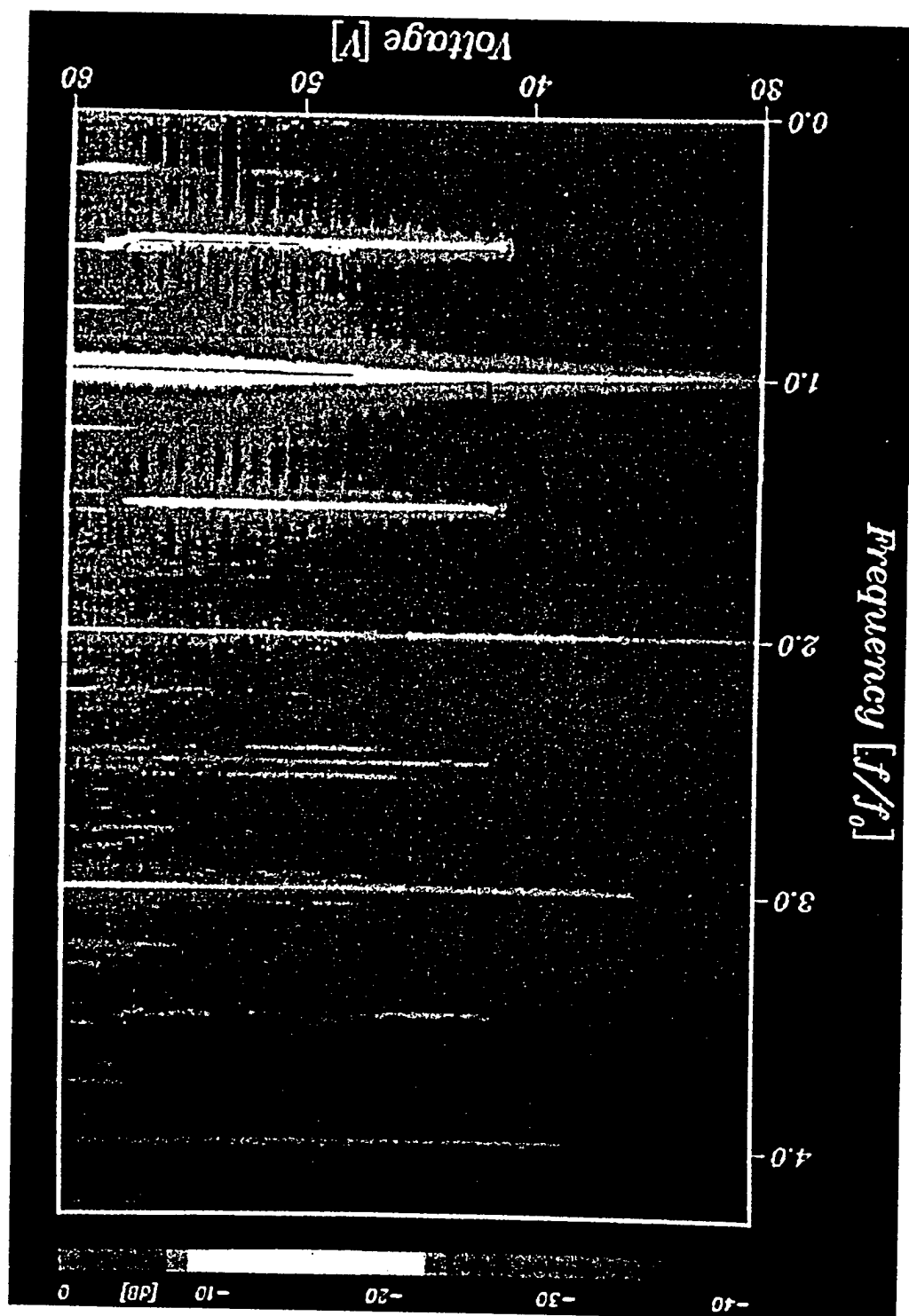

Fig. 2. Experimental arrangement for acoustic measurement of driven liquids.

Fig. 3. Spectral bifurcation diagram of the sound output from sound-irradiated water at $f_0 = 22.9$ kHz. Region of period-doubling sequence.

3.2. Acoustic observations

The experimental arrangement for investigating the sound emission from a sound-irradiated liquid is depicted in Fig. 2. The high acoustic intensity necessary to observe the nonlinear phenomena is obtained with a cylindrical transducer of piezoceramic material (PZT-4) of 76 mm length, 76 mm inner diameter, and 5 mm wall thickness. Depending on the container and the water height above the cylinder, the resonance frequency for half a wavelength across the diameter of the cylinder is slightly different, and was 22.9 kHz in the present case. The liquid used was deionized, air-saturated water with added nuclei (Vestamid X2805, a polycrystalline substance of about the same density as that of water) of size smaller than 40 μm at a concentration of about 15 $\mu g/cm^3$.

The experiment is fully computer-controlled, and consists of a function generator to produce the sinusoidal frequency f_0 of the driving sound field and a power amplifier to obtain the necessary amplitudes. The driving amplitude is used as control parameter, and usually made a linear ramp function. A broadband hydrophone records the sound output from the liquid. After suitable filtering, the pressure time data are digitized with 10 bits at rates of usually 500 kHz to 1 MHz, and stored in a memory of 4K to 128K words.

Figure 3 shows a spectral bifurcation diagram of a measurement, where the sound pressure has been raised within 256 ms from 0 to its maximal value at 60 V transducer voltage. Only the part from 30–60 V is shown. The spectral bifurcation diagram consists of successively plotted power spectra of 4K data windows shifted along the whole 128K dataset. The amplitudes are color encoded. At low amplitudes, the spectrum consists of only sharp lines at the driving frequency f_0 and its harmonics $2f_0$, $3f_0$... At 41 V, the first subharmonic at $f_0/2$, together with its harmonics, appears. Further increase of the driving amplitude leads to a further frequency halving to $f_0/4$ at 50–52 V and again at 54 V. From about 56 V on, more and more broader lines appear and the spectral intensity is distributed over a broad frequency range. The spectrum is then a broadband noise spectrum.

Phase space analysis has been applied to the data. Figure 4 shows trajectories constructed in a phase space that consists of time-shifted sound pressure values $p(t)$. A trajectory in this phase space is given by the set of 3-tupels $\{p(t), p(t+T), p(t+2T)\}$. The time shift constant T is here approximately 1/10 of the driving period T_0. It can be shown [Takens, 1981] that this set yields the same properties as the set of state points in the original phase space. A period-doubling route to chaos with increasing sound pressure is observed (Figs. 4a–4f). Figure 4a shows a period-1 orbit that doubles into a period-2 orbit (b) and a period-4 orbit (c). In (d) the trajectories no longer seem to form a periodic orbit. In (e) this becomes more pronounced, and in (f) the aperiodic feature is clearly visible. The dimension of these objects in phase space can be determined. It is a measure of the active modes of the underlying physical process [Lauterborn & Holzfuss, 1986]. A very suitable method is the determination of the average of pointwise dimensions [Holzfuss & Mayer-Kress, 1986]. A pointwise dimension D at a point in phase space can be calculated by looking at the scaling behavior of the "mass" around this point, that is, the number of points within a certain distance r. The number of points in this ball should scale like

$$N(r) \sim r^D , \quad r \to 0 .$$

Their average dimension and standard deviation are characteristic of the attractor. In Fig. 5, it is seen that the dimension of periodic attractors is, as expected, equal to 1. However, in the broadband noise region, a fractal value of 2.6 is obtained. This surprisingly low value for the sound output of thousands of oscillating bubbles is an evidence for low-dimensional chaos.

For a further justification of this finding, the spectrum of Lyapunov exponents has been calculated from the noise data. Lyapunov exponents are the basic indicator for chaos [Eckmann & Ruelle, 1985]. They measure the average exponential expansion and contraction rates of small volume elements around the attractor. A physical process, whose motions in phase space lie on an attractor with a positive Lyapunov exponent, has the property that, initially, almost similar states may lead to a totally different time evolution.

To calculate Lyapunov exponents from experimental data, a method based on a least-squares approximation of the linearized flow maps has been proposed [Eckmann & Ruelle, 1985; Sano & Sawada, 1985; Holzfuss & Lauterborn 1989]. Figure 6 shows exponents as a function of the embedding (matrix) dimension for noise data at 58 V transducer voltage. With increasing matrix dimension, the calculated values converge to $\lambda_1 = 1.9$, $\lambda_2 = 0$, $\lambda_3 = -3$ bits per period of the driving sound wave. The positive value of the first Lyapunov exponent is a further indication for low-dimensional chaos being the explanation for the observed broadband noise.
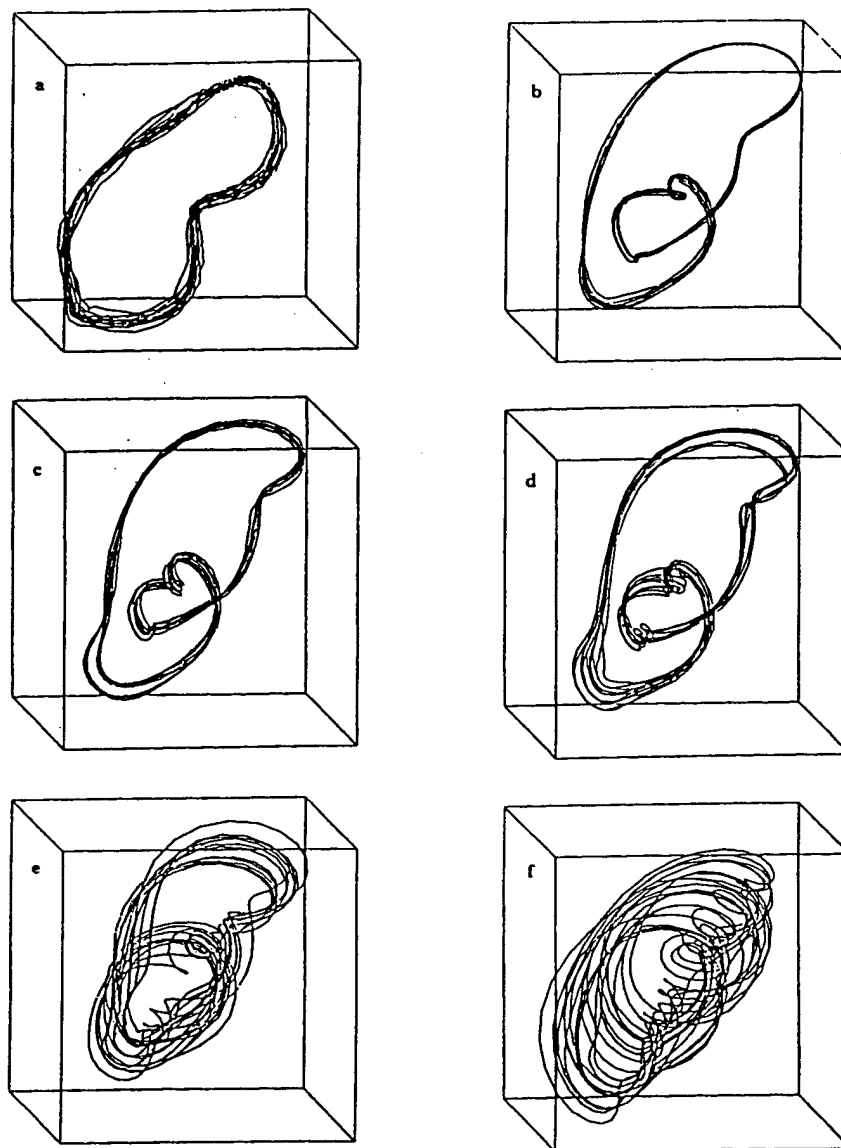
11

Fig. 4. Phase space reconstruction from the sound output of sound-irradiated water at $f_0 = 22.9$ kHz. Projections of trajectories in a phase space, constructed with time-shifted coordinates. The transducer voltage is (a) 30.7 V (b) 52.8 V (c) 55 V (d) 56 V (e) 56.4 V (f) 58.8 V.

3.3. *Holographic observations*

Thus far, only the acoustic signal has been considered. An optic inspection of the liquid inside the piezoelectric cylinder reveals that a structured, branch-like cloud of bubbles is present at high driving (see, e.g., Lauterborn [1986]). It is a straightforward conjecture that it is these bubbles that produce the sound output from the driven liquid. The mechanism of subharmonic production, however, is not clear, e.g., whether the bubble field will period double as a whole or only through a few members of the field. To settle

this question, a detailed, time-resolved visual study of the bubble dynamics is necessary. To this end, an arrangement for high-speed holographic cinematography has been developed [Hentschel & Lauterborn, 1985; Lauterborn & Koch, 1987]. The arrangement is depicted in Fig. 7. A series of coherent light pulses from a cavity-dumped argon ion laser is periodically deflected by an acoustooptic deflector into three distinct directions. Each of the beams is split into object and reference beams to produce a separate hologram on a rotating holographic plate. The sound output can
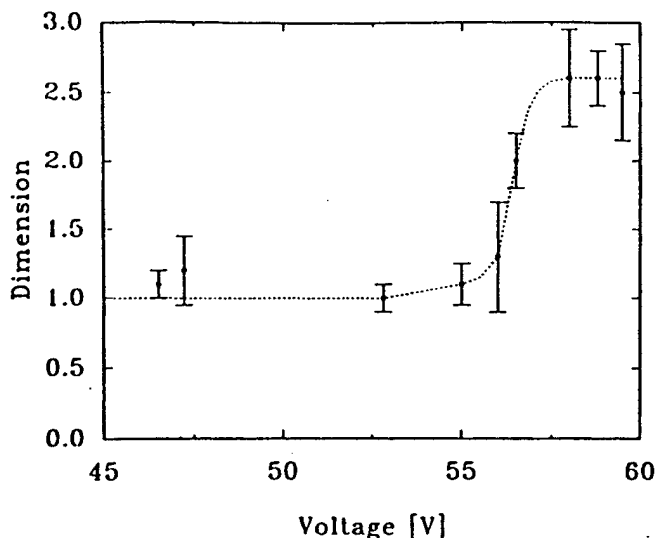
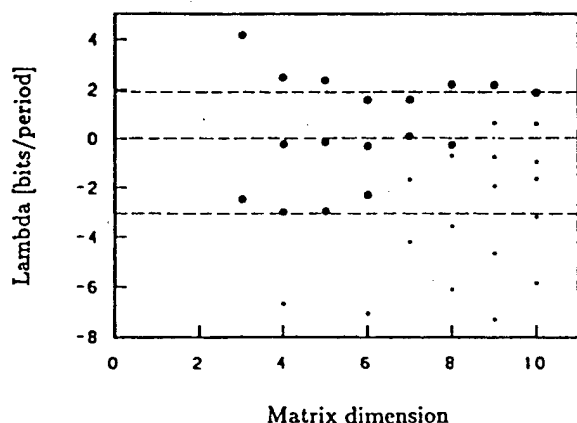Fig. 5. Fractal dimension estimation versus voltage of the driving.

Fig. 6. Lyapunov spectrum versus matrix dimension in the case of a chaotic attractor.

be simultaneously monitored in the same way as given in Fig. 2. The arrangement is capable of taking a hologram every cycle at a fixed phase of the driving or even three holograms per cycle [Lauterborn & Koch, 1987].

Experiments with holographic recording of the bubbles were conducted at a driving frequency of 23.1 kHz, i.e., holograms were taken at a rate of 23 100 holograms per second. Should the bubble field repeat its motion every cycle (period 1), then the same picture (three-dimensional) should reappear with each hologram. When the bubble field undergoes its first period doubling (to period 2), then every second picture (hologram) should look alike, and so on. When there is a

chaotic motion, then the appearance of the bubble field will alter unpredictably from cycle to cycle. This stroboscopic look at the dynamics is very reminiscent of Poincaré sections. The result of this experiment came as a surprise. The whole bubble field seemed to oscillate unisono (as seen in the pictures with, unfortunately, coarse resolution because of the small hologram size of 2×2 mm^2 and therefore large speckle grains). Figure 8 gives an example of the state just after the first period-doubling bifurcation. The row of frames shows reconstructed images from four successive holograms; the lower diagram is the corresponding power spectrum taken simultaneously. Figure 9 is taken from the area of second period doubling, and Fig. 10 gives a long series of seemingly chaotic oscillations. Figure 11 shows the corresponding power spectrum. Of course, as the experiments are difficult to conduct, time-consuming in processing (68 individual reconstructions had to be done just for the frames of Fig. 10), and only a few hundred or at most a few thousand holograms can be obtained in succession, it cannot be made absolutely sure that there are no long-lived transients, long periods or quasiperiodic motions involved. Also, problems in keeping the bubble field stable for some time prohibit the high accuracy necessary to answer these questions better. But the occurrence of a period-doubling sequence gives good confidence that chaotic oscillations indeed occur.

3.4. Theoretical considerations

A theory that can account for the dynamics of a bubble field (e.g., as given in Fig. 10) and its sound radiation has yet to be developed. Most advanced is the knowledge on the behavior of single spherical bubbles oscillating spherically in a sound field [Lauterborn & Parlitz, 1988; Smereka et al., 1987; Parlitz et al., 1990]. Also, it was, until recently, thought that the pressure field obtained from the multipole part of an asymmetric bubble oscillation would decay much faster with radial distance from the bubble than the direct monopole contribution and thus add negligibly to the overall sound radiation. It seems that this view has to be abandoned. Two recent articles claim monopole-like emission of sound by asymmetric bubble oscillations [Longuet-Higgins, 1989a,b]. This then adds a new interesting idea to the sound emission problem that, however, has yet to be worked out for the present purpose. Also, a connection to the Faraday experiment, although on a curved surface, can be envisaged with interacting modes yielding quasiperiodic or chaotic motion on the surface of the bubbles and giving rise to corresponding pressure waves in the liquid. In a next step, many of these bubbles would
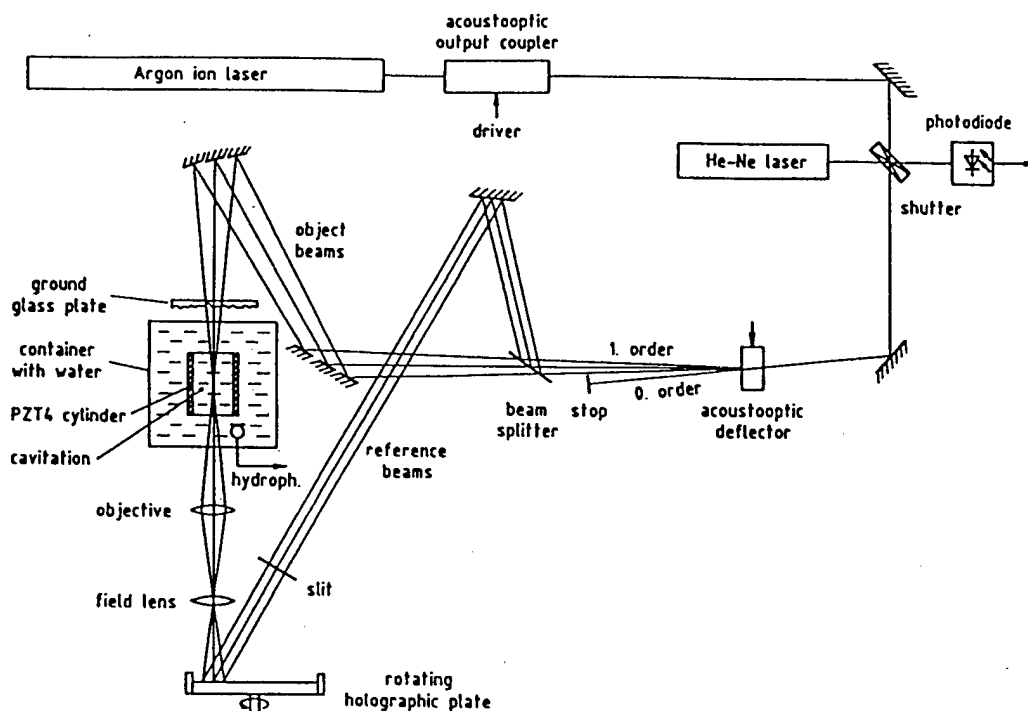
Fig. 7.   Experimental arrangement for high-speed holographic cinematography of the bubble motion in sound-irradiated water.
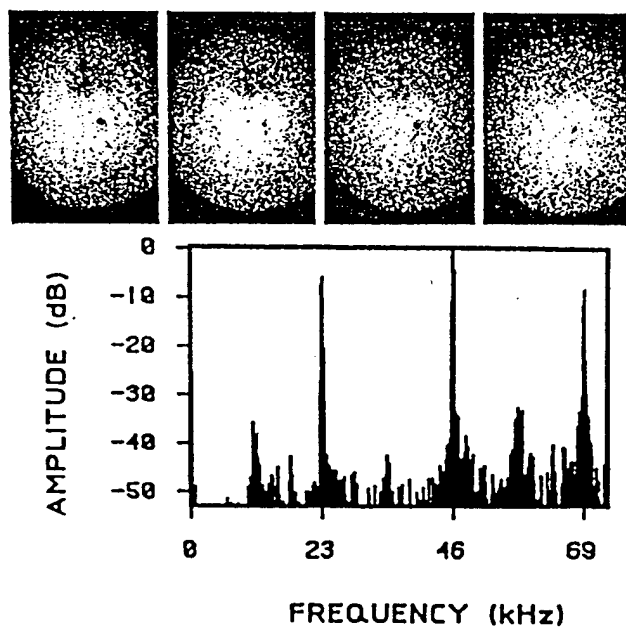
Fig. 8.   Four reconstructed images of the bubble distribution in sound-irradiated water from a holographic series taken at 23 100 holograms per second and the corresponding power spectrum of the sound emitted. First period doubling. (Courtesy of A. Koch).

Fig. 9.   Eight reconstructed images of the bubble distribution in sound-irradiated water from a holographic series taken at 23 100 holograms per second and the corresponding power spectrum of the sound emitted. Second period doubling. (Courtesy of A. Koch).

Fig. 10. Sixty-seven reconstructed images of a bubble distribution in sound-irradiated water from a holographic series taken at 23 100 holograms per second. (Courtesy of A. Koch).

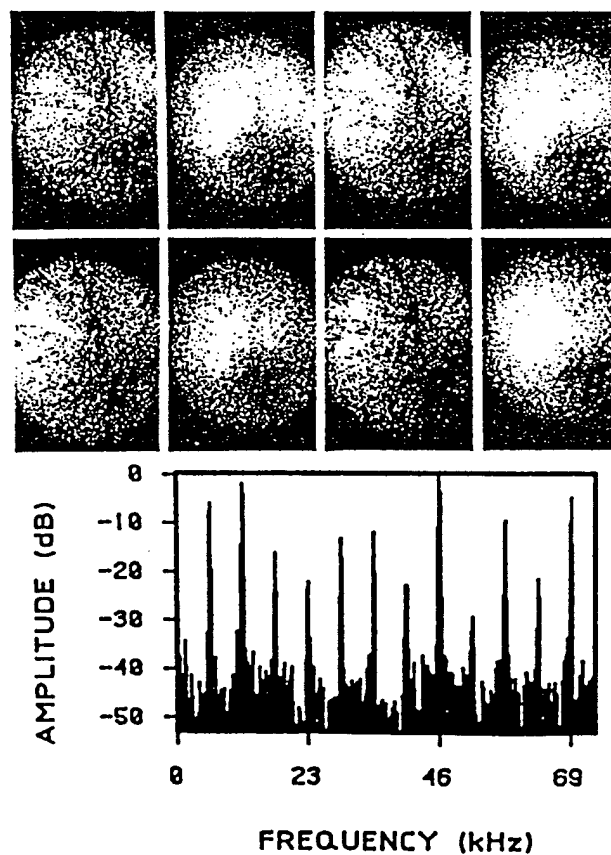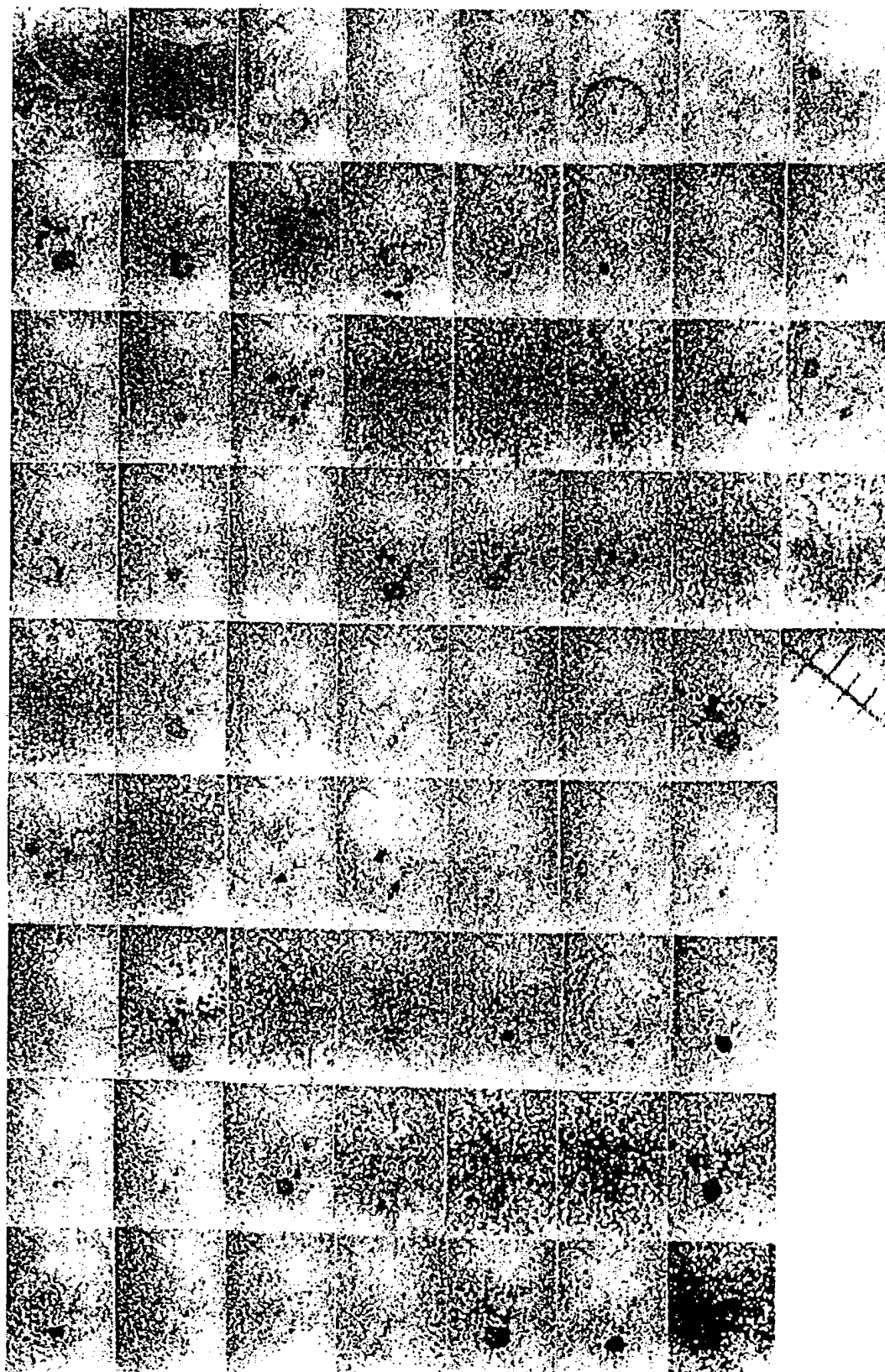have to be put together as a cloud. This is far from being feasible at present. But first attempts with bubble clouds along other lines are encouraging in giving the approximate behavior of a single bubble under suitable conditions [Smereka & Banerjee, 1988].

Thus, single-bubble behavior, being of interest on its own as a nonlinear oscillator, may have some bearing on the problem of sound emission from a bubble cloud. We have investigated several models, all of which show period doubling and chaos [Lauterborn, 1976; Lauterborn & Suchla, 1984; Lauterborn & Parlitz, 1988; Parlitz *et al.*, 1990]. In a similar way, as the experimental data are plotted as a spectral bifurcation diagram (Fig. 3), the theoretical data may also be transferred to spectra of the bubble wall motion and plotted, with the driving sound pressure amplitude as control parameter, in the form of a color-coded spectral bifurcation diagram. This is done in Fig. 12, where the Gilmore model has been used [Gilmore, 1952]. It is seen that areas of period doubling up to 1/8 the driving frequency are easily discernible, as well as chaotic regions with periodic windows (a usual behavior with nonlinear driven oscillators). The mechanism at work in producing the sound output therefore seems to be the widespread period-doubling mechanism and saddle-node bifurcations characteristic of the dynamics of many nonlinear systems.

4. Discussion and Conclusions

The new area of acoustic chaos physics has been reviewed with special emphasis on ultrasonic acoustic chaos. The experiment of irradiating a liquid with sound of a single frequency yields (under favorable circumstances) a period-doubling route to chaos (Fig. 3). The bubbles (or cavities) appearing in the liquid through a breakdown process (cavitation) are responsible for this. They are passive nonlinear oscillators which are periodically driven by the sound field that has also produced them.

The motions of the bubbles are very complex, ranging from pure radial oscillations at low amplitudes to motion with such strong collapses that shock waves are radiated (see Fig. 10). Also, strong deformations of the bubbles occur, leading to disintegration of a single bubble to many new ones [Hentschel & Lauterborn, 1985] and often erratic dancing motion. Strong clustering of bubbles and, often, an arrangement in a branch-like structure called streamers is observed. In experiments in which just a few bubbles were oscillating far away from each other, relatively stable operation of an almost pure shape oscillation at half the driving frequency could be maintained. The corre-
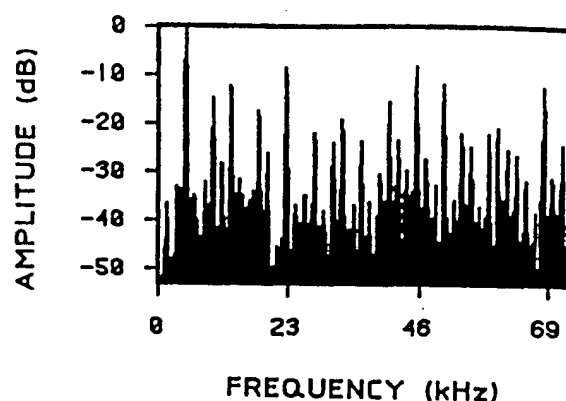
Fig. 11. Power spectrum of the sound output from a sound-irradiated liquid in the chaotic region. (Courtesy of A. Koch).

sponding power spectrum of the sound output from the liquid showed a substantial subharmonic line at one half the driving frequency. This phenomenon exists, besides the period-doubled volume pulsations demonstrated in Fig. 8. Even second period doubling seems to be possible with just shape oscillations of the bubbles. Holograms of better resolution holograms are being prepared to visualize the observations more clearly.

Bubbles are easily forced into surface oscillations (see, e.g., Kornfeld & Suvorov [1944], Strube [1971] and Hullin [1977]). The new finding that these surface oscillations emit strong sound [Longuet-Higgins, 1989a, b], gives rise to new considerations. Surface oscillations driven by the radial mode [Strube, 1971] may then play a bigger role than previously thought, and a more complex view of the sound emission from a sound-irradiated liquid than the simple view of the dominance of radial-mode oscillations of bubbles may emerge.

Acknowledgements

16

Fig. 12.   Spectral bifurcation diagram of numerically calculated spherical bubble oscillations with the driving pressure amplitude as control parameter. Period doubling, chaos and windows of periodic motion are seen. (Courtesy of E. Suchla).
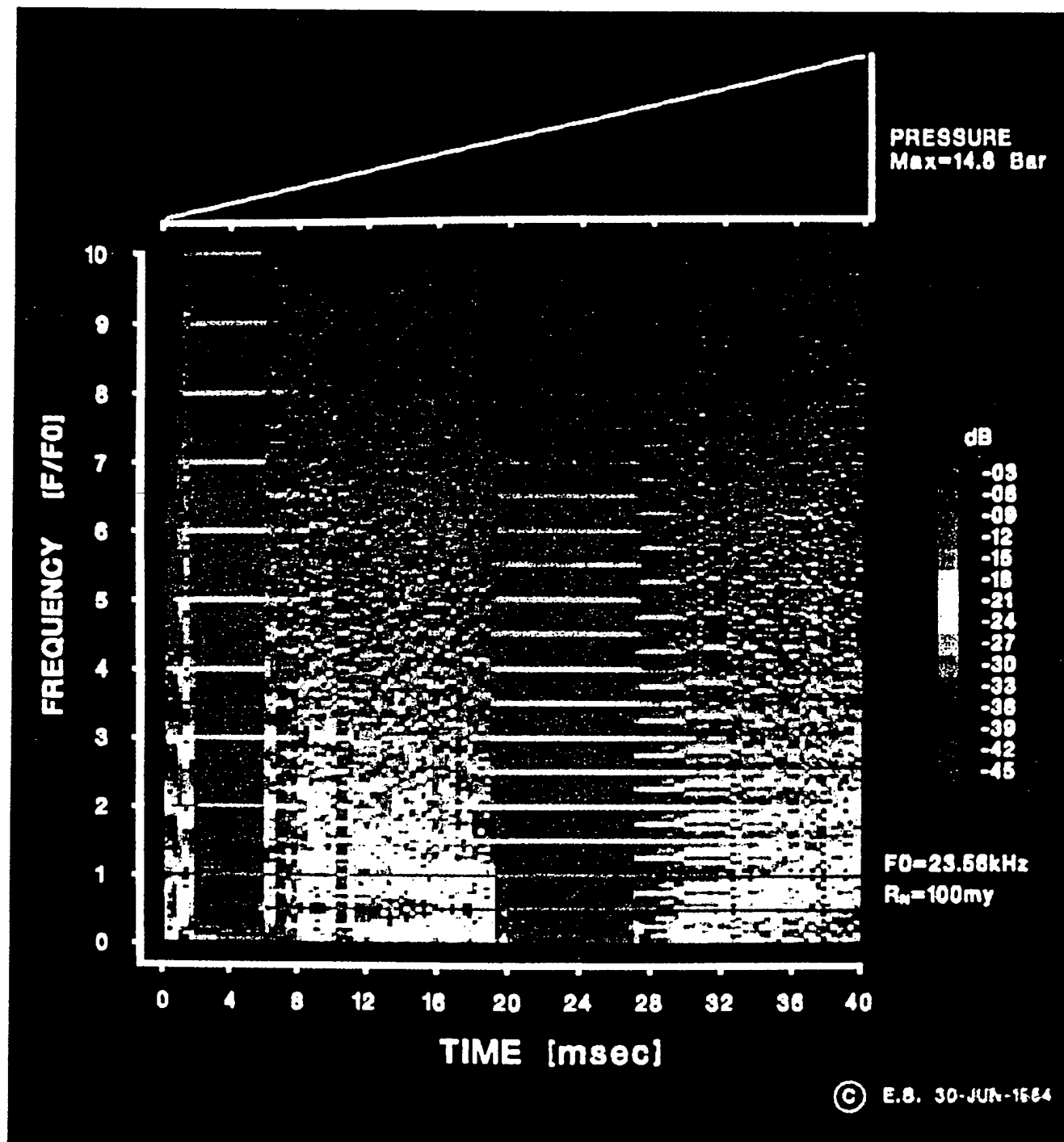
References

Bialek, W. & Wit, H. P. [1984] "Quantum limits to oscillator stability: Theory and experiments on acoustic emissions from the human ear", *Phys. Lett.* 104A, 173–178.

Ciliberto, S. & Gollub, J. P. [1985] "Chaotic mode competition in parametrically forced surface waves", *J. Fluid Mech.* 158, 381–398.

Eckmann, J. P. & Ruelle, D. [1985] "Ergodic theory of chaos and strange attractors", *Rev. Mod. Phys.* 57, 617–656.

Eller, A. & Flynn, H. G. [1969] "Generation of subharmonics of order one-half by bubbles in a sound field", *J. Acoust. Soc. Am.* 46, 722–727.

Esche, R. [1952] "Untersuchung der Schwingungskavitation in Flüssigkeiten", *Acustica* 2, Akust. Beih. 208–218.

Faraday, M. [1831] "On a peculiar class of acoustical figures, and on certain forms assumed by groups of particles upon vibrating elastic surfaces", *Phil. Trans. Roy. Soc. London* 121, 299–340. See appendix: "On the forms and states assumed by fluids in contact with vibrating elastic surfaces", pp. 319–340.

Feldmann, K. T. [1968a] "Review of the literature on Sondhauss thermoacoustic phenomena", *J. Sound Vib.* 7, 71–82.

Feldmann, K. T. [1968b] "Review of the literature on Rijke thermoacoustic phenomena", *J. Sound Vib.* 7, 83–89.

Gibiat, V. [1988] "Phase space representations of acoustical musical signals", *J. Sound Vibr.* 123, 529–536.

Gilmore, F. R. [1952] "Collapse of a spherical bubble in a viscous compressible liquid", California Institute of Technology Report No. 26-4, 1–40.

Gold, T. [1948] "Hearing II. The physical basis of the action of the cochlea", *Proc. Roy. Soc. Edinb.* B135, 492–498.

Güth, W. [1956] "Nichtlineare Schwingungen von Luftblasen in Wasser", *Acustica* 6, 532–538.

Gu, X. M. & Sethna, R. P. [1987] "Resonant surface waves and chaotic phenomena", *J. Fluid Mech.* 183, 543–565.

Hentschel, W. & Lauterborn, W. [1985] "High speed holographic movie camera", *Opt. Engng.* 24, 687–691.

Holzfuss, J. & Lauterborn, W. [1989] "Liapunov exponents from time series of acoustic chaos", *Phys. Rev.* A39, 2146–2152.

Holzfuss, J. & Mayer-Kress, G. [1986] "An approach to error-estimation in the application of dimension algorithms", in *Dimensions and Entropies in Chaotic Systems*, ed. Mayer-Kress, G. (Springer, Berlin) pp. 114–122.

Hullin, C. [1977] "Stabilitätsgrenze pulsierender Luftblasen in Wasser", *Acustica* 37, 64–73.

Kemp, D. T. [1978] "Stimulated acoustic emissions from the human auditory system, *J. Acoust. Soc. Am.* 64, 1386–1391.

Kemp, D. T. [1979] "Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea", *Arch. Otorhinolaryngol.* 244, 37–45.

Keolian, R., Turkevich, L. A., Putterman, S. J., Rudnick, I. & Rudnick, J. A. [1981] "Subharmonic sequences in the Faraday experiment: Departures from period doubling", *Phys. Rev. Lett.* 47, 1133–1136.

Kornfeld, M. & Suvorov, L. [1944] "On the destructive action of cavitation", *J. Appl. Phys.* 15, 495–506.

Lauterborn, W. [1969/70] "Subharmonische Schwingungen von Gasblasen in Wasser", *Acustica* 22, 238–239.

Lauterborn, W. [1976] "Numerical investigation of nonlinear oscillations of gas bubbles in liquids", *J. Acoust. Soc. Am.* 59, 283–293.

Lauterborn, W. [1986] "Acoustic turbulence", in *Frontiers in Physical Acoustics*, ed. Sette, D. (North-Holland, Amsterdam) pp. 123–144.

Lauterborn, W. [1989] "Acoustic chaos", in *Encyclopedia of Physical Science and Technology, 1989 Yearbook* (Academic Press, San Diego) pp. 171–176.

Lauterborn, W. & Cramer, E. [1981] "Subharmonic route to chaos observed in acoustics", *Phys. Rev. Lett.* 47, 1445–1448.

Lauterborn, W. & Holzfuss, J. [1986] "Evidence of a low-dimensional strange attractor in acoustic turbulence", *Phys. Lett.* 115A, 369–372.

Lauterborn, W. & Koch, A. [1987] "Holographic observation of period-doubled and chaotic bubble oscillations in acoustic cavitation", *Phys. Rev.* A35, 1974–1976.

Lauterborn, W. & Parlitz, U. [1988] "Methods of chaos physics and their application to acoustics", *J. Acoust. Soc. Am.* 84, 1975–1993.

Lauterborn, W. & Suchla, E. [1984] "Bifurcation superstructure in a model of acoustic turbulence", *Phys. Rev. Lett.* 53, 2304–2307.

Legge, K. A. & Fletcher, N. H. [1989] "Nonlinearity, chaos, and the sound of shallow gongs", *J. Acoust. Soc. Am.* 86, 2439–2443.

Longuet-Higgins, M. S. [1989a] "Monopole emission of sound by asymmetric bubble oscillations. Part 1. Normal modes", *J. Fluid Mech.* 201, 525–541.

Longuet-Higgins, M. S. [1989b] "Monopole emission of sound by asymmetric bubble oscillations. Part 2. An initial-value problem", *J. Fluid Mech.* 201, 543–565.

Maganza, C., Caussée, R. & Laloë, F. [1986] "Bifurcations, period doublings and chaos in clarinet-like systems", *Europhys. Lett.* 1, 295–302.

McIntyre, M. E., Schumacher, R. T. & Woodhouse, J. [1983] "On the oscillation of musical instruments", *J. Acoust. Soc. Am.* 74, 1325–1345.

Meron, E. & Procaccia, I. [1986] "Theory of chaos in surface waves equals the reduction from hydrodynamics to few-dimensional dynamics", *Phys. Rev. Lett.* 56, 1323–1326.

Miles, J. [1984] "Nonlinear Faraday resonance", *J. Fluid Mech.* 146, 285–302.

Moon, F. C. [1987] *Chaotic Vibrations* (Wiley, New York).

Parker, T. S. & Chua, L. O. [1989] *Practical Numerical Algorithms for Chaotic Systems* (Springer, New York).

Parlitz, U., Englisch, V., Scheffczyk, C. & Lauterborn, W. [1990] "Bifurcation structure of bubble oscillators", *J. Acoust. Soc. Am.* 88, 1061–1077.

Parlitz, U. & Lauterborn, W. [1987] "Period-doubling cascades and devil's staircases of the driven van der Pol oscillator", *Phys. Rev.* A36, 1428–1434.

Pedersen, P. O. [1935] "Subharmonics in forced oscillations in dissipative systems. Part I", *J. Acoust. Soc. Am.* 6, 227–238; Part II, ibid. 7, 64–70 (1935).

Rabinowitz, W. M. & Widin, G. P. [1984] "Interaction of spontaneous otoacoustic emissions and external sounds", *J. Acoust. Soc. Am.* 76, 1713–1720.

Rayleigh [1883a] "On the crispations of fluid resting upon a vibrating support, *Phil. Mag., Ser. 5*, 16, 50–58.

Rayleigh [1883b] "On maintained vibrations", *Phil. Mag., Ser. 5*, **15**, 229–235.

Rayleigh [1887] "On the maintenance of vibrations by forces of double frequency, and on the propagation of waves through a medium endowed with a periodic structure", *Phil. Mag., Ser. 5*, **24**, 145–159.

Riess, P. [1859] "Das Anblasen offener Rohre durch eine Flamme", *Ann. Phys. Chem.* **108**, 653–656.

Rijke, P. L. [1859] "Notiz über eine neue Art, die in einer an beiden Enden offenen Röhre enthaltene Luft in Schwingungen zu versetzen", *Ann. Phys. Chem.* **107**, 339–343.

Sano, M. & Sawada, Y. [1985] "Measurement of the Lyapunov spectrum from a chaotic time series", *Phys. Rev. Lett.* **55**, 1082–1085.

Smereka, P. & Banerjee, S. [1988] "The dynamics of periodically driven bubble clouds", *Phys. Fluids* **31**, 3519–3531.

Smereka, P., Birnir, B. & Banerjee, S. [1987] "Regular and chaotic bubble oscillations in periodically driven pressure fields", *Phys. Fluids* **30**, 3342–3350.

Sigrist, M. W. [1986] "Laser generation of acoustic waves in liquids and gases", *J. Appl. Phys.* **60**, R83–R121.

Smith, C. W., Tejwani, M. J. & Farris, D. A. [1982] "Bifurcation universality for first-sound subharmonic generation in superfluid helium-4", *Phys. Rev. Lett.* **48**, 492–494.

Sondhauss, C. [1850] "Über die Schallschwingungen der Luft in erhitzten Glasröhren und in gedeckten Pfeifen von ungleicher Weite", *Ann. Phys. Chem.* **79**, 1–34.

Strube, H. W. [1971] "Numerische Untersuchungen zur Stabilität nichtsphärisch schwingender Blasen", *Acustica* **25**, 299–313.

Strube, H. W. [1989] "Evoked otoacoustic emissions as cochlear Bragg reflections", *Hear. Res.* **38**, 35–46.

Taconis, K. W., Beenakker, J. J. M., Nier, A. O. C. & Aldrich, L. T. [1949] "Measurements concerning the vapour liquid equilibrium of solutions of He^3 in He^4 below $2.19°K$", *Physica* **15**, 733–739.

Takens, F. [1981] "Detecting strange attractors in turbulence", in *Dynamical Systems and Turbulence*, eds. Rand, D. A. & Young, L.-S. (Springer-Verlag, Berlin) pp. 366–381.

Thompson, J. M. T. & Stewart, H. B. [1986] *Nonlinear Dynamics and Chaos* (Wiley, Chichester).

Tufillaro, N. B. [1989] "Nonlinear and chaotic string vibrations", *Am. J. Phys.* **57**, 408–414.

Wheatley, J. & Cox, A. [1985] "Natural engines", *Phys. Today* **38**, No. 8, 50–58.

Yazaki, T., Sugioka, S., Mizutami, F. & Mamada, H. [1990] "Nonlinear dynamics of a forced thermoacoustic oscillation", *Phys. Rev. Lett.* **64**, 2515–2518.

Yazaki, T., Takashima, S. & Mizutani, F. [1987] "Complex quasiperiodic and chaotic states observed in thermally induced oscillations of gas columns", *Phys. Rev. Lett.* **58**, 1108–1111.

Methods of chaos physics and their application to acoustics

W. Lauterborn and U. Parlitz

Drittes Physikalisches Institut, Universität Göttingen, Bürgerstrasse 42-44, D-3400 Göttingen, Federal Republic of Germany

This article gives an introduction to the research area of chaos physics. The new language and the basic tools are presented and illustrated by examples from acoustics: a bubble in water driven by a sound field and other nonlinear oscillators. The notions of strange attractors and their basins, bifurcations and bifurcation diagrams, Poincaré maps, phase diagrams, fractal dimensions, scaling spectra, reconstruction of attractors from time series, winding numbers, as well as Lyapunov exponents, spectra, and diagrams are addressed.

PACS numbers: 43.10.Ln, 05.45. + b, 43.25.Yw, 43.50.Yw

INTRODUCTION

The last 10 years have seen a remarkable development in physics that may be succinctly described as the upsurge of "chaos."[1-16] This is, at first sight, really puzzling, as the notion of chaos implies irregularity and unpredictability, whereas physics is usually thought to be a science devoted to finding the laws of nature, i.e., its order and harmony. How, then, may chaos have become a subject of serious investigation in physics—and not only physics? This is just the new insight—that law and chaos do not exclude each other, that even simple deterministic laws may describe chaotic, i.e., unpredictable and irregular, motion. Thus not only law and order, but also law and chaos, go together and, even more so, it seems that law and chaos are as important a combination as law and order. This statement may be derived from the fact that chaotic motion is intimately related to nonlinearity and the realm of nonlinearity by far exceeds that of linearity.

This article is an attempt to acquaint the reader with the ideas and methods that lead to the above statements. The basic notions are given without resorting to too much mathematics. It is hoped that this approach will also be honored by those readers to whom this is not the first exposure to the subject.

I. ATTRACTORS

Theoretical chaos physics starts with evolution equations that describe the dynamic development of the state of a system (a model). These may be continuous models

$$\dot{x} = f_\mu(x), \quad x \in \mathbf{R}^m, \quad m \geqslant 1, \tag{1}$$

or discrete ones

$$x_{n+1} = g_\mu(x_n), \quad x_n \in \mathbf{R}^m, \quad m \geqslant 1, \quad n = 0,1,\dots. \tag{2}$$

The state of the system is given by the m-dependent variables $x(t) = [x_1(t), x_2(t), \dots, x_m(t)]$ or $x_n = (x_1^{(n)}, x_2^{(n)}, \dots, x_m^{(n)})$, respectively. The index μ indicates that the system depends on a parameter μ (often it will be several parameters). The dynamic laws (1) or (2) determine how a given state $x(t)$ or x_n develops. This evolution can be viewed when the states of the system are displayed as points in a *state space* $\mathbf{R}^m$. In the continuous case, the temporal (= dynamic) evolution then leads to a curve in this space called *trajectory* or also orbit (Fig. 1). In the discrete case, a sequence of points is obtained, usually called an *orbit*. The state space in nonlinear dynamics as introduced above is a generalization of the usual phase space of Hamiltonian dynamics. When p and q are the generalized coordinates and momenta of a Hamiltonian system, then $x = (p, q) \in \mathbf{R}^m$ with m necessarily even. General nonlinear dynamical systems may have an odd-dimensional state space.

An important question is how a set of initial conditions (a volume of the state space $\mathbf{R}^m$) evolves as time proceeds. According to the theorem of Liouville, a volume stays constant in conservative systems whereas it shrinks in dissipative ones. Here, only dissipative systems will be treated. In this case the question almost poses itself, as to how the volume shrinks and how the limit set of points in state space looks, to which a given volume shrinks. This simple question cannot yet be answered in general as obviously an unknown number of different limit sets are possible. The limit sets have been given the name *attractor* as trajectories out of whole volumina of state space move towards these sets, i.e., seem attracted by them. The set of initial conditions (points in state space) moving upon evolution towards a given attractor is called its *basin*.

What is already known about attractors and their properties? A certain classification can already be given. It often happens that all trajectories in state space move towards a
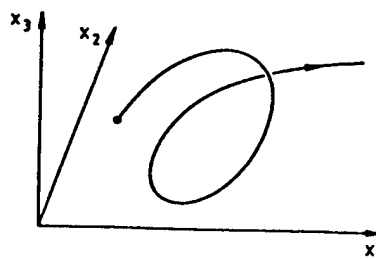

FIG. 1. A trajectory in state space.

20

single point, a *fixed point* [Fig. 2(a)]. This means that the system does not alter with time; it has come to rest. In the language of physicists, this is an equilibrium position. A standard example is a pendulum that has come to rest after some time of oscillation due to friction.

A more complex possibility is that the limit set consists of a closed trajectory that is scanned again and again. An attractor of this kind is called a *limit cycle* [Fig. 2(b)]. Limit cycles regularly occur with driven oscillators. The standard example is the attractor of the van der Pol oscillator. In physics, any sine wave (or square wave, etc.) generator displays an example of a limit cycle. The next kind of attractor fills an area (a two-dimensional surface) in a, e.g., three-dimensional, state space. This may happen if the system oscillates with two incommensurable frequencies. This attractor constitutes a *torus* [Fig. 2(c)]. A trajectory on the torus is a *quasi-periodic* motion. Systems with this property also exist experimentally (see, e.g., Ref. 12). These three types of attractors have been known for a long time.

Quite new is a further kind of attractor, called *strange* or *chaotic attractor* [Fig. 2(d)]. In the continuous case, an at least three-dimensional state space is necessary for a strange attractor to occur. The properties of strange attractors are not yet totally explored. An important property is the *dimension* of the strange attractor, which usually turns out to be *fractal*, i.e., not an integer. In Sec. VIII, we discuss how the dimension of general sets can be defined and determined in practical situations. A further property is that strange attractors obviously possess self-similar structures; i.e., on magnifying the attractor, partial structures repeat again and again on a finer and finer scale. The notion of *self-similarity* seems to play an important part in chaos physics as does the notion of fractal dimension. In Sec. X, we discuss how it may be brought about by the dynamic law by stretching and fold-
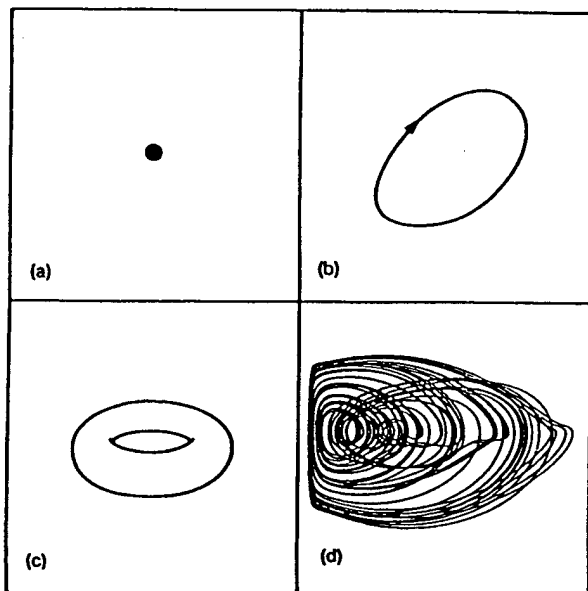
ing a volume of state space. Such objects obviously belong to the deeper inner structure of nature.[17] It may be interesting to note that the discovery of strange or chaotic attractors gradually came through theoretical arguing and that it is mainly through models with chaotic behavior that it has become possible to interpret measurements that were long known in the language of chaos physics. Acoustics has supplied a prominent, and one of the first, example in the form of acoustic cavitation noise[18-20] and related experiments.[21]

A dynamical system may possess several attractors simultaneously that are reached starting from different initial conditions in state space. The space of initial conditions is then divided into different areas, the basins of attraction, each of which belongs to its corresponding attractor. One speaks of *coexisting attractors*. Obviously, any type of attractor so far known can coexist with any other type including the same type. Thus a system may have several fixed points or several chaotic attractors and any mixture. An example for coexisting limit cycles is the resonance curve of a driven nonlinear oscillator where the maximum of a position coordinate of the limit cycle is plotted versus the frequency of a driver. At higher driving, it attains the appearance of a breaking wave (Fig. 3). Different oscillatory states are obtained depending on the way the curve is tracked. This phenomenon is well known as hysteresis. Examples for coexisting chaotic attractors are, for instance, found in Ref. 22 where the single-valley Duffing oscillator is explored.

Several questions concerning coexisting attractors can immediately be posed, such as, e.g., how many attractors a given system may have. This question is usually not easily answered. It may happen that a system possesses infinitely many coexisting attractors. For driven nonlinear oscillators (e.g., the bubble oscillator), the number of coexisting attractors grows rapidly when the damping is decreased.

The basins of attraction usually do not have a simple appearance. Even in the case of just two coexisting attractors, the boundaries of the two basins may be incredibly intertwined and even become a fractal set. An example of typical basins of attraction is taken from the Duffing equation $\ddot{x} + d\dot{x} - x + x^3 = f \cos \omega t$, which is a damped nonlinear oscillator with a two-valley potential driven by a harmonic

FIG. 2. Types of attractors: (a) fixed point, (b) limit cycle, (c) torus, (d) projection of a strange attractor.
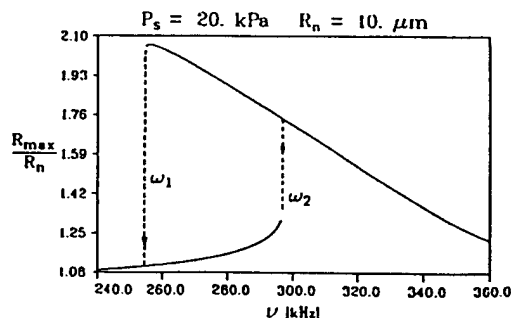
FIG. 3. A resonance curve of a bubble in water driven by a sound field. For the model used, see Eq. (9). Radius of the bubble at rest $R_n = 10\ \mu m$, sound-pressure amplitude 20 kPa (0.2 bar). In the region between ω_1 and ω_2, two coexisting attractors are present that are reached from different initial conditions.

1976   J. Acoust. Soc. Am., Vol. 84, No. 6, December 1988   W. Lauterborn and U. Parlitz: Chaos acoustics   1976

21

force of amplitude f and frequency ω. For a damping constant $d = 0.2$, a forcing amplitude $f = 1$, and a forcing frequency $\omega = 0.85$, this oscillator has three stable attractors whose basins of attraction are shown in Fig. 4 in black, grey, and white. The coordinates in the plane are $(x, v = \dot{x})$ and are the initial conditions with which the solution of the Duffing equation was started at $t = 0$. A set of 320 by 320 initial points has been used. Each point has been colored black or grey or left white according to the attractor to which the solution curve tends. The attractors are two period-2 stationary solutions and one period-1 stationary solution. The black and white areas are the basins belonging to the period-2 attractors and the grey area belongs to the period-1 attractor. The five big dots represent the three attractors. These points are given by stroboscopically illuminating the solution curve $[x(t), v(t)]$ at times $t_n = n\, 2\pi/\omega$. This leads to one point for the period-1 attractor and two points each for the two period-2 attractors. The black basin belongs to the period-2 attractor represented by the two white dots, the white basin to the period-2 attractor represented by the two black dots in the white area, and the grey basin to the period-1 attractor represented by the black dot in the grey area. The reader interested in the question of basin boundaries may consult Ref. 23 and from there explore the state of the art.

II. BIFURCATIONS

When doing experiments, it is found that the system investigated normally depends on several parameters. In a typical measurement, usually only one of the parameters (pressure, temperature, voltage, current, etc.) is altered to learn about the reaction of the system to the alteration. In theoretical language, one considers a one-parameter family of systems. The question then is how an attractor or coexisting attractors alter when a parameter is varied. In chaos physics, such a parameter is called a *control parameter*. Systems with different values of the control parameter are different systems and may have totally different attractors. Therefore, there must be parameter values at which the type
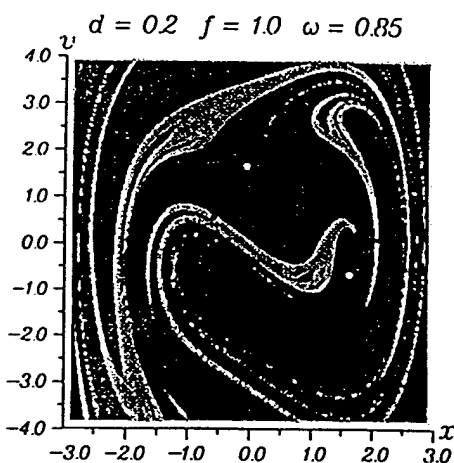
or gross appearance of an attractor switches to another one, or even just disappears, or is generated. This change, including birth and death, is called *bifurcation*. The set of parameter values at which a bifurcation occurs is called *bifurcation set*. It is thus a subset of *parameter space*, which, in a generalization of the above notions, may be high dimensional.

There are three basic types of local bifurcation, the *Hopf bifurcation*, the *saddle-node* or *tangent bifurcation*, and the *period-doubling* or *pitchfork bifurcation*. These bifurcations are called local bifurcations, as the phenomena associated with them can be studied by linearizing the system about a fixed point or periodic orbit in the immediate vicinity of the bifurcation point (of a control parameter). Figure 5 shows an example for each of the types of bifurcation. The standard example for a Hopf bifurcation is the onset of a self-excited oscillation in the van der Pol oscillator $\ddot{x} + \mu(x^2 - 1)\dot{x} + \omega^2 x = 0$ at $\mu = 0$ [Fig. 5(a)]. In this case, a fixed point changes to a limit cycle. Via Hopf bifurcation, a limit cycle may also change to a (two-dimensional) torus.

A saddle-node bifurcation occurs at the points of the resonance curves with the driving frequencies ω_1 and ω_2 in Fig. 3. At these points, one of the two attractors loses its stability and "jumps," in reality very slowly moves, towards the other attractor. Figure 5(b) shows this change in (projected) state space according to the jump at ω_2. A limit cycle of low amplitude changes to a limit cycle of larger amplitude. It is also possible that a limit cycle is replaced through a saddle-node bifurcation by a chaotic attractor. Also, via a saddle-node bifurcation, totally new oscillation frequencies may be introduced into a system (e.g., subharmonics). These "new" oscillation frequencies, e.g., of period 3, are due to coexistent attractors that take over at the bifurcation point.

The last type of bifurcation, the period-doubling bifurcation, only operates on periodic orbits. Its importance has become clear only in the last few years (see Refs. 3 and 4) and has stressed the importance of oscillatory systems for our understanding of nature. At a period-doubling bifurcation point, as the name states, a limit cycle of a given period T changes to a limit cycle of exactly double the period, $2T$. This appears peculiar, and even more peculiar is that this type of bifurcation preferentially occurs in the form of cascades; i.e., when a period doubling has occurred, it is very likely that, upon further altering the control parameter, a further period-doubling bifurcation occurs yielding $4T$, and so on. Indeed, via an infinite cascade of period doublings, a chaotic attractor can be obtained. This leads us to the more sophisti

$$d = 0.2 \quad f = 1.0 \quad \omega = 0.85$$

FIG. 4. Basins of attraction for the double-valley Duffing equation $\ddot{x} + d\dot{x} - x + x^3 = f\cos\omega t$ for $d = 0.2$, $f = 1$, $\omega = 0.85$. There are three attractors with their three basins. (Courtesy of V. Englisch.)
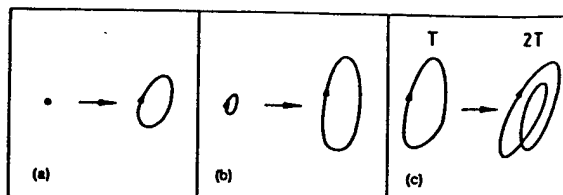

FIG. 5. Examples for the three types of local bifurcations: (a) Hopf bifurcation (fixed point → limit cycle), (b) saddle-node bifurcation (limit cycle → limit cycle), (c) period-doubling bifurcation (limit cycle of period T → limit cycle of period $2T$).

22

cated question of possible *sequences of bifurcations* when a
parameter of a system is changed. In the context of chaos
physics, such sequences are called *routes to chaos* and certain
scenarios are observed.

III. ROUTES TO CHAOS

It has been found that each of the local bifurcations may
give rise to a distinct route to chaos, and all three basic routes
have already been observed.[12] These routes are of impor-
tance because it is often difficult to conclude from just irreg-
ular measured data whether this is the outcome of intrinsic
chaotic dynamics of the system or simply noise in the mea-
suring system (outer disturbances). When, upon altering
the control parameter, one of the three basic routes is ob-
served, then this strongly supports the idea that the system is
a chaotic one producing the irregular output through its very
dynamics itself. In the context of measurements, an alterna-
tive way to distinguish between intrinsic and extrinsic noise
has been developed. This method is discussed in Sec. IX.

The scenario based on a sequence of Hopf bifurcations is
called *quasiperiodic route to chaos*, as a system with incom-
mensurate frequencies undergoes quasiperiodic oscillations
(Fig. 6). This route is connected with the names of Ruelle,
Takens, and Newhouse.[24,25] It is a somewhat surprising
route because, starting from a fixed point, the three-dimen-
sional torus generated after three Hopf bifurcations is not
stable in the sense that there exists an arbitrarily small per-
turbation of the system (alteration of parameters) for which
the three-torus gives way to a chaotic attractor. This route to
chaos has been found experimentally in the flow between
rotating cylinders (Taylor–Couette flow) and in Rayleigh–
Bénard convection where a liquid layer is heated from be-
low.[3,4]

The route to chaos mediated by saddle-node or tangent
bifurcations comes in different types, but all with the appear-
ance of a direct transition from regular to chaotic motion.
The most prominent type is called the *intermittency route to
chaos*. This route is connected with the names of Pomeau
and Manneville.[26,27] It only needs a single saddle-node bifur-
cation and is not easily visualized in its properties in a single
diagram (Fig. 7). It is in a sense really a route to chaos (and
not just a jump), as in the immediate vicinity after the bifur-
cation point μ_c, the trajectory contains long time intervals of
(almost) regular oscillation (so-called laminar phases) with
only short bursts into irregular motion. The period of the
oscillations equals approximately that before chaos has set
in. With increasing distance $|\mu - \mu_c|$ from the bifurcation
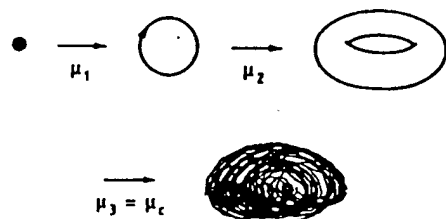point μ_c into the chaotic region, these laminar phases be-
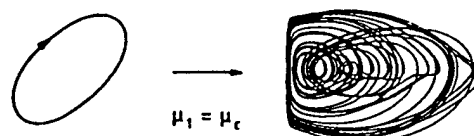

FIG. 7. Intermittency route to chaos via a saddle-node bifurcation.

come shorter and shorter, and the intervals of visibly chaotic
oscillations larger and larger, until the regular oscillation
intervals disappear. Chaos is really developed only at μ val-
ues at some distance from μ_c. This route has, for instance,
been observed in Rayleigh–Bénard experiments. Besides the
intermittency route, there is a different type of transition to
chaos connected with saddle-node bifurcations. It consists of
a direct transition from a regular attractor (fixed point, limit
cycle) to a coexisting chaotic one without the phenomenon
of intermittency described above. This type is usually en-
countered in systems with many coexisting attractors, as in
the case of bubbles in a liquid driven by a sound field.

The route to chaos encountered with the period-dou-
bling bifurcation is called the *period-doubling route* to chaos
(Fig. 8). This route is connected with many names, with
Sharkovskii, Grossmann, Thomae, Coullet, Tresser, and
Feigenbaum being the most prominent ones.[28-31] Aperiodi-
city is introduced here in steps, as every period doubling
transforms a limit cycle at first only into a limit cycle of
doubled period. But when the sequence of successive period
doublings consists of infinitely many doublings, the limit
will be a period of infinity, i.e., an aperiodic motion. This, of
course, can only happen at a finite value of the control pa-
rameter μ, when the intervals in μ between successive dou-
blings get smaller at a sufficiently rapid rate. This is indeed
the case. Period doubling is governed by a universal law that
holds in the vicinity of the bifurcation point to chaos μ_c.
Actually, there are several laws. One of these states that
when the ratio δ_n of successive intervals of μ, in each of
which there is a constant period of oscillation, is taken,

$$\delta_n = (\mu_n - \mu_{n-1})/(\mu_{n+1} - \mu_n), \tag{3}$$

where μ_n is the bifurcation point for the period from $2^n T$ to
$2^{n+1} T$, then in the limit $n \to \infty$ a universal constant is ob-
tained,[30,31] which for usual physical systems has the value

$$\lim_{n \to \infty} \delta_n = \delta = 4.6692 \cdots . \tag{4}$$

This number is called the *Feigenbaum number* δ, because

FIG. 6. Quasiperiodic route to chaos via a sequence of Hopf bifurcations.
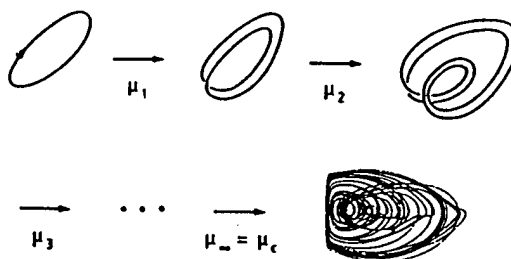
FIG. 8. Period-doubling route to chaos via an infinite cascade of period-
doubling bifurcations.

1978   J. Acoust. Soc. Am., Vol. 84, No. 6, December 1988   W. Lauterborn and U. Parlitz: Chaos acoustics   1978

23

Feigenbaum discovered its universality. The number had been found previously in the *inverse cascade* consisting of bands of *periodic chaos* converging towards the accumulation point μ_∞ from the opposite side.[29] It is not known whether this number can be expressed by other numbers like π or e, or is a totally new number of a similar kind. At present, it can only be determined numerically to some accuracy (like π and e). The period-doubling route to chaos has been found in many experiments by now, but among the first was a purely acoustical experiment,[18-20] the acoustic cavitation noise, which will be discussed in greater detail in a forthcoming article.

Period doubling has been found experimentally in significantly different systems[1-13] and in areas as different as physics (hydrodynamics, acoustics, optics), electronics, chemistry, biology, and physiology. A large class of physical systems that is considered of special importance is the driven nonlinear oscillators[9] (bubble oscillator,[32-36] Duffing oscillator,[22] van der Pol oscillator,[37] Toda oscillator[38]). They all show period-doubling and saddle-node bifurcations to chaos (the van der Pol oscillator, also Hopf bifurcations) and display common features connected with their resonance properties. The bubble oscillator will be described in more detail in a separate article.

IV. BIFURCATION DIAGRAMS

Several methods are available to handle experimental data in the attempt to determine which route to chaos may apply. They are not all similarly well suited to each route so that usually one should try all of them. In the intermittency route to chaos, the directly measured time dependence of the variable considered is taken to observe the continuous shortening of the laminar phases with change of the control parameter. In the quasiperiodic route, the time dependence of a variable usually has no specific features that would easily be detectable. In this case the Fourier spectrum should be calculated, which immediately detects the new frequencies appearing upon alteration of the control parameter. The period-doubling route to chaos, too, is best observed in the spectrum of the data, as the successive appearance of lines at half the lowest line (and their harmonics) is very characteristic. But in this case also a plot of the time dependence often yields good hints.

The methods recommended so far stem from usual data analysis. They are not quite satisfactory for coping with the problem of chaotic motion. Therefore, chaos research has invented and introduced its own specific methods to display its results.

Among these methods are the *bifurcation diagrams* that have become a powerful and standard tool in visualizing the properties of a system. In a bifurcation diagram, the attractors of a system are plotted versus the control parameters. This is the idea that, however, usually cannot be fully realized due to the dimension of the attractors and also of the parameter space. As for visualization purposes, normally only the plane of the paper is available; just one coordinate of the attractor (a projection) is plotted versus a single control parameter. This works for discrete systems (2). In the case

of continuous systems, a discretization (see Secs. V and VI below) is necessary.

The standard example of a bifurcation diagram is therefore given by one-dimensional iterated maps $x_{n+1} = f_\mu(x_n)$, $x_n \in [a,b]$; $x_n, a, b \in R$, depending on a single control parameter μ, $\mu \in R$, as, in this case, the attractors are at most one dimensional. This map shows up, for instance, in (strongly) damped oscillatory systems when Poincaré sections are taken (see Sec. V). Figure 9 shows a bifurcation diagram of the so-called logistic parabola in the form $x_{n+1} = 4\mu x_n(1-x_n)$. For small μ, the attractor is a fixed point. It splits into a periodic attractor of period 2, then period 4, etc., until at the accumulation point μ_∞ the period 2^∞, i.e., aperiodicity, is obtained. Afterwards, these periods are present in the form of bands that combine, in a reverse way to period doubling, to a single chaotic band.[29] In the chaotic region, parameter intervals with periodic attractors appear, e.g., of period 3. Most of the known properties of this system are collected in Refs. 11 and 39; see also Refs. 3 and 4.

A variant of this kind of bifurcation diagram has been introduced by us in the context of acoustic chaos[18] and called *spectral bifurcation diagram*.[33,20] Of course, also in this case only a single coordinate of an attractor can be handled, whose power spectrum is plotted versus the control parameter. Since a power spectrum itself needs two dimensions to be plotted, additional difficulties appear which may be overcome by using grey scales or, even better, color graphics (see Ref. 1, color plate VI or Ref. 20, plate I). Spectral bifurcation diagrams are especially well suited to experimental systems, where external noise usually is hard to avoid. This noise is distributed over a wide spectral range, whereas the energy in the system is concentrated in a few lines when, e.g., period doubling adds new lines. These then stick out from the noise background and are easily detected. In acoustics, spectral bifurcation diagrams are well known as "visible speech" when time is considered as a "control parameter."

V. POINCARÉ SECTIONS AND POINCARÉ MAPS

When trying to display bifurcation diagrams, as mentioned above, difficulties arise coming from the dimension of the space needed for this purpose. This problem had already been encountered by Poincaré in the context of coping with the problem of the stability of the solar system. He invented what today is called a *Poincaré section*, whereby one dimen-
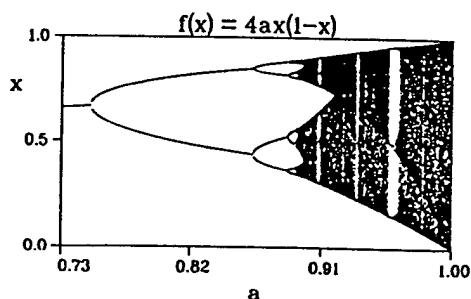

FIG. 9. The bifurcation diagram of the logistic parabola in the form $x_{n+1} = 4\mu x_n(1-x_n)$.

24

sion can be gained, and a continuous system of the kind in Eq. (1) is transferred to a discrete system of the kind in Eq. (2). When investigating high-dimensional systems, this is of little help, but when working with low-dimensional, especially three-dimensional, systems, the limit of visualization of the properties of a system is shifted for quite a large class, among them the driven one-dimensional oscillators. When having a three-dimensional state space, a Poincaré section is simply a plane S (a hyperplane in higher dimensional systems) in the state space that is intersected by all trajectories transversally (Fig. 10). Usually such a plane may only exist locally, but global Poincaré sections are also encountered in certain special systems, e.g., in our driven bubble oscillators. We therefore consider only global Poincaré sections here. Section planes are well suited to investigate the stability of periodic orbits. In chaos research they are now frequently used to display strange attractors, whereby only the section points of the trajectory of the attractor in the plane S are plotted. In Fig. 11, a strange bubble attractor, i.e., an attractor of a driven bubble in a liquid, as it appears in a Poincaré section, is given. The section points arrange themselves on lines folded over and over again. They hop around on this structure in a hardly describable manner. To show its fractal nature and self-similarity, as pointed out in Sec. I, a blowup is given in Fig. 11(b) that reveals the occurrence of the same structure on a finer and finer scale.

It is possible to consider the dynamics of a given continuous system in the section plane S only. When a point $Q_1 \in S$ is taken, it is imaged via the dynamics of the system to the point $Q_2 = P(Q_1) \in S$. In this way, a continuous dynamical system is transferred to a discrete dynamical system, given by a map from S to S. This map is called a *Poincaré map P* or also *first return map*. It is immediately seen that a periodic orbit of the continuous system becomes a fixed point of a corresponding (iterated) Poincaré map. One dimension has been saved in this way. A quasiperiodic orbit consisting of two incommensurate frequencies then looks like a limit cycle in the Poincaré section (a cut transverse through a two-dimensional torus). This "limit cycle," however, is made up of points hopping around, not by a smooth periodic trajectory in the section plane.
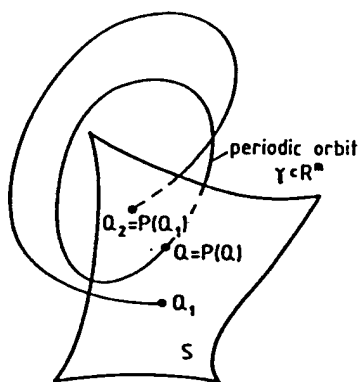

FIG. 10. Poincaré section plane S and the Poincaré map P. The section point Q of a periodic orbit γ is a fixed point of the Poincaré map P.
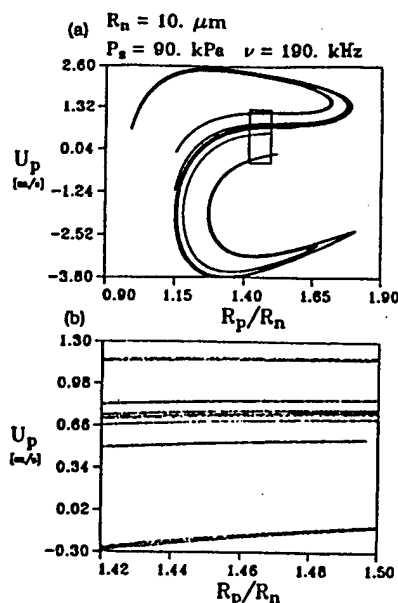

FIG. 11. A strange bubble attractor in a Poincaré section plane. (a) Total view, (b) exploded view that indicates the self-similarity of the band structure.

With the help of the Poincaré section method, the period-doubling route to chaos can adequately be displayed without resorting to spectral representations. Upon the first period-doubling bifurcation, the fixed point splits into a periodic orbit consisting of two points that are imaged back and forth. At each successive period doubling, each of the previous points splits into two new points until in the limit at $\mu = \mu_\infty = \mu_c$ the point set of an aperiodic attractor is obtained. For a bifurcation diagram, usually the points of a Poincaré section are used, whereby one again is forced to take only one coordinate of a point in the section for simple visualization.

The Poincaré map P defined on a surface S determines where the points of S are imaged under P. Thus P considers a whole set of initial conditions simultaneously. In physics, usually only single trajectories can be followed (e.g., when measuring a pressure in dependence on time). The trajectory may then be described in discrete form through the series $P(x_0), P\{P(x_0)\},...,$ x_0 being the initial condition. This leads to iterated maps [compare Eq. (2)]:

$$x_{n+1} = P(x_n), \quad n = 0,1,2,... \ . \tag{5}$$

The simplest map P with nontrivial dynamics that may be encountered will be a one-dimensional map with some function f:

$$x_{n+1} = f(x_n), \quad x_n \in R, \quad n = 0,1,... \ . \tag{6}$$

When a control parameter μ is introduced for comparison with experiments, one is led to a family of maps f_μ:

$$x_{n+1} = f_\mu(x_n), \quad x_n \in R, \quad \mu \in R, \quad n = 0,1,... \ . \tag{7}$$

Thus the Poincaré map connects continuous dynamical systems (differential equations) with iterated maps. Iterated maps have long been investigated by pure mathematicians who collected a wealth of beautiful results that now find

25

applications in physics. With respect to the corresponding differential equations, iterated maps are much simpler and can be investigated much more easily both analytically and numerically. Yet the same richness in behavior can be expected as in the original differential equation. How involved a behavior must be envisaged is strikingly demonstrated by the "simple" example of the quadratic map or logistic parabola

$$x_{n+1} = 4\mu x_n (1 - x_n), \quad x_n \in [0,1], \quad \mu \in [0,1], \quad (8)$$

the bifurcation diagram of which, in a sense, can completely be given and is plotted in Fig. 9.

One-dimensional maps like those of Eq. (7) can be constructed from the Poincaré map P only in those cases where the (strange) attractor in the section plane resembles almost a (thin) curve. Then either a projection in some suitable direction will do, or some coordinates along the curve must be introduced with respect to which a one-dimensional map may be formulated. Such maps are often called *reduced* (*Poincaré*) *maps* or, as suggested by us, *attractor maps*.[33] Figure 12(a) shows a Poincaré section plane where a strange attractor can be seen for a bubble with radius at rest of $R_n = 10\,\mu$m, driven at a sound-pressure amplitude of 276 kPa and a frequency of 530 kHz. It is noticed that the section points make up eight short line segments, and thus a one-dimensional map may be constructed for each of them. This is indeed possible, as the dynamics on the attractor are both regular *and* chaotic. The regularity consists in the fact that, from one section of the chaotic trajectory to the next, the section points go around from one segment to the next in the manner indicated until all eight segments have been visited. Then the sequence starts again. The chaos in the dynamics comes from the fact that on a segment the points come back

in an irregular way for which no long-term prediction can be made. For this type of behavior, known from the inverse cascades of the logistic parabola, the term *periodic chaos* has been coined.[29] Figure 12(a) thus shows a period-8 chaotic attractor. In this case, strong lines in the Fourier spectrum of the motion at $\frac{1}{8}$ the driving frequency and their harmonics appear (see Sec. VI). If only every eighth section point were plotted, only one line segment would show up. Such a plot is called a *subharmonic Poincaré section plot of order eight*.

From one line segment, which, to be sure, is only approximately a line, a reduced Poincaré map or attractor map may be constructed in the following way. Every eighth section point is taken and the points of this sequence are plotted versus the previous one. This means that x_{n+8} is plotted versus x_n, x_n being the original iteration points with $n = 1$, 9, 17,... . When one coordinate, in this case the radius of the bubble, is taken, a map like that shown in Fig. 12(b) is obtained for a certain segment. There will be eight different attractor maps of this kind, depending on the starting point, i.e., the segment chosen. The type of map given in Fig. 12(b) is called a *subharmonic attractor map of order eight*. It strongly resembles the logistic parabola (8). When a parameter in the original differential equation is altered, it may happen that the corresponding attractor in the Poincaré section plane alters in such a way that the corresponding (subharmonic) attractor map alters like a logistic parabola when producing period doubling. Then period doubling in the continuous system may be said to occur as in the logistic parabola. This follows from the universality of the scaling laws governing period doubling.[3]

Whenever the attractor is more complicated, two-dimensional maps must be constructed and investigated. Since the time that the connection between iterated maps and continuous dynamical systems (as used by physicists) has been clearly noticed, not only mathematicians but also physicists (and other scientists) work intensely on the properties of iterated maps, with much computer work going on. In any case, the occupation with iterated maps is strongly recommended for those who wish to get a deeper understanding of chaos. The newcomer may start with the article of May in *Nature*.[11]

VI. DEMONSTRATION OF SOME METHODS OF CHAOS PHYSICS CHOOSING A BUBBLE OSCILLATOR

In this section, we pause to demonstrate some of the methods coherently on a bubble oscillator. The bubble oscillator used is a nonautonomous differential equation of second order of the form[35]

$$\left(1 - \frac{\dot{R}}{c}\right)R\ddot{R} + \dot{R}^2\left(\frac{3}{2} - \frac{\dot{R}}{2c}\right) = \frac{P}{\rho} + \frac{1}{\rho c}\frac{d}{dt}(RP), \quad (9)$$

with

$$P(R,\dot{R},t) = P_i(R) - 2\sigma/R - 4\mu(\dot{R}/R) - P_{stat} + P_v - P_s \sin(2\pi\nu t)$$

and

$$P_i(R) = (P_{stat} - P_v + 2\sigma/R_n)(R_n/R)^{3\kappa},$$

where $R = R(t)$ is the radius of the bubble at time t, R_n is the
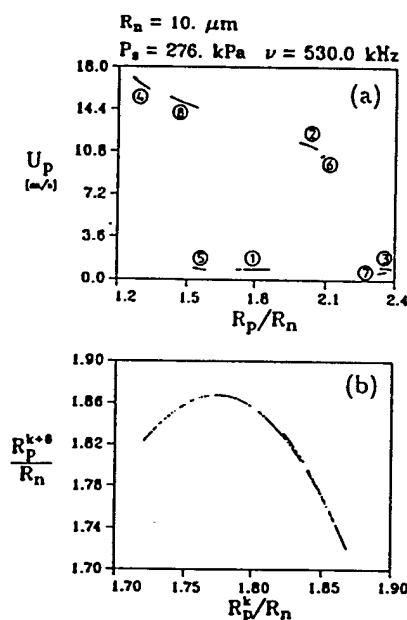
FIG. 12. Period-8 chaotic attractor of a bubble oscillator in a Poincaré section plane (a) and a subharmonic attractor map of order eight (b). The encircled numbers in (a) indicate the succession of the section points.

26
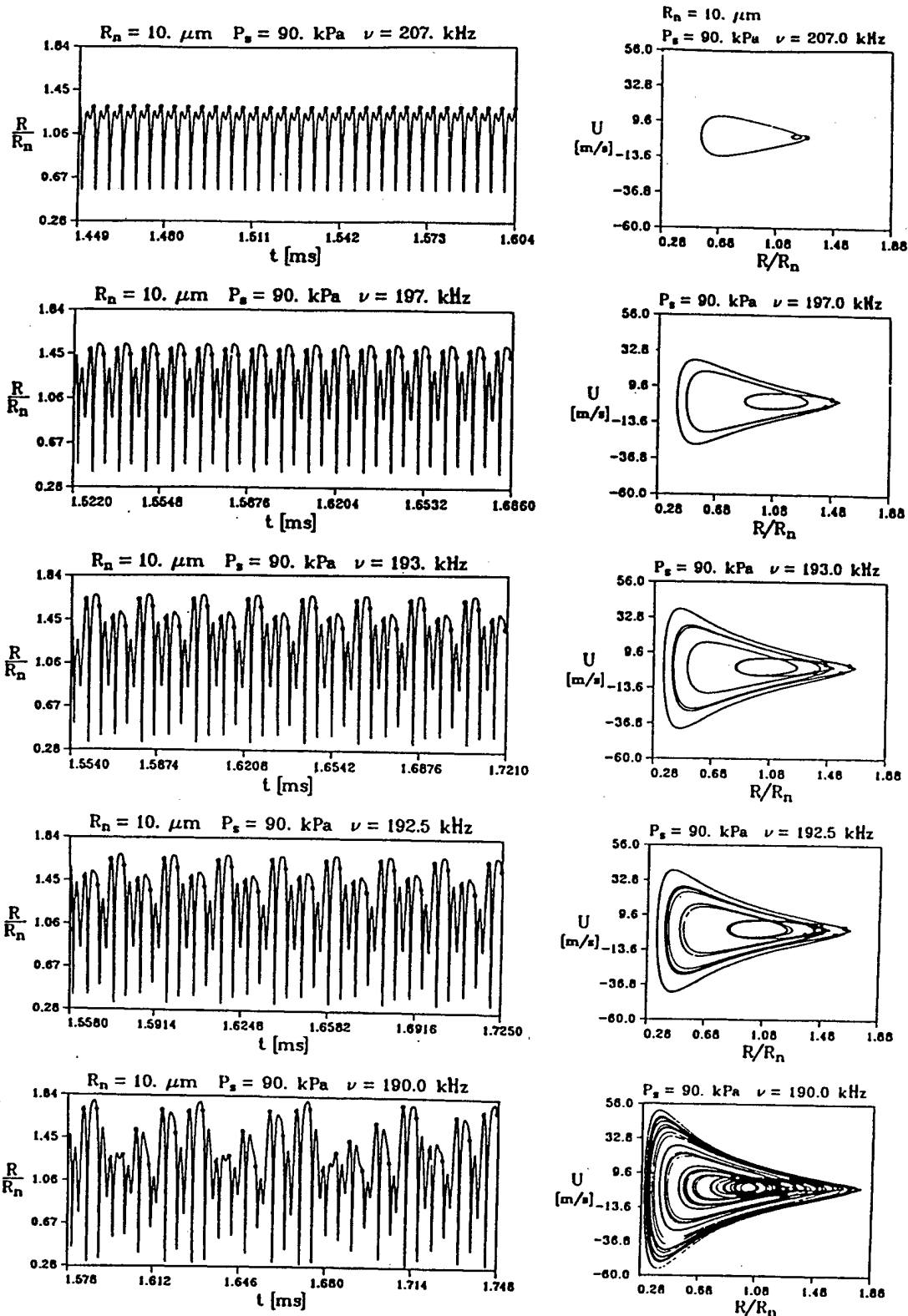
FIG. 13. Period-doubling route to chaos demonstrated by the attractors for a bubble oscillator. Left column: radius–time solution curves; middle left column: trajectories in state space; middle right column: Poincaré section plots; right column: power spectra. Radius of bubble at rest $R_n = 10\,\mu m$, sound-pressure amplitude 90 kPa (0.9 bar), driving frequency 1st row: 207 kHz; 2nd row: 197 kHz, 3rd row: 193 kHz, 4th row: 192.5 kHz, 5th row: 190 kHz.

27

FIG. 13. (*Continued.*)

W. Lauterborn and U. Parlitz: Chaos acoustics
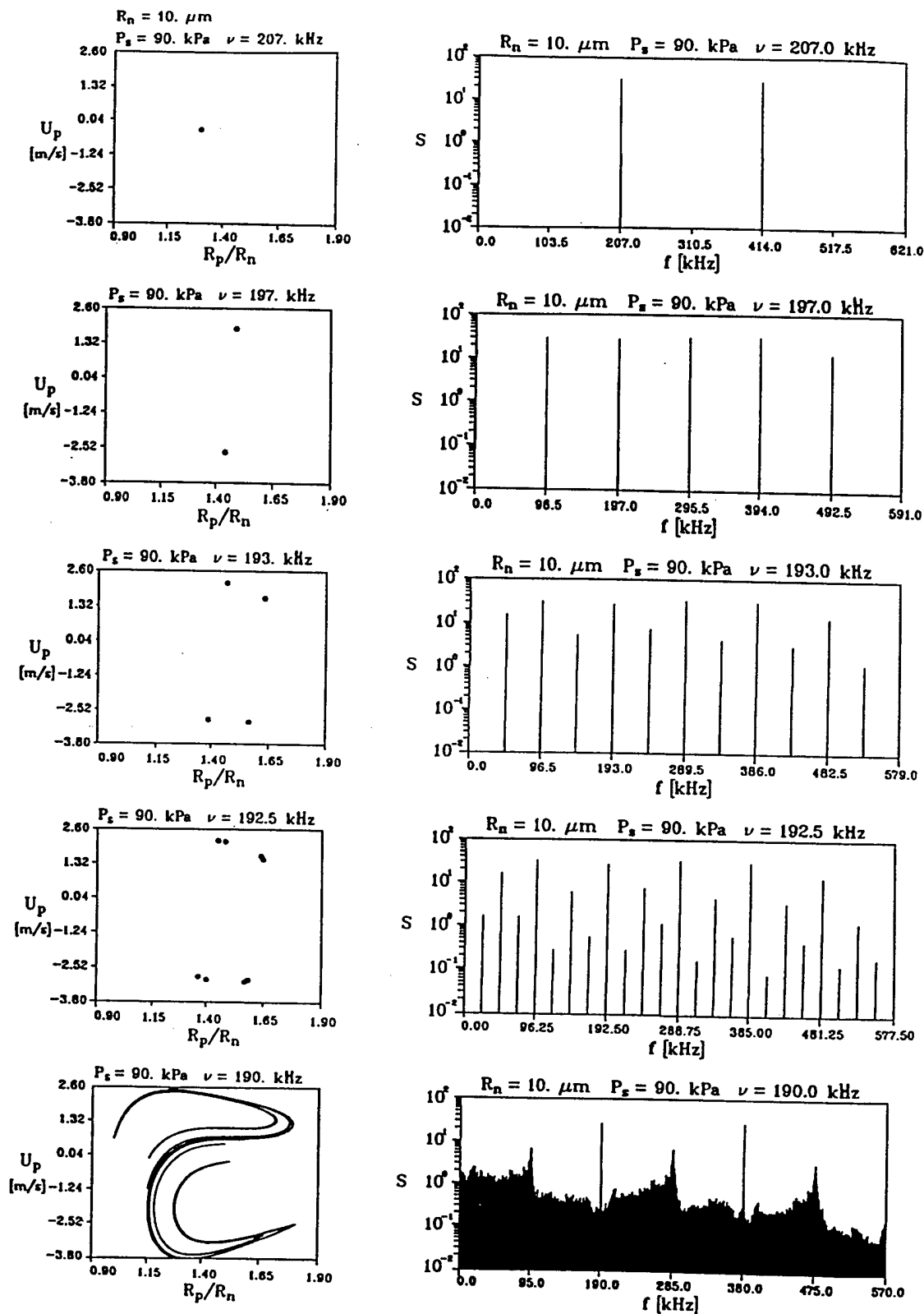
radius of the bubble at rest, ν is the frequency of the driving sound field, P_s is the amplitude of the driving sound field, $P_{stat} = 100$ kPa is the static pressure, $P_v = 2.33$ kPa is the vapor pressure, $\sigma = 0.0725$ N/m is the surface tension, $\rho = 998$ kg/m^3 is the density of the liquid, $\mu = 0.001$ N s/m^3 is the viscosity, $c = 1500$ m/s is the sound velocity, and $\kappa = 4/3$ is the polytropic exponent of the gas in the bubble. Here, $\dot{R}$ denotes differentiation of the radius with respect to time t. The model (9) describes a spherical gas bubble of radius $R(t)$ in water, set into motion by a sound field of sinusoidal time dependence constant at any one time all over the bubble surface. For this article, it may suffice to just give the bubble model without discussing its derivation or relation to other bubble models. A more thorough description will be given elsewhere, together with a detailed discussion of the chaotic properties of some of its solutions.

When the sound-pressure amplitude (control parameter) of the driving sound field is increased at constant frequency, or when the frequency (another control parameter) of the driving sound field is altered at constant sound-pressure amplitude, peculiar things may happen with the radial (!) oscillation of the bubble (Ref. 35; see, also, Refs. 32 and 34). Despite periodic excitation, chaotic oscillations are encountered for some parameter values. An example where this happens via a period-doubling route is given in Fig. 13, which displays 20 diagrams in four columns. A bubble of radius at rest $R_n = 10 \mu$m, driven at a sound-pressure amplitude of 90 kPa (0.9 bar), has been chosen. The frequency of the driving sound field is lowered from $\nu = 207$ kHz to $\nu = 190$ kHz. In five steps, the stationary solutions and their spectra (after transients have decayed) are plotted for $\nu = 207, 197, 193, 192.5,$ and 190 kHz, the same for all diagrams in a row. The dots in the diagrams of the radius–time curves (left column) correspond to a certain phase of the driving sound field. Their interval thus corresponds to the period T of the driving. The rows contain to the left the radius time curves, followed by phase space (velocity versus radius) curves (trajectories), Poincaré section plots, and power spectra of the radius–time curves. In the first row, the bubble oscillates in a stationary state with the period of the driving. In the language of chaos theory, we have a limit cycle as an attractor which has the period T of the driving. In the radius–time plot, the thick dots therefore all lie at the same level R/R_n. The limit cycle trajectory in the second column also is marked by just one dot, as it exactly repeats after one period. With periodically driven oscillators, a global Poincaré plane of section can be defined by a fixed phase of the driving. The plane then is made up of the radius R_P of the bubble (given here normalized with R_n as R_P/R_n) and its velocity U_P, where the index P stands for Poincaré to notify that it is a section plane, in contrast to the second column where the velocity all along the trajectory is plotted. Thus the first diagram in the third column simply contains a single point, the one section point of the limit cycle. The appearance of only one point in the section plane indicates that the limit cycle has the period T of the driving, as is also learned from the other diagrams. The first picture in the last column gives the corresponding power spectrum. As expected, the lowest frequency in the spectrum is $\nu = 1/T$, but

higher harmonics are present due to the nonlinear nature of the oscillation. In the second row, at $\nu = 197$ kHz, we see in the radius–time curve to the left that the two points now lie on two horizontal lines, indicating that the oscillation only repeats after two periods of the driving. The corresponding trajectory is again a limit cycle, but of more complex form. The two thick points indicate period $2T$ for the limit cycle, as do just the two points in the Poincaré section plane. The power spectrum of the radius–time curve to the right now has a lowest frequency of $\nu = 1/(2T)$. Again, as the oscillation is nonlinear, the harmonics of $\nu = 1/(2T)$ show up; i.e., lines at $\nu = 3/(2T), 5/(2T),...$ are newly introduced into the spectrum with respect to the first row. Clearly, this second row indicates that period doubling has taken place. In the third and fourth rows, it is demonstrated how period doubling proceeds, leading to rapidly more complex trajectories making up the limit cycle. The spectrum is filled up with new lines exactly between two old lines of the spectrum at each period doubling. In this way, the irregular, chaotic state in the last row is reached. As the trajectory is aperiodic, the radius–time curve never repeats, and only gives a short segment of the actual possible oscillation sequence of larger and smaller oscillations. The trajectory now forms part of a strange attractor created out of a limit cycle. Only a few revolutions are plotted, as otherwise a whole area would have turned black, leaving no discernible structure. Also, the points indicating the elapse of one period of the driving have been omitted so as not to disturb the picture. The spectrum now contains some amount of noise which is intrinsic, i.e., coming from the deterministic dynamics itself. The Poincaré section plot contains many more section points of the attractor than correspond to the turns of the trajectory plotted in the last row (second column) and thus best displays the involved nature of the chaotic attractor. However, the way points are imaged from one part of the attractor to another cannot be given by this type of static picture.

The reader will surely agree that by simply looking at a radius–time curve (a solution curve of a differential equation) in the chaotic region, the statement that intrinsic aperiodicity is present cannot be made. However, that no periodic solutions should be present in certain parameter regions has been argued nevertheless from repeated and strong trials,[32] since these regions cannot be overlooked. Even as early as 1969 it had been observed that these regions occur near regions where subharmonics are present, and a connection between subharmonics and noise has been conjectured.[40]

The regions of chaos are best visualized in bifurcation diagrams, as then the route to chaos can automatically be discerned. Figure 14 just gives one example of a bifurcation diagram where the normalized radius of the bubble at a certain phase of the driving sound field (when transients have decayed) is plotted versus the driving frequency as control parameter. The plot has been obtained in the following way. First, a maximum of 100 oscillations is calculated to let transients die out. Then 100 points of the radius coordinate of the Poincaré section plane (given by constant phase of the driving) are plotted. When there is a limit cycle of period T, these 100 points will neatly fall one upon the other and just one point will show up in the diagram. Then, starting from this
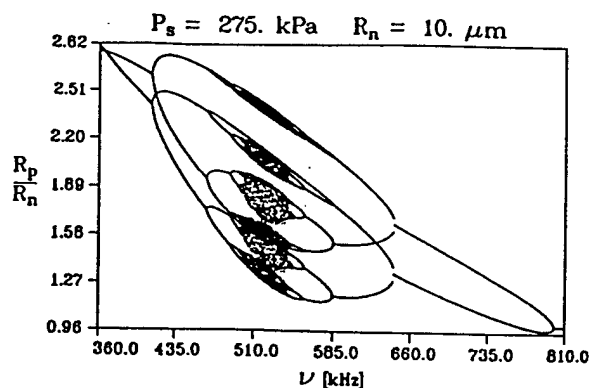
P_s = 275. kPa R_n = 10. μm

FIG. 14. Bifurcation diagram of a bubble oscillator.

FIG. 15. Phase diagram of a bubble oscillator (just a few bifurcation lines from the actually infinitely many).

value as the initial condition, the next series of oscillations is calculated for a slightly increased (or decreased) frequency yielding the next attractor. If it is of period $2T$, two points will show up in the diagram. In the case of a chaotic attractor, 100 points will be plotted, scattered along a vertical line at the given frequency. In this way, the properties of the system have been made visible for 1350 frequency points. Figure 14 shows an interesting typical picture with period doubling to and from chaos making up a complicated "bubble" structure.

Again, the reader is pointed to a forthcoming article for more details. There, the growth of these bubbles and their distribution in parameter space along resonance horns yielding a superstructure[22] of bifurcations will be discussed. This superstructure is conjectured to be universal in some sense and for a certain class of driven nonlinear oscillators.

VII. PHASE DIAGRAMS (PARAMETER SPACE DIAGRAMS)

When there is more than one control parameter in a system, its properties can only be given in a series of bifurcation diagrams, where one parameter is chosen as control parameter, the other ones held fixed and only changed from one diagram to the next. Experience shows that it is not easy to grasp the gross properties of a system from such sequences. A more condensed view combining such sequences would be desirable. This can be done with the help of *phase diagrams* or, equivalently, *parameter space diagrams*. The name "phase diagram" comes from thermodynamics, where in a pV diagram the areas are marked where, e.g., a liquid or gaseous phase is present, and curves are plotted to denote their boundaries. At the points of the curves, a phase transition takes place and also coexisting phases are known. In exactly the same sense, phase diagrams of (nonlinear) dynamical systems are to be understood. Theoretically, they are the plot of the bifurcation set in parameter space together with the indication of the kind of attractor ("phase") in the areas created by the curves or planes of the bifurcation set. To calculate even a fairly complete phase diagram of a dynamical system is a time consuming task and usually needs hours of computer time. An early example of a phase diagram for a bubble oscillator is shown in Fig. 15, taken from
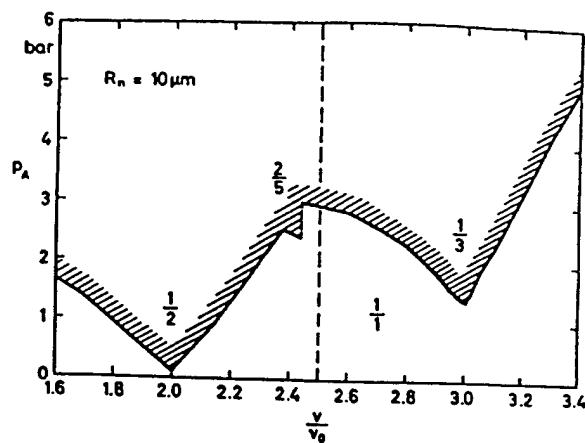
Ref. 32. In the parameter plane spanned by the (normalized) driving frequency v/v_0 and the sound-pressure amplitude P_A, one curve belonging to the bifurcation set of the bubble oscillator (see Ref. 32 for the equation used) is drawn, separating regions of period-1, period-2, period-3, and period-5 oscillation. Below the curve, period-1 oscillations, i.e., oscillations with the same period as the driving (period-1 limit cycles), occur; above this curve, but only very near to it, period-2, -3, and -5 limit cycles are present. The region above the curve will be further divided by a complicated infinite set of bifurcation lines, since period doubling sets in and further saddle-node bifurcations occur. One more phase diagram using the same bubble model as in Ref. 32 is given in Ref. 34. Very detailed phase diagrams of the Duffing oscillator $\ddot{x} + d\dot{x} + x + x^3 = f\cos \omega t$ and the Toda oscillator $\ddot{x} + d\dot{x} + e^x - 1 = f\cos \omega t$ can be found in Refs. 21 and 38, respectively. Phase diagrams can also be measured. An example of a measured phase diagram of a simple electronic oscillator is given in Ref. 41.

The determination of phase diagrams of dynamical systems is one of the main tasks of chaos research as, in a sense, they give a complete qualitative and even partly quantitative overview of possible behavior of a nonlinear dynamical system. Special methods are presently being developed to locate and follow bifurcation curves in parameter space, to speed up the calculations. Due to increase in computer speed and availability of computer time, the near future will see a quickly growing set of phase diagrams for various systems.

VIII. FRACTAL DIMENSIONS

The notion of the dimension of an object (physical or mathematical) had long occupied physicists and mathematicians until a solution came into sight. Cantor and Poincaré both put great effort into this question but failed. It was not until Hausdorff's article, "Dimension und äusseres Mass" ("Dimension and external measure"), that a satisfactory definition could be put forward.[42] This definition of the dimension of a set of points naturally leads to *fractal dimensions*, i.e., dimensions that are not just natural numbers. For a long time, sets with these properties (fractals) were thought to be

purely mathematical objects until their obviously abundant occurrence in nature was demonstrated by Mandelbrot.[17] Chaos physics has strongly expanded the importance of fractals and their dimensions. Chaotic attractors are known for their usually peculiar shapes, which point to fractal dimensions. This raises the question of how to determine fractal dimensions, especially when they are encountered in experiments. Indeed, concepts have been developed whereby fractal dimensions can be determined both from numerically calculated strange attractors and from measured data.[43-53]

An infinity of different definitions of a dimension has been introduced having their value in describing the inhomogeneity of the attractor. Only a simplified notion of the *Hausdorff dimension*, called *capacity d_0*, and the most often used *correlation dimension d_2* will be discussed in more detail here.

First, the definition of the dimension d_0 of a set of points $A \subset R^m$ is given:

$$d_0 = \lim_{r \to 0} \log M(r)/\log(1/r), \qquad (10)$$

where r is the edge length of an m-dimensional cube and $M(r)$ is the lowest number of m-dimensional cubes of edge length r to cover the given set A (a chaotic attractor). When this definition is applied to a point, a line, an area, and a volume, the dimensions $d_0 = 0, 1, 2$, and 3 are obtained as they should be. But the definition is much more powerful. Also, Cantor sets now get a dimension, usually fractal as it turns out. Figure 16 gives the standard example of a Cantor set whereby, starting with the unit interval [0,1], the middle third without the end points is successively taken out of the remaining intervals. In the figure, to the left, the edge length r is given that is conveniently taken to cover the set, and to the right the number M of cubes (intervals) that is needed to cover the set. According to the definition of d_0, one then easily gets, by simply inserting the sequence as given in the figure into Eq. (10),

$$d_0 = \lim_{k \to \infty} \frac{\log 2^k}{\log 3^k} = \frac{\log 2}{\log 3} = 0.6309 \cdots, \qquad (11)$$

a noninteger number. This is the fractal dimension of the Cantor set.

It turns out that for systems with a high-dimensional state space the direct use of the definition of d_0 (box counting) is not practical in numerical applications. From the next members of dimensions, the information dimension d_1,
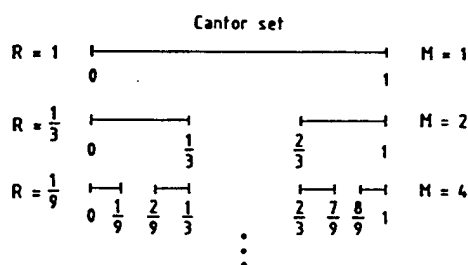
the correlation dimension d_2, and the higher-order dimensions $d_3, d_4, \ldots$, the correlation dimension d_2 is the most attractive from the experimentalist's point of view.[43,44] For the sequence of dimensions, the relations $\cdots d_n > d_{n-1} > \cdots > d_2 > d_1 > d_0$ hold, and often the d_n are nearly all the same. Thus d_2, which is very convenient to determine numerically, is a good estimate of "the" fractal dimension of a strange attractor.

The *correlation dimension* is defined as

$$d_2 = \lim_{r \to 0} \log C_m(r)/\log r, \qquad (12)$$

where r again is the edge length of an m-dimensional cube and $C_m(r)$ is the so-called correlation sum

$$C_m(r) = \lim_{N \to \infty} \frac{1}{N^2} \sum_{\substack{k,l=1 \\ k \neq l}}^{N} H(r - \|p_k^{(m)} - p_l^{(m)}\|). \qquad (13)$$

In this expression, N is the number of points in R^m of the (strange) attractor available from some calculations or measurements, H is the Heaviside step function [$H(x) = 0$ for $x < 0, H(x) = 1$ for $x > 0$ and, used here, $H(0) = 0$], p_k are the points of the attractor, and $\|\cdot\|$ is a suitable norm, e.g., the Euclidean norm. The dimension does not depend on the norm; therefore, any norm may be chosen that is most convenient for numerical computation. In our determination of cavitation noise attractors, we used d_2 together with the maximum norm.[47] Equation (12) immediately shows that d_2 can be determined from the slope of the curve obtained when $C_m(r)$ is plotted versus r, each on a logarithmic scale.

An example for the determination of d_2 is given in Fig. 17 for the strange bubble attractor of Fig. 11. In Fig. 17(a), the log $C(r)$ vs log r curve is plotted with 100 000 points of the attractor in the Poincaré section plane. Figure 17(b) shows the local slope of the curve in Fig. 17(a) whereby "local" means a fit of the slope over a region of a quarter of a decade. It is seen that a plateau in the local slope only occurs for values of log r between -1.5 and -0.5 giving a fractal dimension of $d_2 = 1.3 \pm 0.1$. For larger r's the gross structure of the attractor naturally leads to a decreasing local slope until at r's above the size of the attractor the slope is zero, because all points of the attractor fit into the cube of edge length r. At low r's the following happens. Because of the finite precision in the calculations, the exact location of the attractor points in the Poincaré section plane is only known to a limited number of digits. Therefore, the attractor looks more and more noisy on smaller and smaller scales. For these reasons the local slope for small r's tends to 2, which is the dimension of the Poincaré section plane. Thus only a region between small and large values of r is available for the determination of the fractal dimension. The fractal dimension of $d_2 = 1.3$ is valid for the attractor in the Poincaré section plane. To get the fractal dimension of the whole attractor, one dimension has to be added giving $d_2 = 2.3$. Some remarks may be in order to warn the reader who wants to apply this formalism for fractal dimension determination. The method must be used very carefully, especially with experimental data when used in conjunction with reconstructed attractors (see Sec. IX). The errors are usually very large.

Cantor set

$R = 1$ $M = 1$
0   1

$R = \frac{1}{3}$ $M = 2$
0   $\frac{1}{3}$ $\frac{2}{3}$ 1

$R = \frac{1}{9}$ $M = 4$
0 $\frac{1}{9}$ $\frac{2}{9}$ $\frac{1}{3}$ $\frac{2}{3}$ $\frac{7}{9}$ $\frac{8}{9}$ 1

FIG. 16. Cantor set construction by successively taking out the middle third without its end points.

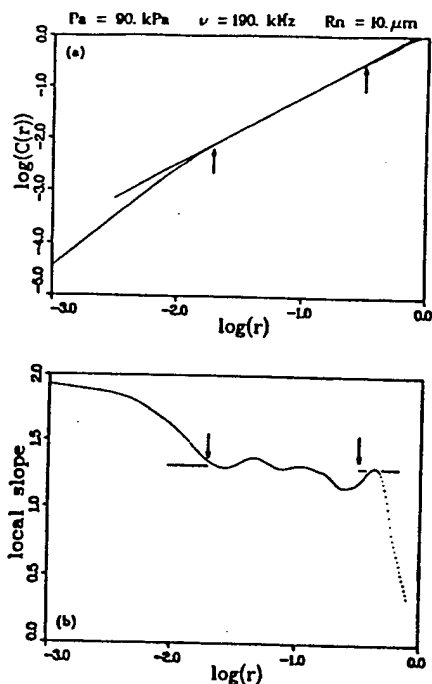Pa = 90. kPa ν = 190. kHz   Rn = 10. μm

FIG. 17. Determination of the correlation dimension d_2 for the strange bubble attractor given in Fig. 11. (a) Correlation sum $C(r)$ versus the cube edge length r on a doubly logarithmic scale, (b) local slope [fit to the curve in (a) over a range of a quarter of a decade] versus r. The fit region is from -1.7 to -0.5. In this region the local slope has a plateau giving $d_2 = 1.3 \pm 0.1$.

The state of the art of the techniques available is discussed in Ref. 46.

Thus far we have discussed two types of dimensions, d_0 and d_2, out of the series d_n, $n = 0,1,2,...$. The definition of dimension can even be extended to d_q, where q is any real number.[48] To an experimentalist, this generalization may seem rather sophisticated and beyond anything that can be measured. The opposite is true.[49-54] Only the full set $\{d_q, q \in R\}$, or equivalently the (smooth) scaling spectrum $f(\alpha)$ of (local) scaling indices α on the attractor,[49] describes the global structure of the attractor satisfactorily and simultaneously in a measurable way. The scaling spectrum $f(\alpha)$ has a quite simple meaning. Take a point of a (strange) set (attractor) and a small sphere around it. Then the number of points of the set inside the sphere will scale with some exponent (index) α when the radius of the sphere goes to zero. Different points of the set may have different indices α. The spectrum $f(\alpha)$ characterizes the strength of a scaling index α or more precisely, $f(\alpha)$ is the global (Hausdorff) dimension of the subset of points of the set (attractor) with scaling index α. The relation $f(\alpha) \leq d_0$ holds, d_0 being the global dimension of the total set. Thus the maximum f_{max} of the spectrum yields the Hausdorff dimension. The spectrum of the Cantor set mentioned above is a line spectrum with just a single line at $\alpha = d_0 = \log 2/\log 3$. The dimension $f(\alpha)$ at $\alpha = d_0$ then must have the same value $f(\alpha = d_0) = d_0$, as every point of the Cantor set scales with the same α; thus $f(\alpha) = d_0 \int \delta(\alpha - d_0) \, dx$, where δ is the Dirac δ function. Sets of this kind showing uniform scaling behavior are called self-similar (in a strict sense) and are *simple fractal sets*.

Strange attractors usually have a more complicated scaling behavior. However, according to the scaling spectrum $f(\alpha)$, they can be thought of as a "superposition" of simple fractal sets given by the subsets of uniform scaling behavior with index α. They are therefore called *multifractal sets* or, briefly, *multifractals*. For special cases, the scaling spectrum seems to be universal, i.e., the same for all systems, e.g., at the accumulation point of period doubling.[49,50] The experiments so far confirm the theory,[52-54] and it seems that the experimental determination of $f(\alpha)$[51-54] may be a better way to characterize a strange attractor than a single dimension as, for instance, d_2.

IX. RECONSTRUCTION OF ATTRACTORS

In mathematical models of physical dynamical systems, the dynamic evolution is visualized in the state space whose dimension is given by the number of dependent variables (see Sec. I and Fig. 1). In experiments, the state space is usually not known beforehand and often only one variable of the system can be measured, e.g., a velocity component at a point in the liquid in a Rayleigh–Bénard experiment (a liquid layer heated from below). Thus only a projection of a trajectory of the system with a usually high-dimensional state space onto a single coordinate axis is given. One may ask whether, under these circumstances, experiments make sense at all, when only such minor information about a system is obtainable. Fortunately, it can be shown that one variable already contains most of the information about the total system and not just a minor part.[55,56] It may at first sight seem really astonishing that one coordinate of a trajectory in a high-dimensional state space should contain almost all information to determine the other coordinates of the trajectory. There is an intuitive way to understand why this may be possible. The single variable considered develops in time not due to its own isolated dynamical law but (usually) is coupled to all the other dependent variables of the system. Its dynamics, therefore, reflects the influence of all the other variables which in turn react to the influence of the variable considered. This mutual interaction lets a single variable contain the dynamics of all the other ones. An example may help the reader. Take a simple oscillator: a mass on a spring. The state space is the usual phase space given by the coordinates elongation x and velocity $v = \dot{x}$. From our knowledge of the dynamics of the system, we know that with the knowledge of $x(t)$ (one coordinate only!) we also know $v(t) = \dot{x}(t)$. Thus, when only one coordinate is measured, say the elongation x, then from $x(t)$ the velocity $v(t)$ (the other coordinate) can be determined.

The example also reveals the difficulties in an actual application. It may be possible in principle to reconstruct the dynamical evolution of the other variables from just one but where do we know the construction law from? In the above example the velocity $v(t)$ is given from $x(t)$ through the operation of differentiation. If we would have measured $v(t)$ instead of $x(t)$, we should apply integration instead of differentiation to get $x(t)$ out of $v(t)$. Thus the information may be there but hopelessly encrypted. At this state of affairs it came as a surprise, that a simple general solution could be

32

found to this problem,[55-57] which may be formulated as the problem of reconstructing an attractor of a dynamical system from a time series of one (measured) variable only. Theory states: For almost every state variable $p(t)$ and for almost every time interval T, a trajectory in an n-dimensional state space can be constructed from the (measured) values $[p(kt_s), k = 1,2,...,N]$ by grouping n values to n-dimensional vectors:

$$\mathbf{p}_k^{(n)} = [p(kt_s), p(kt_s + T),..., p(kt_s + (n-1)T)], \quad (14)$$

where t_s is the sampling interval at which samples of the variable p are taken. It is advisable to choose the delay time T as a multiple of t_s, $T = lt_s$, $l \in N$, to avoid interpolation. The above construction yields a point set $A^{(n)} = (\mathbf{p}_k^{(n)}, k = 1, 2,...,N - n)$ in the embedding space $\mathbf{R}^n$, which represents the attractor. The sampling time t_s and the delay time T must be chosen appropriately according to the problem under investigation (see Ref. 46 for details). When t_s and also T are too small, then from one sample to the next there is little variation and the points $\mathbf{p}_k^{(n)}$ all lie on a diagonal in the embedding space. On the other hand, when t_s and T are too large, then the point set $A^{(n)}$ and the attractor obtained by connecting consecutive points $\mathbf{p}_k^{(n)}$ get a fuzzy appearance. In both cases, the reconstruction and visualization of the attractor are not very helpful. Figure 18 shows an example of the influence of the delay time T on the reconstruction of a calculated

chaotic bubble attractor. The same equation as that given in Sec. VI has been used. For the reconstruction, only the calculated radii of the bubble have been taken and embedded in a three-dimensional state space by using four different delay times T of $\frac{1}{16}$, $\frac{1}{4}$, $\frac{1}{2}$, and 1 in units of the period T_0 of the driving sound field. The bubble had a radius at rest of $100 \mu m$ and was driven by a sinusoidal sound wave of amplitude 310 kPa and frequency 22.9 kHz. The reconstruction is best at $T = \frac{1}{4}T_0$. This is the same value as for a simple harmonic wave of period T_0 where the elongation x and the velocity $v = \dot{x}$ are 90° out of phase, i.e., by $\frac{1}{4}T_0$.

The chosen dimension n is called the *embedding dimension*. Which dimension n should be taken is not known beforehand when the system is not known sufficiently (as usual in real experiments). Sometimes it happens that a three-dimensional state space is sufficient as embedding space. This, of course, is true for mathematical models with an *a priori* three-dimensional state space as, e.g., one-dimensional driven oscillators (Duffing, van der Pol, Toda, Morse, spherical bubble). In the context of acoustics, a three-dimensional embedding space was found sufficient in representing an acoustic cavitation noise attractor.[47,20] In this experiment, a liquid is irradiated with sound of high intensity and the sound output from the liquid is measured. The measurement yields pressure–time samples $p(kt_s)$ that may be used to construct an attractor. An example is given in Fig. 19. The reconstructed attractor in a three-dimensional em-
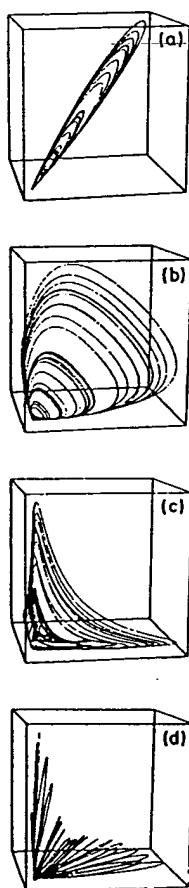

FIG. 18. Four reconstructions of a numerically obtained bubble oscillator from radii data only with different delay times T of (a) $\frac{1}{16} T_0$, (b) $\frac{1}{4} T_0$, (c) $\frac{1}{2} T_0$, (d) $1 T_0$, T_0 being the period of the driving sound wave of frequency 22.9 kHz and amplitude 310 kPa. The bubble has a radius at rest of $R_n = 100 \mu m$. The sampling rate t_s is 1 μs. For better visualization the transformed coordinates $R^* = \exp(2R/R_n)$ have been used. (Courtesy of J. Holzfuss.)
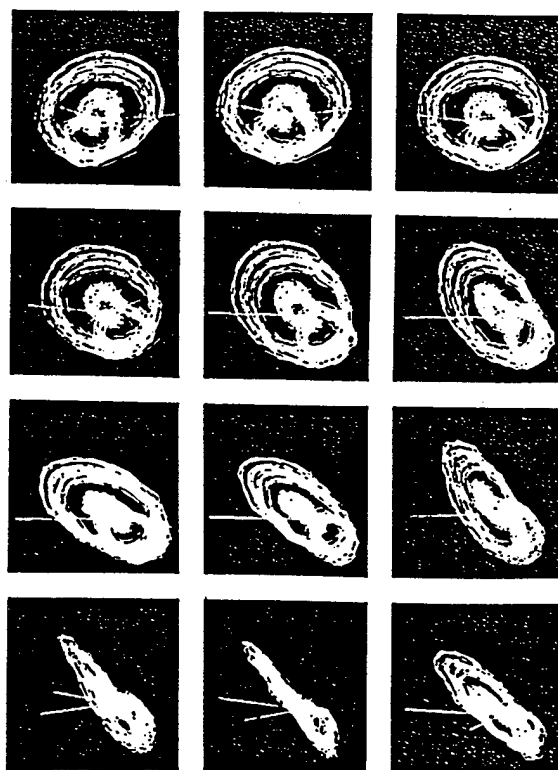

FIG. 19. Example of a reconstructed acoustic cavitation noise attractor from sampled pressure values in a three-dimensional embedding space. The attractor is shown from different directions for visualizing its spatial structure. Sampling time is $t_s = 1 \mu s$, and delay time is $T = 5 \mu s$.

bedding space is viewed from different directions to show its three-dimensional structure and almost flat overall appearance. The very pronounced structure suggests that the measured acoustic noise is of quite simple deterministic origin. A model with a three-dimensional state space should be sufficient. To deduct the structure of the equations from the attractor, however, is beyond the state of present knowledge. It is interesting to note that a very similar attractor has been found by Roessler[58] in a mathematically constructed model of hyperchaos (see Sec. X for a definition of hyperchaos).

Once an attractor has been reconstructed, its properties can be investigated. Of special interest is the (fractal) dimension of the attractor, which may be determined using the methods of Sec. VIII. Indeed, as the embedding dimension n is not known for a real experimental system under investigation, n is successively increased, and the dimension of the point set $A^{(n)}$ is determined. When the system is of deterministic origin with a low-dimensional state space, then at some n the (fractal) dimension of $A^{(n)}$ will stabilize at some definite value (smaller than n). The largest fractal dimension of $A^{(n)}$, $n = 1,2,...$, obtained in this way then determines the relevant number of (nonlinear) degrees of freedom (number of dependent variables) of the dynamic system investigated. In this way, the dimension of an acoustic cavitation noise attractor has been determined to be about 2.5 (Ref. 47).

X. LYAPUNOV EXPONENTS AND LYAPUNOV SPECTRA

Chaotic systems exhibit *sensitive dependence on initial conditions*. This expression has been introduced to denote the property of a chaotic system, that small differences in the initial conditions, however small, are *persistently* magnified because of the dynamics of the system, so that in a finite time the system attains totally different states. It is not difficult to envisage this property with systems that are not bounded, like unstable linear systems. But physical systems are in general bounded, and it is not at all obvious how a persistent magnification of small differences is brought about. It seems that a sensitive dependence on initial conditions can only occur through a *stretching* and *folding* process of volumes of state space under the action of the dynamics. This process is depicted in Fig. 20. A persistent simple stretching would expand one direction more and more without bounds [Fig. 20(a)]. Neighboring points thus get more and more distant.
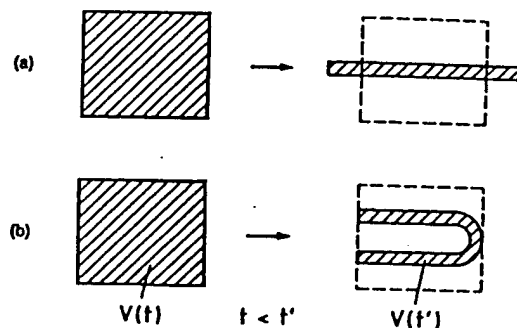
This expansion in one direction can take place in a bounded volume of state space only when an additional folding process occurs [Fig. 20(b)]. The transition from $V(t)$ to $V(t')$, $t' > t$ (see Fig. 20), may be viewed as a map, a so-called "horseshoe map." Maps with similar properties are the baker's transformation and Arnold's cat map.[1] When iterated they should give the simplest examples of a dynamical system with sensitive dependence on initial conditions in two dimensions. This is the reason why they are intensely studied.

The notion of sensitive dependence on initial conditions is made more precise through the introduction of Lyapunov exponents and Lyapunov spectra. Their definition can nicely be illustrated.[59] Take a small sphere in state space encircling a point of a given trajectory (Fig. 21). The points of the sphere can be viewed as initial points of trajectories. This sphere is shifted in state space and deformed due to the dynamics so that at a later time a deformed sphere is present. To properly make use of this idea, mathematically an infinitesimal sphere and its deformation into an ellipsoid with principal axes $r_i(t)$ $i = 1,2,...,m$ (m = dimension of the state space) are considered. The *Lyapunov exponent* λ_i may then, *cum grano salis*, be defined by

$$\lambda_i = \lim_{t \to \infty} \lim_{r_i(0) \to 0} \frac{1}{t} \log \frac{r_i(t)}{r_i(0)}. \qquad (15)$$

The set $\{\lambda_i, i = 1,...,m\}$, whereby the λ_i usually are ordered $\lambda_1 > \lambda_2 > \cdots > \lambda_m$, is called the *Lyapunov spectrum*. It is to be recalled that the $r_i(t)$ should stay infinitesimally small. Then the linearized local dynamics applies that has to be taken along a nonlinear orbit. This is the meaning of $t \to \infty$. A strict mathematical definition resorts to linearized flow maps and may be found in Ref. 15. When $\lambda_1 > 0$, a time-dependent "direction" exists, in which the system expands. The system is then said to be chaotic (when, additionally, it is bounded). All along a trajectory, neighboring trajectories will retreat. A Lyapunov exponent is a number which, due to the limit $t \to \infty$, is a property of the whole trajectory. To get an idea of how uniformly neighboring trajectories recede from a given trajectory, Lyapunov exponents can be defined for pieces of trajectories to identify the most chaotic parts.

In dissipative systems, the final motion takes place on attractors. Besides the fractal dimensions (or the scaling spectrum), the Lyapunov spectrum may serve to characterize these attractors. With the help of the Lyapunov exponents, the motion on a chaotic attractor can be made more precise, as the following relations and definitions hold (continuous systems, $\lambda_1 > \lambda_2 > \cdots > \lambda_m$):

FIG. 20. Stretching and folding of a volume of state space. (a) Stretching only, (b) stretching and folding.
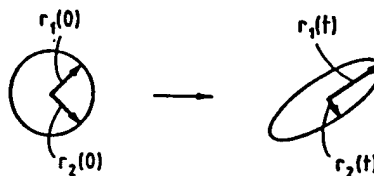
FIG. 21. Notions for the definition of Lyapunov exponents. A small sphere in state space is deformed to an ellipsoid indicating expansion or contraction of neighboring trajectories.

1989 J. Acoust. Soc. Am., Vol. 84, No. 6, December 1988

W. Lauterborn and U. Parlitz: Chaos acoustics 1989

34

$\lambda_1 < 0 \rightarrow$ fixed point;

$\lambda_1 = 0 \rightarrow$ limit cycle ($\lambda_2 < 0$);

$\lambda_1 = \lambda_2 = 0 \rightarrow$ quasiperiodic attractor on a

two-dimensional surface ($\lambda_3 < 0$);

$\lambda_1 > 0 \rightarrow$ chaotic attractor ($\lambda_2 = 0$);

$\lambda_1 > \lambda_2 > 0 \rightarrow$ hyperchaotic attractor

$(\lambda_3 \text{ may be} >, =, \text{or} < 0)$.

The calculation of Lyapunov spectra is an area of active research. Lyapunov spectra can also be calculated from (measured) time series via the reconstruction of attractors described in Sec. IX. The reader who wants to dig deeper into this subject may consult Refs. 59–62 and 15 and the references therein.

XI. LYAPUNOV DIAGRAM

When the Lyapunov spectrum is plotted in dependence on a control parameter, a *Lyapunov diagram* is obtained. The regions of chaotic dynamics are easily recognized by the largest Lyapunov exponent λ_1 being greater than zero. Figure 22 shows an example of what a typical Lyapunov diagram looks like. Again the comparatively simple case of the logistic parabola has been taken where there is only one Lyapunov exponent λ. In the case of one-dimensional maps, λ is simply given by[39]

$$\lambda(\mu) = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \log |f'_\mu [f_\mu^{(k-1)}(x_0)]|, \qquad (16)$$

f'_μ means the derivative of f_μ with respect to x, $f_\mu^{(k)}$ is the k th iteration of f_μ. For the logistic parabola, $f_\mu(x) = 4\mu x(1-x)$ and $f'_\mu(x) = 4\mu(1-2x)$ hold.

XII. WINDING NUMBERS

Winding numbers (or rotation numbers) were originally introduced by Poincaré to describe periodic and quasiperiodic trajectories that are part of an invariant torus in phase space. An n-dimensional torus T is given as the Cartesian product of n unit circles S^1, $T = S^1 \times S^1 \times \cdots \times S^1$. Thus any trajectory on the torus may be projected onto these circles resulting in some motions on the circles that are parametrized by n angles α_i ($i = 1,...,n$). In Fig. 23, the construction of a two-dimensional torus is shown. In the two-dimen-
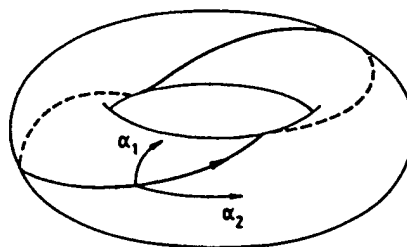
FIG. 23. A two-dimensional torus and its two angle variables α_1 and α_2. A periodic trajectory with a winding number of 2 is plotted.

sional case, the winding number w is defined as the limit (17) of the ratio of the angle variables α_i for $t \to \infty$:

$$w = \lim_{t \to \infty} \alpha_1(t)/\alpha_2(t). \qquad (17)$$

In higher dimensions, more than one ratio exists and a whole set of winding numbers may be defined. Periodic orbits possess rational winding numbers and are called resonant.[63,64]

The essential prerequisite of this definition of winding numbers is the existence of a so-called invariant torus in state space. A torus in state space is called invariant when any trajectory of the system that starts on the torus stays on it for all times. Such tori are usually created by Hopf bifurcations (see Sec. II) and occur, for example, in self-excited systems like the driven van der Pol oscillator.[37] If a parameter is altered, usually a typical sequence of periodic and quasiperiodic orbits on the torus occurs. As in the case of period doubling, this sequence is governed by universal scaling laws.[65,66] The one-dimensional map that is referred to when dealing with such mode-locking sequences is the sine circle map (18):

$$\Theta_{n+1} = f(\Theta_n) \bmod 1, \quad n = 0,1,2,..., \qquad (18)$$

with

$$f(x) = x + \Omega + (K/2\pi)\sin 2\pi x, \quad x \in \mathbb{R}, \qquad (19)$$

where Ω and K are parameters.

In this case, the state space itself is a one-dimensional torus $T = S^1$ (unit circle), and the winding number is given as

$$w = \lim_{n \to \infty} f^{(n)}(\Theta_0)/n, \qquad (20)$$

where $f^{(n)}$ denotes the n-fold iterate of f and Θ_0 is some initial value. The winding number w counts the mean number of revolutions per iteration.

Figure 24(a) shows a bifurcation diagram of the circle map. Between the periodic windows quasiperiodic orbits occur. The construction law of the self-similar pattern of periodic windows can best be seen in the *winding number diagram* in Fig. 24(b), which is an example of a so-called devil's staircase. Within a periodic window the winding number is constant, yielding one step of the staircase. Given two periodic windows with winding numbers, $w_1 = n_1/m_1$ and $w_2 = n_2/m_2$, another periodic window with winding number $w = (n_1 + n_2)/(m_1 + m_2)$ can always be found between these windows. When continued, this construction leads to a so-called Farey tree of periodic windows.
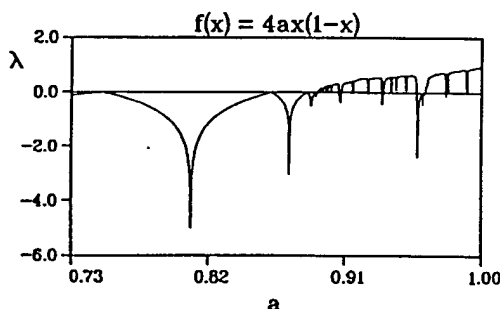

FIG. 22. Lyapunov diagram for the logistic parabola. Chaotic dynamics is indicated by a Lyapunov exponent greater than zero.

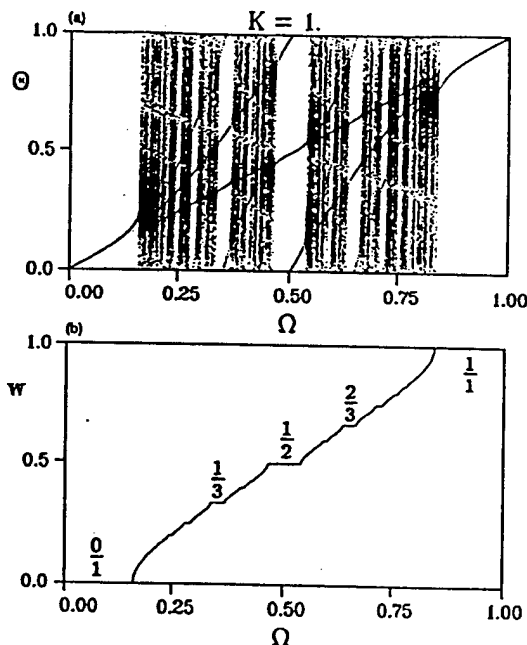FIG. 24. A bifurcation diagram (a) and a winding number diagram (b) of the sine circle map. The winding number diagram is an example of a devil's staircase. The staircase is really devilish because between every two steps there are infinitely many other steps and climbing up or down the staircase from step to step actually is impossible.

As mentioned above, winding numbers can only be defined in those cases where the trajectory is part of an invariant torus. But there are many systems, like the periodically driven Duffing oscillator or the bubble oscillator, where the existence of such an invariant torus can be definitely excluded. This means that the definition of the winding number given above cannot be applied. We therefore introduced a similar quantity called generalized winding number[67,37] to classify the resonances of this type of systems. As in the case of the Lyapunov exponents, we consider a trajectory γ' that starts in the vicinity of a given orbit γ. But now we are not interested in the divergence or convergence of these trajectories, but in the way they are twisted around each other. A frequency Ω may be attached to the orbit γ that gives the mean number of twists of γ' about γ per unit time. To compute this *torsion frequency* Ω, the linearized dynamics along the whole orbit has to be considered (for details, see Ref. 67). If we choose as unit time the period T_0 of the oscillation, the number of twists is called the *torsion number n* of the closed orbit. Torsion numbers may be used to classify resonances and bifurcation curves in the parameter space of nonlinear oscillators.[67,38] If the solution becomes aperiodic, the torsion number cannot be defined anymore, but in this case the ratio $w = \Omega/\omega$ of the torsion frequency Ω and the driving frequency ω of the oscillator still exists. We call this ratio w the (generalized) winding number, because it equals the winding number introduced above in those cases where the trajectory is part of an invariant two-dimensional torus. For period-doubling cascades, two recursion schemes exist for the winding numbers $w_1, w_2, w_3, \ldots$, at the period-doubling bifurcation points of the control parameter.[37,38] The winding

number of a chaotic solution describes some aspects of the folded geometry of the strange attractor. Details of the procedure to compute winding numbers of nonlinear oscillators are given in Refs. 37 and 67.

XIII. CONCLUDING REMARKS

The main new methods of chaos physics have been presented in a tutorial manner and illustrated with examples from acoustics, especially driven bubble oscillations and acoustic cavitation noise. The methods described have been invented to characterize irregular motion from deterministic systems more specifically than by, for instance, their Fourier spectra and correlations which are intrinsically linear concepts. In this context, chaos physics suggests that the notion of a *degree of freedom* must be revised. In conventional physics, a degree of freedom is connected with a linear mode (a harmonic oscillator). A Fourier spectrum with many lines is interpreted as coming from a system with as many harmonic oscillators and thus as many degrees of freedom. One of the results of chaos physics is that there are nonlinear systems with just a three-dimensional state space, for instance the driven bubble oscillator, that will give rise to broadband Fourier spectra (a sign of their irregular behavior). Thus they obviously have just three "degrees of freedom" instead of infinitely many as suggested by the Fourier spectrum because only three coordinates are needed to specify their state completely.

Chaos physics also affects the notion of *randomness*. Randomness is no longer a domain of high-dimensional systems too large to be properly described by a set of deterministic equations and initial conditions. A random-looking motion may well be the outcome of a deterministic system with a low-dimensional state space. Deterministic equations are thus far more capable in describing nature than previously thought provided that they are nonlinear. Indeed, *nonlinearity* is the necessary basic ingredient for this capability.

As nonlinear oscillators are the natural extension of the harmonic oscillator that plays a fundamental role in physics, chaos physics will find one of its main applications in the area of nonlinear oscillatory systems. The investigations there will center around the laws that are valid in the fully nonlinear case. A few universal laws have already been found, and many more are waiting for their discovery. On a local scale in parameter space, the most prominent universal phenomenon is period doubling. On a global scale, it is to be expected that the bifurcation superstructure[22,32,33,38] of resonances of nonlinear oscillators is of universal nature. Chaos physics is a rapidly growing field with applications all over the different areas of physics and even extending to chemistry, biology, medicine, ecology, and economy. The methods described in this article may help in the dissemination of these ideas, and the authors would be proud if some readers would be attracted to the fascinating and rewarding field of chaotic dynamics.

ACKNOWLEDGMENTS

1991   J. Acoust. Soc. Am., Vol. 84, No. 6, December 1988   W. Lauterborn and U. Parlitz: Chaos acoustics   1991

36

Göttingen, for many stimulating discussions, V. Englisch for supplying us with a beautiful example of basins of attraction, and J. Holzfuss for letting us use his reconstructed attractors. The computations have been carried out on a SPERRY 1100 and a VAX 8650 of the Gesellschaft für wissenschaftliche Datenverarbeitung, Göttingen, and on a CRAY X-MP of the Konrad-Zuse-Zentrum für Informationstechnik, Berlin.

[1] H. G. Schuster, *Deterministic Chaos* (Physik Verlag, Weinheim, 1984).

[2] P. Bergé, Y. Pomeau, and Ch. Vidal, *L'Ordre dans le Chaos* (Hermann, Paris, 1984).

[3] P. Cvitanović (Ed.), *Universality in Chaos* (Hilger, Bristol, England, 1984).

[4] Hao Bai-Lin (Ed.), *Chaos* (World Scientific, Singapore, 1984).

[5] A. V. Holden (Ed.), *Chaos* (Manchester U. P., Manchester, England, 1986).

[6] F. C. Moon, *Chaotic Vibrations* (Wiley, New York, 1987).

[7] J. M. T. Thompson and H. B. Stewart, *Nonlinear Dynamics and Chaos* (Wiley, Chichester, England, 1986).

[8] T. S. Parker and L. O. Chua, "Chaos: A tutorial for engineers," Proc. IEEE 75, 982–1008 (1987).

[9] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcation of Vector Fields* (Springer, New York, 1983).

[10] D. Campbell and H. Rose (Eds.), *Order in Chaos* (North-Holland, Amsterdam, 1983).

[11] R. M. May, "Simple mathematical models with very complicated dynamics," Nature 261, 459–467 (1976).

[12] J. P. Eckmann, "Roads to turbulence in dissipative dynamical systems," Rev. Mod. Phys. 53, 643–654 (1981).

[13] E. Ott, "Strange attractors and chaotic motions of dynamical systems," Rev. Mod. Phys. 53, 655–671 (1981).

[14] M. W. Hirsch, "The dynamical systems approach to differential equations," Bull. Am. Math. Soc. 11, 1–64 (1984).

[15] J. P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," Rev. Mod. Phys. 57, 617–656 (1985).

[16] W. Lauterborn, "Solitons and deterministic chaos: new concepts in nonlinear acoustics," in *Fortschritte der Akustik*, edited by D. Guicking (FASE/DAGA'82, Göttingen, 1982), pp. 47–68.

[17] B. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, New York, 1983).

[18] W. Lauterborn and E. Cramer, "Subharmonic route to chaos observed in acoustics," Phys. Rev. Lett. 47, 1445–1448 (1981).

[19] W. Lauterborn, "Acoustic chaos," Phys. Today 40 (1), S-4–S-5 (January 1987).

[20] W. Lauterborn, "Acoustic turbulence," in *Frontiers in Physical Acoustics*, edited by D. Sette (North-Holland, Amsterdam, 1986), pp. 123–144.

[21] C. W. Smith, M. J. Tejwani, and D. A. Farris, "Bifurcation universality for first-sound subharmonic generation in superfluid helium-4," Phys. Rev. Lett. 48, 492–494 (1982).

[22] U. Parlitz and W. Lauterborn, "Superstructure in the bifurcation set of the Duffing equation $\ddot{x} + d\dot{x} + x + x^3 = f\cos(\omega t)$," Phys. Lett. A 107, 351–355 (1985).

[23] C. Grebogi, E. Kostelich, E. Ott, and A. Yorke, "Multi-dimensioned intertwined basin boundaries: basin structure of the kicked double rotator," Physica D 25, 347–360 (1987).

[24] D. Ruelle and F. Takens, "On the nature of turbulence," Commun. Math. Phys. 20, 167–192 (1971); also 23, 343–344 (1971).

[25] S. Newhouse, D. Ruelle, and F. Takens, "Occurrence of strange axiom—A attractors near quasiperiodic flow on T^m, $m > 3$," Commun. Math. Phys. 64, 35–40 (1978).

[26] Y. Pomeau and P. Manneville, "Intermittent transition to turbulence in dissipative dynamical systems," Commun. Math. Phys. 74, 189–197 (1980).

[27] P. Manneville and Y. Pomeau, "Different ways to turbulence in dissipative dynamical systems," Physica D 1, 219–226 (1980).

[28] A. N. Sharkovskii, "Coexistence of cycles of a continuous map of a line into itself," Ukr. Mat. Z. 16, 61–71 (1964).

[29] S. Grossmann and S. Thomae, "Invariant distributions and stationary correlation functions of one-dimensional discrete processes," Z. Naturforsch. 32(a), 1353–1363 (1977).

[30] P. Coullet and J. Tresser, "Iterations of endomorphisms and renormalization group," C.R. Hebd. Séances Acad. Sci. Ser. A 287, 577 (1978), in French.

[31] M. J. Feigenbaum, "Quantitative universality for a class of nonlinear transformations," J. Stat. Phys. 19, 25–52 (1978).

[32] W. Lauterborn, "Numerical investigation of nonlinear oscillations of gas bubbles in liquids," J. Acoust. Soc. Am. 59, 283–293 (1976).

[33] W. Lauterborn and E. Suchla, "Bifurcation superstructure in a model of acoustic turbulence," Phys. Rev. Lett. 53, 2304–2307 (1984).

[34] U. Parlitz and W. Lauterborn, "Periodic and chaotic bubble oscillations," in *Proceedings of the 12th International Congress on Acoustics (ICA 12)*, Toronto, 24–31 July 1986, edited by T. F. W. Embleton et al. (Beauregard, Canada, 1986), Vol. III, paper I4–7.

[35] W. Lauterborn and U. Parlitz, "On the bifurcation structure of bubble oscillators," *Proceedings of the XIth International Symposium on Nonlinear Acoustics*, edited by V. Kedrinskii (Institute of Hydrodynamics, Novosibirsk, 1987), pp. 71–80.

[36] P. Smereka, B. Birnir, and S. Banerjee, "Regular and chaotic bubble oscillations in periodically driven pressure fields," Phys. Fluids 30, 3342–3350 (1987).

[37] U. Parlitz and W. Lauterborn, "Period-doubling cascades and devil's staircases of the driven van der Pol oscillator," Phys. Rev. A 36, 1428–1434 (1987).

[38] T. Kurz and W. Lauterborn, "Bifurcation structure of the Toda oscillator," Phys. Rev. A 37, 1029–1031 (1988).

[39] P. Collet and J. P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems* (Birkhäuser, Basel, 1980).

[40] W. Lauterborn, "Subharmonic oscillations of gas bubbles in water," Acustica 22, 238–239 (1969/70), in German.

[41] T. Klinker, W. Meyer-Ilse, and W. Lauterborn, "Period doubling and chaotic behavior in a driven Toda oscillator," Phys. Lett. A 101, 371–375 (1984).

[42] F. Hausdorff, "Dimension und äusseres Mass (Dimension and exterior measure)," Math. Ann. 79, 157–179 (1919), in German.

[43] P. Grassberger and I. Procaccia, "Characterization of strange attractors," Phys. Rev. Lett. 50, 346–349 (1983).

[44] A. Ben-Mizrachi, I. Procaccia, and P. Grassberger, "The characterization of experimental (noisy) strange attractors," Phys. Rev. A 29, 975–977 (1984).

[45] J. D. Farmer, E. Ott, and J. A. Yorke, "The dimension of chaotic attractors," Physica D 7, 153–180 (1983).

[46] G. Mayer-Kress (Ed.), *Dimensions and Entropies in Chaotic Systems—Quantification of Complex Behavior* (Springer, Berlin, 1986).

[47] W. Lauterborn and J. Holzfuss, "Evidence of a low-dimensional strange attractor in acoustic turbulence," Phys. Lett. A 115, 369–372 (1986).

[48] H. G. E. Hentschel and I. Procaccia, "The infinite number of generalized dimensions of fractals and strange attractors," Physica D 8, 435 (1983).

[49] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman, "Fractal measures and their singularities: The characterization of strange sets," Phys. Rev. A 33, 1141–1151 (1986).

[50] D. Bensimon, M. H. Jensen, and L. P. Kadanoff, "Renormalization-group analysis of the global structure of the period-doubling attractor," Phys. Rev. A 33, 3622–3624 (1986).

[51] M. H. Jensen, L. P. Kadanoff, A. Libchaber, I. Procaccia, and J. Stavans, "Global universality at the onset of chaos: Results of a forced Rayleigh–Bénard experiment," Phys. Rev. Lett. 55, 2798–2801 (1985).

[52] J. A. Glazier, M. H. Jensen, A. Libchaber, and J. Stavans, "Structure of Arnold tongues and the $f(\alpha)$ spectrum for period doubling: Experimental results," Phys. Rev. A 34, 1621–1624 (1986).

[53] Z. Su, R. W. Rollins, and E. R. Hunt, "Measurement of $f(\alpha)$ spectra of attractors at transition to chaos in driven diode resonator systems," Phys. Rev. A 36, 3515–3517 (1987).

[54] J. A. Glazier, G. Gunaratne, and A. Libchaber, "$f(\alpha)$ curves: experimental results," Phys. Rev. A 37, 523–530 (1988).

[55] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L.-S. Young (Springer, Berlin, 1981), pp. 366–381.

[56] R. Mañé, "On the dimension of the compact invariant sets of certain nonlinear maps," in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L.-S. Young (Springer, Berlin, 1981), pp. 230–242.

[57] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," Phys. Rev. Lett. 45, 712–716 (1980).

[58] O. E. Roessler, "Chaotic oscillations: an example of hyperchaos," Lect. Appl. Math. 17, 141–156 (1979).

[59] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining

37

Lyapunov exponents from a time series," Physica D 16, 285–317 (1985).

[60]J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, and S. Ciliberto, "Liapunov exponents from time series," Phys. Rev. A 34, 4971–4979 (1986).

[61]M. Sano and Y. Sawada, "Measurement of the Lyapunov spectrum from a chaotic time series," Phys. Rev. Lett. 55, 1082–1085 (1985).

[62]I. Shimada and T. Nagashima, "A numerical approach to ergodic problems of dissipative dynamical systems," Progr. Theor. Phys. 61, 1605–1616 (1979).

[63]M. Henon, "Numerical exploration of hamiltonian systems," in *Chaotic Behavior of Deterministic Systems*, edited by G. Iooss, R. H. G. Helleman, and R. Stora (North-Holland, Amsterdam, 1983), pp. 53–170.

[64]M. Berry, "Semiclassical mechanisms of regular and irregular motion," in *Chaotic Behavior of Deterministic Systems*, edited by G. Iooss, R. H. Helleman, and R. Stora (North-Holland, Amsterdam, 1983), pp. 171–271.

[65]M. H. Jensen, P. Bak, and T. Bohr, "Transition to chaos by interaction of resonances in dissipative systems. I. Circle maps," Phys. Rev. A 30, 1960–1969 (1984).

[66]I. Procaccia, S. Thomae, and C. Tresser, "First-return maps as a unified renormalization scheme for dynamical systems," Phys. Rev. A 35, 1884–1900 (1987).

[67]U. Parlitz and W. Lauterborn, "Resonances and torsion numbers of driven dissipative nonlinear oscillators," Z. Naturforsch. 41a, 605–614 (1986).

1993   J. Acoust. Soc. Am., Vol. 84, No. 6, December 1988   W. Lauterborn and U. Parlitz: Chaos acoustics   1993

38

EXPERIMENTAL NONLINEAR PHYSICS

WERNER LAUTERBORN, THOMAS KURZ and ULRICH PARLITZ
Drittes Physikalisches Institut,
Universität Göttingen,
D-37073 Göttingen, Germany

The review gives and account of the historical development, the current state and possible future developments of experimental nonlinear physics, with emphasis on acoustics, hydrodynamics and optics. The concepts of nonlinear time-series analysis which are the basis of the analysis of experimental outcomes from nonlinear systems are explained and recent developments pertaining to such different fields as modeling, prediction, nonlinear noise reduction, detecting determinism, synchronization, and spatio-temporal time series are surveyed. An overview is given of experiments on acoustic cavitation, a field rich of nonlinear phenomena such as nonlinear oscillations, chaotic dynamics and structure formation, and one of the first physical systems to exhibit period-doubling and chaos in experiment.

1. Introduction

In his essay "Tractatus logico-philosophicus", Ludwig Wittgenstein writes:

"The world is all that is the case".

A physicist will easily agree but may feel unhappy about the amount of information given. But are there better solutions? Yes, there are. For instance, when a physicist at the turn of the century (1900) would have been asked, in view of the unification of mechanics and thermodynamics, the answer may have been:

"The world is statistical".

Ten years later, around 1910, impressed by the results of Einstein, the same physicist may have answered:

"The world is relative".

Twenty years later, around 1930, when looking at the achievements of quantum mechanics, the still living physicist may have added:

"The world is quantal".

What to add today, near the turn of the next century? Is there something left in fundamental science to be discovered besides statistics, relativity, and quantality? If not, this would be strange and even disappointing. To be sure, physics has not yet come to an end, and we see the next fundamental step in science prospering, that is nonlinearity. Nonlinearity is fundamental for all processes in nature, except perhaps quantum physics [Heisenberg, 1967]. However, for a long time it has not been recognized as a fundamental fact. Why? It seems to have been the general opinion that no universal laws may be found for nonlinear phenomena, leaving them as a set of individual, independent problems. Then nothing can be learned for a specific nonlinear problem from solving the other ones. This situation has changed drastically in the last decade. Nonlinearity now is recognized as being fundamental in almost any area of physics, notably hydrodynamics, optics, acoustics, and extends to chemistry, biology, ecology.

Thus, today, a physicist will add:

"The world is nonlinear",

and put this sentence on an equal footing with the three previous ones.

The basic hypothesis of a physicist (not long ago called philosopher) is that the natural phenomena surrounding us are governed by laws and that

a language can be invented to pin down these laws. We call that basic research, and it worked well so far for the benefit of mankind, although it has been a long way from ancient times to arrive at this insight. For instance, Descartes worked hard to give hints on how to proceed in science to make progress. But why do we want to understand nature? One reason is our intellectual curiosity. The second reason is that we can make use of the laws for better mastering our life, for instance by building houses, beating enemies with improved arms, moving faster than with our legs, transmitting information, etc. This works because with every law detected we gain the ability to predict the behavior of the respective system.

When we collect more and more laws of nature, will we, in the end, be able to predict the future? The answer is no, and the reason is nonlinearity. Even simple deterministic laws may have such complicated dynamics that a prediction over longer periods of time is impossible. Thus determinism does not imply (practical) predictability. And what do we do facing this insight? We now do experiments to gain insight into our limitations set by nonlinearity. This review is an attempt to at least partially follow the ways experimentalists have gone and will go further to elucidate the nature of unpredictability.

2. Historical Notes

The theory of nonlinear dynamics to a great extent has been developed in the context of celestial mechanics and started with the notion of a differential equation (see Diacu and Holmes [1996]). But celestial mechanics relies more on observation than actual experiments and therefore is not really suited for proving ideas and concepts of nonlinear physics in general. For instance, it does not comprise dissipative nonlinear systems with their attractors and self-organizing properties opening up a new and different realm of physics.

Looking at the history of physics over the past 300 years, say, there are quite a few occasions where experimental findings have preceded theoretical development. This is particularly true for nonlinear experimental physics where in certain areas typical nonlinear dynamical phenomena such as period-doubling, bifurcations or even chaos, structure formation, and self-organization have been observed early but could not be explained with the conceptions then available. Nowadays, this situation has changed remarkably. The numerical and analytical analysis of nonlinear models has evolved rapidly, not least due to the availability of ever more powerful computers. Numerous theoretical concepts have been developed to describe and evaluate nonlinear dynamical systems, involving sophisticated numerical algorithms. In fact, computing power starts to compete with experiments. However, as Poincaré — the father of nonlinear dynamics — puts it in his book *Science and Hypothesis* [1903]: "Experiment is the sole source of truth". Therefore the occupation with experiment should not be neglected and should be balanced with the advancement of theory.

Nonlinear effects in nature are most evident in the flow and oscillation phenomena of liquids and gases. In fact, the nonlinear properties of fluid flow leading to structure formation and turbulence are encountered in our daily life. The experimental areas of physics in the last century being mainly mechanics, electromagnetism and thermodynamics, it is understandable that mainly hydrodynamical and acoustic phenomena, accessible by the experimental means of that time, are among the first nonlinear effects to be discovered experimentally.

2.1. *Acoustics*

It seems that the oldest experiments in nonlinear physics stem from acoustics, in particular musical acoustics [Beyer, 1975]. Indeed, as the production of sound involves oscillations that can be generated only through nonlinear processes, all musical instruments (including the human voice) are devices of nonlinear dynamics, and devices on a scale for easily doing experiments with them. The first experiments and experiments at all are done to generate an understanding that cannot be obtained just by reasoning and imagination. One of the oldest nonlinear devices where such an understanding was sought is a musical instrument called Aeolian harp described in 1650 by Athanasius Kircher. In the Aeolian harp wind flowing past a string sets the string into vibrations as well as the air around it giving the sensation of a musical tone. Along the way of investigations done to understand this phenomenon we find the von Karman vortex street and the Reynolds number, a number describing the transition from laminar to turbulent flow. The simple string suspended in a flow of wind leads us to the problem of turbulence, a problem still to be solved despite the efforts of our greatest minds.

A further basically nonlinear phenomenon was first discovered in musical acoustics as reported by Sorge in 1745 and Tartini in 1754 (see Beyer [1975]). They found that two musical tones of high intensity cause the sensation of a lower tone being the difference in frequency of the two original ones (Fig. 1). Long debates arose how to explain the phenomenon: Is it simply the beats, does it originate in the course of propagation, or is it produced in the ear? Beats were excluded by Helmholtz [1870] on account of the sum frequency also being present. Helmholtz attributed his "combination frequencies" to the nonlinearities within the ear as we believe with him today. Even more combination frequencies can be heard and are actively investigated, in particular the lower cubic difference tone $(2f_1 - f_2$ in Fig. 1). However, the propagation medium itself can generate combination frequencies, no ear required [Thuras *et al.*, 1935]. Today, this is standard knowledge and is used in optics to frequency double light from einfrared to green, for instance, and even to produce white light from focused femtosecond pulses.

Systems that are periodically driven develop peculiar responses when the system is nonlinear. Most remarkable is the appearance of exactly half the driving frequency in the response, now a common observation announcing chaotic behavior. It was first noted in an acoustics experiment done by Faraday [1831]. Starting from the investigation of sound-emitting, vibrating surfaces by means of Chladni figures, Faraday also sprinkled water instead of sand onto his vibrating plates, and extended the work to complete layers of fluid to investigate the "beautifully crispated appearance" of the liquid layer. His aim was the "progress of acoustical philosophy". To be able to watch the motion of the fluid layer, he enlarged his vibrating plates to lower the frequency of oscillation, and ultimately came up

with a board eighteen feet long, upon which a liquid layer of three quarters of an inch in depth and twenty-eight inches by twenty inches in extent could be vibrated vertically. Then, by ordinary inspection, he could observe that the heaps of liquid making up the crispations were oscillating in a sloshing motion to and fro between neighboring heaps of liquid. He states: "Each heap (identified by its locality) recurs or is reformed in two complete vibrations of the sustaining surface" and adds in a footnote: "A vibration is here considered as the motion of the plate, from the time that it leaves its extreme position until it returns to it, and not the time of its return to the intermediate position". This result was confirmed by Rayleigh [1883a]. Today, the powerful methods of nonlinear dynamics and computerized experimental instrumentation are applied to this spatio-temporal system (see Bechhoefer *et al.* [1995] and references cited therein).

A second instance, where subharmonics were observed early, is the Melde experiment [Melde, 1860]. In this experiment, the tension of a string is modulated periodically by fastening it to the prong of a tuning fork. Under suitable conditions, the string will vibrate at half the driving frequency. Rayleigh [1883b, 1887] developed a theory for parametrically driven systems of this kind.

Acoustics has also supplied one of the first experiments at all showing chaotic dynamics. When sound of high intensity is passed through a liquid, the liquid may rupture, giving rise to the phenomenon of acoustic cavitation, whereby bubbles appear and are set into complicated motions. The analysis of this phenomenon has revealed that acoustic cavitation is a chaotic system [Lauterborn & Cramer, 1981; Smith *et al.*, 1982; Lauterborn & Holzfuss, 1986, 1991; Holzfuss & Lauterborn, 1989]. Moreover, it is a complex, spatio-temporal
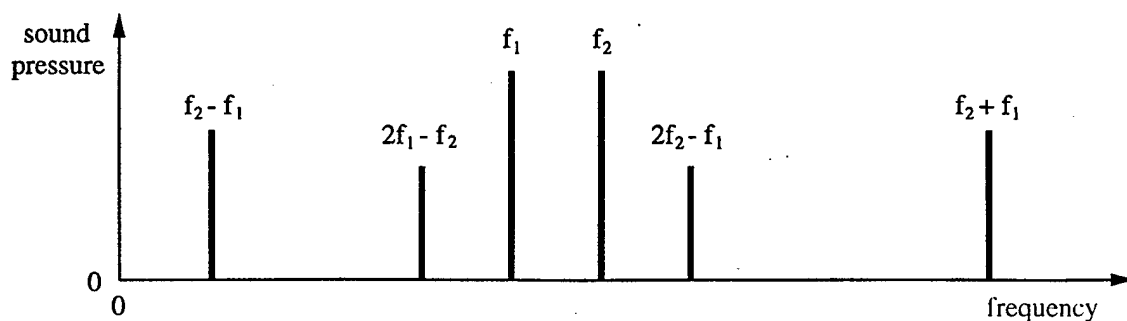

Fig. 1. Combination tones produced from two primary frequencies, f_1 and f_2.

system showing moving three-dimensional filaments. Acoustic cavitation appears, for instance, in underwater acoustics, when sonar systems are driven too hard, in medical applications, where kidney stones are destroyed by extracorporeal shock wave lithotripsy [Coleman et al., 1987; Philipp et al., 1993; Carnell et al., 1995], and in ultrasonic cleaning (see, e.g. [Young, 1989; Leighton, 1994]). This type of nonlinear dynamics originating from acoustic waves will be discussed in Sec. 4.

Sound may be generated by heat, a field covered by thermoacoustics [Swift, 1988, 1995]. Thermoacoustic oscillations come in two main varieties: Sondhauss oscillations [Sondhauss, 1850; Feldmann, 1968a] and Rijke oscillations [Rijke, 1859; Feldmann, 1968b]. The Sondhauss oscillation occurs when the closed end of a gas-filled pipe is heated (externally or internally), or, conversely, when the open end is cooled (then called Taconis oscillations [Taconis et al., 1949]). The Rijke oscillation occurs when an internal grid located in the lower half of a vertical pipe is heated. Both ends must be open to allow for a flow of gas (self-generated or enforced) through the pipe. Sound is also produced when the grid, this time being located in the upper half of the vertical pipe, is cooled [Riess, 1859].

Taconis oscillations with steep temperature gradients and large temperature difference show interesting dynamics [Yazaki et al., 1987]. The oscillation may period double, develop quasiperiodic oscillations through the appearance of a second incommensurate frequency and also chaotic oscillations. This behavior is attributed to mode competition similar to the Faraday experiment of Ciliberto and Gollub [1985]. When Taconis oscillations are confronted with periodic, externally imposed acoustic oscillations, the whole set of nonlinear dynamical phenomena seems to occur [Yazaki et al., 1990; Yazaki, 1993], as encountered, when self-excited systems are subject to periodic driving (see, e.g. the driven van der Pol oscillator [Parlitz & Lauterborn, 1987; Mettin et al., 1993]). These acoustic oscillations are not just of scientific interest, but have potential as "natural engines" [Wheatley & Cox, 1985]. The thermoacoustic oscillator may also be used as a musical instrument [Müller & Lauterborn, 1995].

In musical acoustics a variety of musical instruments [McIntyre et al., 1983] have the potential to be played with chaotic sound emission although this is usually avoided. Among them are string instruments [Tuffilaro, 1989; Müller & Lauterborn, 1996], woodwind instruments [Maganza et al., 1986; Gibiat, 1988; Idogawa et al., 1993] and gongs [Legge & Fletcher, 1989]. Also, nonlinear oscillators and their attractors may define new classes of sounds to compose with. Indeed, new musical instruments may be generated by systematically exploring nonlinear dynamical systems. For instance, Chua's circuit [Madan, 1993] with its large class of different attractors displays interesting features for composers [Mayer-Kress et al., 1993].

Chaotic dynamics has also been found in connection with the speech production process. To produce voice sounds the vocal folds are set into vibration through the air flow between them. Models of vocal folds are presently investigated to learn about their bifurcation and chaos properties [Herzel, 1993; Steinecke & Herzel, 1995], and bifurcations and chaos have been found experimentally in the cries of newborn infants [Mende et al., 1990]. Also hearing presents interesting physical examples of nonlinear dynamics (see [Lauterborn, 1996] for some information on this topic).

As an example for this acoustics section, the simple, bowed string is chosen. The string may exhibit a wealth of different motions when bowed at different speeds and pressure. Figure 2 shows time series of string velocities as measured with a laser vibrometer, corresponding embeddings and Poincaré section plots for four different types of motion. It is seen that period doubling occurs as well as quasiperiodic motion. Even chaotic oscillations may be observed [Müller & Lauterborn, 1996]. Those readers not yet familiar with the basic notions of chaos physics may consult some book or basic review ([Bergé et al., 1984; Guckenheimer & Holmes, 1983; Moon, 1992; Ott, 1993; Thompson & Stewart, 1986; Schuster, 1988; Lauterborn & Parlitz, 1988], to name just a few).

2.2. Hydrodynamics

At about the time of Faraday's investigation, another important branch of nonlinear physics was initiated in hydrodynamics: The first observation (and subsequent experimental investigation) of solitary waves by Scott Russell [1844, 1895]. He observed the undisturbed propagation over several miles of a single-humped elevation in a water channel, launched by the stopping motion of a boat. This gave the first experimental evidence of nonlinear wave solutions with particle-like character that we now call solitons. From this observation, and its
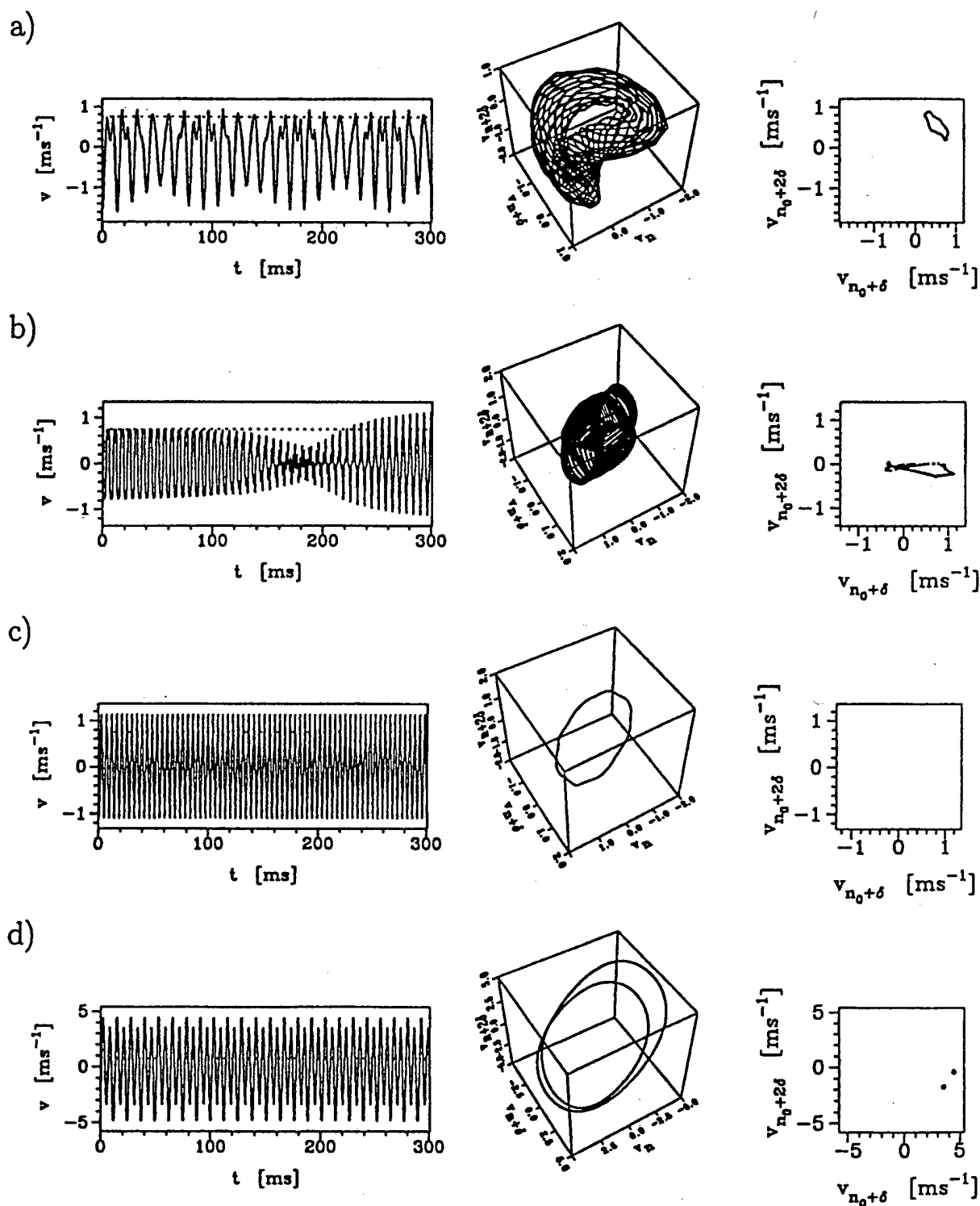
Fig. 2. Motion of a bowed string. Velocity versus time (left column), embedding into a three-dimensional space (center column), Poincaré section plot (right column). See [Müller & Lauterborn, 1996] for more details.

modeling by Korteweg and de Vries [1895] at the end of the century, it took several decades until finally, initiated by the numerical work of Zabusky and Kruskal [1965], an elaborate theory of soli-

tons and solitary wave solutions was established [Remoissenet, 1996].

Hydrodynamics with its complex spatiotemporal dynamics and its outstanding problem of

turbulence has ever been an active field of experimentation in nonlinear physics. It gained new thrust with the emergence of nonlinear dynamics and the availability of new experimental tools. The pioneering works of Lorenz [1963] and of Ruelle and Takens [1971] shed new light on the problem of turbulence in confined geometries. New theoretical concepts like bifurcations and strange attractors called for controlled experiments of hydrodynamical systems below or at the onset of turbulence. At the same time, new equipment such as Laser–Doppler anemometers enabled experimentalists to measure, e.g. the flow velocity at certain points or to control very exactly the critical parameters of an experiment. Thus, in the early 70's several experimentalists started to investigate hydrodynamical systems from the viewpoint of nonlinear dynamics. The most popular experiments were devoted to Rayleigh–Bénard convection and Taylor–Couette flow.
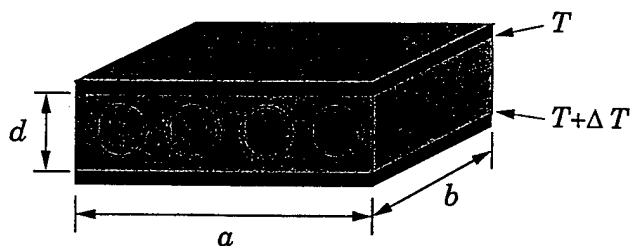

Fig. 3. Schematic of a Rayleigh–Bénard convection cell. The fluid between the two parallel plates is heated from below.

In a Rayleigh–Bénard system a fluid contained between parallel plates is heated from below (Fig. 3). For small heat input diffusive heat transport takes place, and after some time a constant temperature difference between both plates can be observed. If, however, the temperature gradient is increased, an instability occurs. Due to the heating the density of the medium near the

1. Description de l'appareil thermique à simple plateau. — Le premier appareil (!), destiné à produire

Fig. 2.

Échelle : ¼.
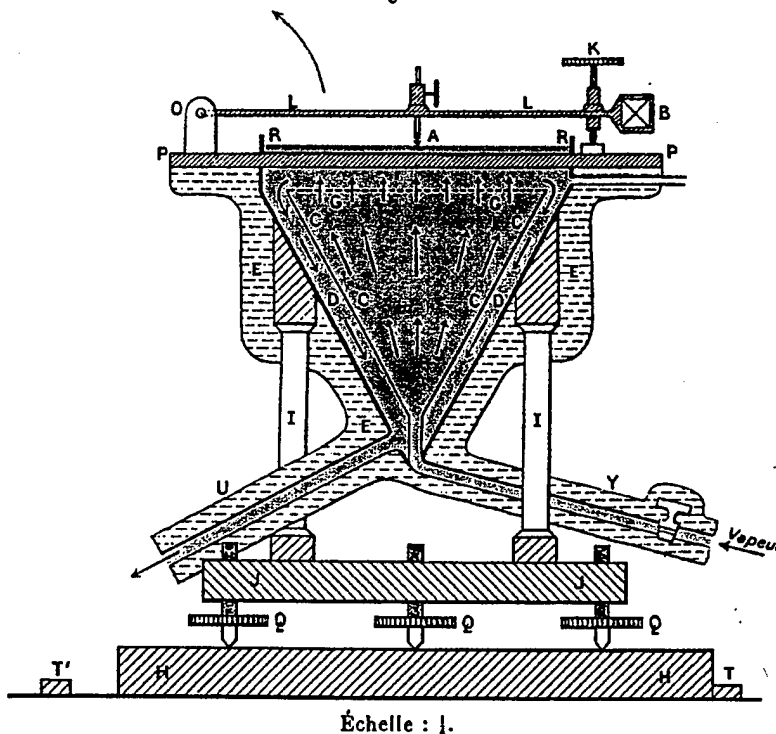
les tourbillons dans une nappe mince d'un liquide non volatil à 100°, comportait simplement (*fig.* 2).

Fig. 4. The apparatus used by Bénard [1901] in his studies on convection.

warmer plate decreases and the fluid moves upward, while the colder parts of the fluid move downward. This motion organizes itself as convection cells that provide a new heat transfer mechanism which is more effective than heat conduction. Such convection cells were observed experimentally for the first time by Henri Bénard [1901]. The first theoretical description was once again given by Lord Rayleigh. Figure 4 from the original paper of Bénard is reproduced here to give an appeal of the experimental techniques at the turn of the century. The Rayleigh–Bénard system was used by many experimentalists to investigate theoretical predictions in nonlinear dynamics. Among the first were Libchaber and Maurer [1980, 1982], who used liquid helium and who took the temperature gradient as a control parameter to verify the period-doubling route to chaos. Ahlers and Behringer [1978] studied the onset of turbulence in a cylindrical layer of fluid, Gollub and Benson [1980] found quasiperiodicity, intermittency and other nonlinear phenomena using a rectangular layer of water, and Swinney and Gollub [1978] observed a quasiperiodic route to chaos, later investigated with sophisticated methods from nonlinear dynamics by Glazier and Libchaber [1988]. A review on the topic of Rayleigh–Bénard convection has been given by Behringer [1985].

The second, well investigated type of fluid motion, Taylor–Couette flow, takes place between two concentric cylinders that rotate independently with angular velocities ω_i (inner cylinder) and ω_o (outer cylinder). In most of the experiments one of the cylinders (in general the outer cylinder) is fixed and does not rotate (Fig. 5). When the rotation frequency of the inner cylinder ω_i is increased, a hierarchy of instabilities is observed. First some toroidal rolls occur, that for larger values of ω_i start to oscillate periodically, then quasiperiodically and finally chaotically. This quasiperiodic route to chaos was observed for the first time in a Taylor–Couette system by Gollub and Swinney [1975].

Hydrodynamics and acoustics meet in the phenomenon of cavitation, that constitutes an experimentally very appealing but also complex arena for different kinds of nonlinear phenomena, from nonlinear oscillations to shock wave propagation, to structure formation, and to light generation on ultrashort time scales, called sonoluminescence (see, e.g. [Crum, 1994] for a readable introduction to the last subject). In particular, after the experimental achievement of Gaitan [Gaitan & Crum, 1990] to stably trap single, oscillating and light emitting bubbles in a liquid-filled resonator, the phenomenon
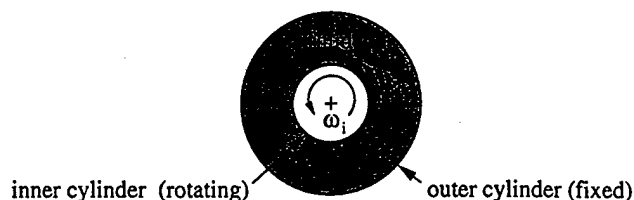

Fig. 5. Schematic of the Taylor–Couette experiment (top view). A fluid is contained between two concentric cylinders, where the outer cylinder is fixed and the inner cylinder is rotating at the angular velocity ω_i.

inner cylinder (rotating)   outer cylinder (fixed)

of single-bubble sonoluminescence has received much attention lately. Upon collapse the oscillating bubble generates very short light pulses of less than 100 ps duration which are spaced very regularly [Barber & Putterman, 1991; Barber *et al.*, 1992]. Under certain conditions, however, the sequence of pulse separations has been shown to display chaotic behavior [Holt *et al.*, 1994]. Many other experimental findings, for instance, the sensitive dependence of the light emission on parameters such as temperature or gas concentration of the liquid, add to the puzzle of this still poorly understood phenomenon.

Other areas, where hydrodynamics, acoustics and also electromagnetic theory come into play are magnetohydrodynamics and plasma physics, where, e.g. ion acoustic waves occur and where instabilities and complex motion abound. Due to their relevance in technological applications such as laser physics and inertial or laser fusion a large number of experiments have been performed, but an account of these experiments is beyond the scope of this article.

2.3. *Optics*

Besides in acoustics and hydrodynamics, nonlinear wave interactions underlie virtually all kinds of nonlinear phenomena in optics. In optics, nonlinearity is brought about by the interaction of photons with matter at high light intensity. As the nonlinear susceptibilities usually are very small, experimental nonlinear optics flourished only after the invention of the laser which provided the necessary light intensities.

Due to the smallness of their coefficients, nonlinear (nonresonant) optical interactions typically require a macroscopic interaction length or feedback by a resonator. If a nonlinear material is placed in a ring or Fabry-Perot type resonator, typically bistability or multistability of the transmission function is observed [Bowden, 1981; Gibbs, 1986].
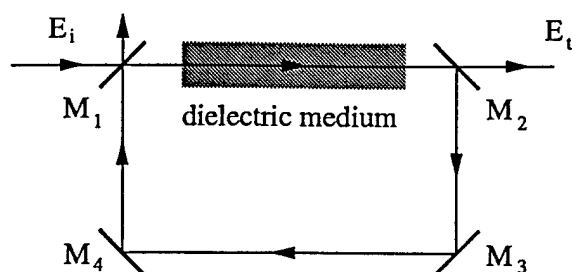
Fig. 6. Ring cavity of Ikeda with mirrors M_1 to M_4 (M_1 and M_2 semi-transparent) and with nonlinear dielectric medium. E_i is the incident field amplitude, E_t the transmitted amplitude.

Ikeda [1979] was the first to realize that the interaction between the time-delayed feedback signal and the incoming light in a nonlinear material placed in a ring resonator (Fig. 6) may lead to oscillations or to chaotic intensity fluctuations. Several experimental realizations followed his suggestion (e.g. [Gibbs et al., 1981]), confirming the theoretical predictions.

Compared to passive feedback systems, the laser itself constitutes a highly nonlinear dynamical system. With its instabilities and nontrivial dynamics it has become an excellent laboratory for experimental nonlinear science. In its simplest form (single-mode, homogeneously broadened, class B) the laser can be modeled by a two-dimensional system of rate equations. To achieve a non-trivial dynamics in this case, additional degrees of freedom have to be added such as periodic modulation of parameters (loss, pump), optical feedback, or excitation of several longitudinal or transversal modes. A derivation of the laser rate equations and examples of chaotic dynamics with periodic pump modulation can be found in [Lauterborn et al., 1995].

During the last two decades, a variety of aspects of nonlinear laser dynamics were studied in detail (see [Weiss & Vilaseca, 1991]). One of the first experiments to detect laser chaos was presented by Arecchi et al. [1982] who found period doubling and chaotic dynamics in a Q-modulated CO_2 laser. At about the same time, Weiss and King [1982] and Weiss et al. [1983] reported period doubling in a cw He-Ne laser, and Gioggia and Abraham [1983] described routes to chaos in an inhomogeneously broadened, single-mode xenon laser. Weiss et al. [1985] investigated an FIR ammonia laser system (class A) that may be described by the Lorenz–Haken model (which was derived from the laser equations in the bad cavity limit by Haken [1975]). A multitude of experiments followed, where semi-

conductor lasers gained increasing attention (e.g. [Mork et al., 1990]). They can easily be modulated and controlled, and arrays of semiconductor lasers integrated on a single chip open up new perspectives in the experimental investigation of coupled optical oscillators.

Lasers are physical systems where basic research on nonlinear dynamics bears on technological applications. For instance, for high-bit-rate communication as well as ultraprecise measurement applications laser instabilities and dynamical chaos are detrimental and have to be avoided. Here, it has been demonstrated that techniques of chaos control may be used to improve laser characteristics in a desired way [Roy et al., 1992].

As in future experiments more complex systems, e.g. those with a large number of degrees of freedom or spatially extended systems, will be addressed, the laser will continue to be one of the main subjects of experimental nonlinear science. By appropriate laser design it is easy to excite several longitudinal or transversal modes, leading to mode competition and chaos, transverse pattern formation, filamentation or optical turbulence [Abraham et al., 1987]. In all these issues of current interest the laser will be with us as an experimental testbed for some time to come.

In optics not only bounded geometries as in the laser are encountered, but usually wave propagation in extended passive media has to be considered. Then, nonlinear wave interaction effects, as are sum- and difference frequency generation and optical parametric amplification, take place and they were, in fact, investigated soon after the invention of the laser. With intense light, due to the interplay between dispersion and nonlinearity, new stable wave forms are possible, such as optical solitons in fibers. Predicted theoretically by Hasegawa and Tappert [1973] they were found experimentally in 1980 [Mollenauer et al., 1980], implemented in a soliton laser in 1984 [Mollenauer & Gordon, 1984] and nowadays enter technological application in fiber soliton transmission systems. As light pulses are getting shorter and more intense, the dynamics of solitons in waveguides, also in the form of spatial solitons in waveguides or light bullets in bulk materials, is receiving continued attention due to their potential for future all-optical information processing systems.

Experimental nonlinear optics has gained new thrust by the development of table-top high power femtosecond laser systems. This will promote research in highly nonlinear light-matter interactions as well as ultrashort dynamical processes in the
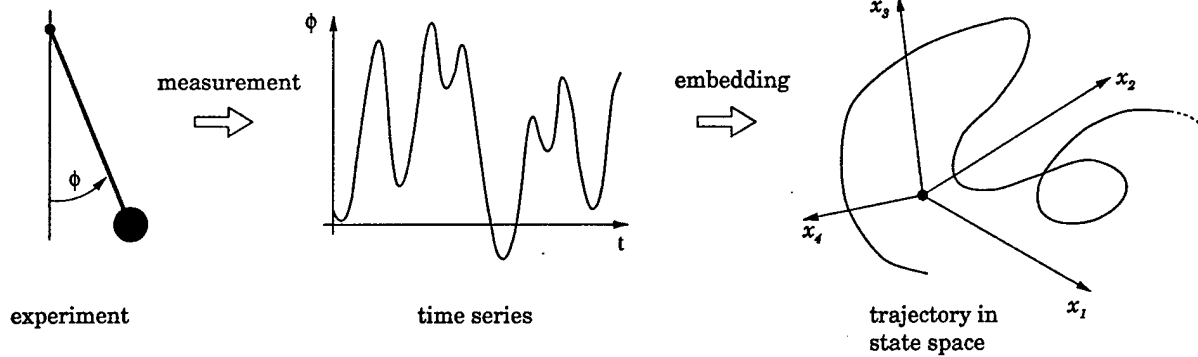
Fig. 7. The way from experiment to a trajectory in state space.

foreseeable future. Here, a wealth of nonlinear physical phenomena is waiting for their prospection; ranging from plasma–laser interaction at solid or liquid interfaces, to violent dielectric breakdown in liquids and gases, to the study of multiphoton-processes, i.e. of atoms in strong electromagnetic fields, and to table-top high energy particle accelerators, to name only a few.

In these historical notes, only a few milestones in the development of experimental nonlinear physics could be mentioned. There are more systems, for instance in mechanics, semiconductor and solid-state physics, microwaves, etc., that show irregular dynamics. Moreover, many other nonlinear systems, not only those in physics, were investigated experimentally, for example electronic oscillators, chemical reactions, physiological or biological systems. A complete review of all experiments in nonlinear science is beyond the scope of this article, and the reader is referred to the literature (e.g. [Gaponov-Grekhov & Rabinovich, 1992; Vohra *et al.*, 1992; Ditto *et al.*, 1994; Moon, 1992; Mullin, 1993; Thompson & Stewart, 1986; Holden, 1986; Madan, 1993]).

3. Nonlinear Time Series Analysis

In theory, i.e. in mathematical models of dynamical systems, the time evolution is described by a trajectory in state space whose integer dimension is given by the number of (dependent) variables of the model. In experiments, often just one quantity is measured as a function of time (a *time series*), and the state space usually is not known. How, then, to arrive at the attractor(s) that may characterize the system?

The gap between the theoretical notions and observable quantities was filled when Packard, Crutchfield, Farmer and Shaw [1980] published

their paper "Geometry from a time series" and Takens [1981] gave a mathematical justification of this approach. Takens proved that it is possible to (re)construct, from a scalar time series only, a new attractor that is diffeomorphically equivalent to the (in general unknown) attractor in the original state space of the experimental system (Fig. 7). This idea was essential for making tools of nonlinear dynamical systems theory amenable to the investigation of experimental data and thus marks the birth of deterministic nonlinear time series analysis [Grassberger *et al.*, 1991; Abarbanel *et al.*, 1993].

Essentially two methods for reconstructing the state space are available: *Delay coordinates* and *derivative coordinates*. Derivative coordinates were used by Packard *et al.* [1980] and consist of higher order derivatives of the measured time series. Since derivatives are susceptible to noise, derivative coordinates usually are not very useful for experimental data. Therefore we will discuss the method of delay coordinates only.

3.1. *Embedding*

Let M be a smooth (C^2) m–dimensional manifold that constitutes the (original) state space of the dynamical system under investigation and let $\phi^t : M \to M$ be the corresponding flow (see [Guckenheimer & Holmes, 1983] for a definition). Suppose that we can measure some scalar quantity $s(t) = h(\mathbf{x}(t))$ that is given by the measurement function $h : M \to \mathbb{R}$, where $\mathbf{x}(t) = \phi^t(\mathbf{x}(0))$. Then one may construct a *delay coordinates map*

$$F : M \to \mathbb{R}^d$$
$$\mathbf{x} \mapsto \mathbf{y} = F(\mathbf{x})$$
$$= (s(t),\, s(t - t_l),$$
$$s(t - 2t_l), \dots, s(t - (d-1)t_l) \tag{1}$$

47

that maps a state **x** from the original state space M to a point **y** in a *reconstructed state space* $\mathbb{R}^d$, where d is the *embedding dimension* and t_l gives the *delay time* (or *lag*) used. Takens [1981] proved that for $d \geq 2m + 1$ it is a generic property of F to be an embedding of M in $\mathbb{R}^d$, i.e. $F : M \to F(M) \subset \mathbb{R}^d$ is a (C^2-) diffeomorphism. Generic means that the subset of pairs (h, t_l) which yield an embedding is an open and dense subset in the set of all pairs (h, t_l). This theorem was generalized by Sauer *et al.* [1991] and Sauer and Yorke [1993] in two ways. Firstly, they replaced the condition $d \geq 2m + 1$ by $d > 2d_0(A)$ where $d_0(A)$ denotes the capacity (or box–counting) dimension of the attractor $A \subset M$. This is a great progress for experimental systems that possess a low-dimensional attractor (e.g. $d_0(A) < 5$) in a very high-dimensional space (e.g. $m = 100$). In this case, the theorem of Takens [1981] says that only for very large embedding dimensions d (e.g. $d \geq 201$) we can expect a diffeomorphic equivalence for sure, whereas with the condition of Sauer *et al.* a much smaller d will suffice (e.g. $d > 10$). The second improvement of Takens' embedding theorem by Sauer *et al.* refers to the term "generic" that was replaced by "prevalent" which essentially means "almost all (h, t_l)" will yield an embedding. This modification of the original theorem was necessary, because examples of open and dense (=generic) sets have been found that are rather "thin". Furthermore, Sauer *et al.* showed that for dimension estimation an embedding dimension $d > d_0(A)$ suffices. In this case the delay coordinates map F is, in general, not one-to-one, but the points where trajectories intersect are negligible for dimension calculations. More details about the reconstruction of states, in particular in the presence of noise, may be found in [Casdagli *et al.*, 1991; Gibson *et al.*, 1992]. The reconstruction from interspike intervals is discussed by Sauer [1994].

In the following we will assume that the scalar signal $s(t) = h(\mathbf{x}(t))$ is sampled with a sampling time t_s. The resulting time series $\{s^n\}$ with $s^n = s(nt_s)$ is used to reconstruct the states

$$\mathbf{y}^n = (s^n, s^{n-l}, s^{n-2l}, \ldots, s^{n-(d-1)l}) \qquad (2)$$

for $n = 1, \ldots, N$. The symbol l denotes the delay time or lag in units of the sampling time (i.e. $t_l = lt_s$).

3.2. *Estimating suitable reconstruction parameters*

The reconstruction using delay coordinates is based on two parameters: The embedding dimension d and the delay time t_l. When the dimension d is chosen too small the conditions given in the embedding theorems are not fulfilled. For d too large, on the other hand, practical problems occur due to the fixed amount of points, that constitute thinner and thinner sets in $\mathbb{R}^d$ when d is increased. Concerning the choice of the delay time t_l the theorems exclude only a few discrete (non–generic) values. Thus, different t_l lead to reconstructions of the attractor that are diffeomorphically equivalent. They are, however, geometrically different. When the delay time t_l is too small, the coordinates $y_i^n = s^{n-(i-1)l}$ of each reconstructed state $\mathbf{y}^n$ do not significantly differ from one another and therefore the state points are scattered along the diagonal [Fig. 8(a)]. In this case, any investigation of a possible fractal structure of the attractor (e.g. a dimension estimation) becomes difficult because one has to zoom into the cloud of points investigating scales that are (much) smaller than the diameter of the cloud perpendicular to the diagonal. On very small scales, however, such investigation becomes almost impossible due to the limited number of data points and the unavoidable presence of noise. The situation improves when t_l is increased. Then the attractor unfolds and its inner structure becomes "visible" already on larger scales as can be seen in [Fig. 8(b)]. When the delay time t_l is increased further, the reconstructed attractor is folded more and more [Fig. 8(c)]. The origin of this folding is the "stretch and fold" mechanism that is typical for chaotic systems with sensitive dependence on initial conditions. Additional folds in the reconstruction are unwanted because they bring states close together in the reconstructed state space that are not close together in the original state space. In particular in the presence of noise, where state points are additionally shifted from the "right places", this may lead to cases where states are wrongly addressed. Even without noise, a too large value of t_l may lead to a folding that yields intersections of the attractor where the reconstruction is not one-to-one anymore. In this case, one may increase the dimension d to "repair" the embedding, but of course it is more sensible to decrease t_l. These examples show that a proper choice of t_l is crucial for any further investigation of the data.
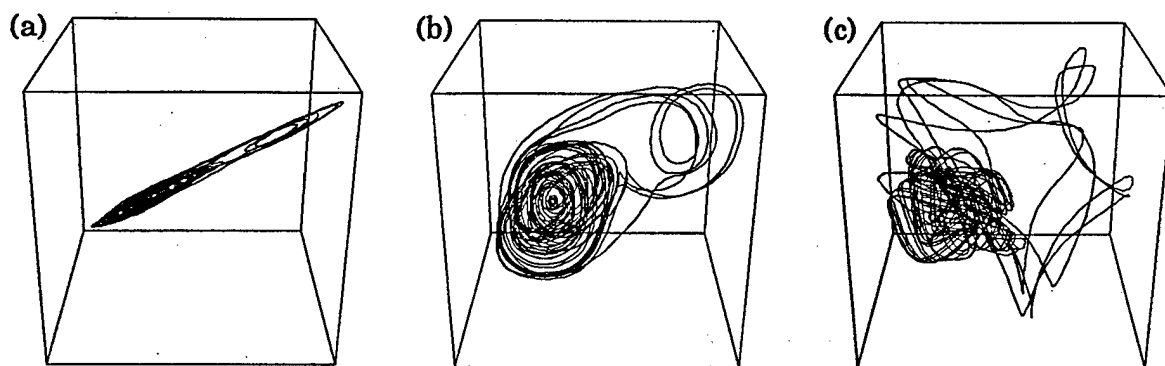
Fig. 8. Three reconstructions with different delay times t_l for experimental data obtained from a chaotically oscillating, periodically driven pendulum.

Several algorithms have been suggested to find "optimal" values for d and t_l. Most of the methods for determining d are based on continuity tests for the induced flow in the reconstruction space [Čenys & Pyragas, 1988; Buzug et al., 1990; Alecsić, 1991; Buzug & Pfister, 1992; Gao & Zheng, 1993, 1994; Huerta et al., 1995] or for the embedding itself [Liebert et al., 1991; Kennel et al., 1992]. The main idea is to check whether closely neighboring points are mapped to neighboring points. The methods differ with respect to their concrete implementation. To find optimal values for t_l one class of algorithms directly considers the geometry of the reconstructed attractor, in particular its unfolding from the diagonal [Kember & Fowler, 1993; Rosenstein et al., 1994]. The most popular approach consists in a minimization of the redundancy of the coordinates of the reconstructed states using (linear) autocorrelation functions or information theoretic concepts like mutual information [Frazer & Swinney, 1986; Frazer, 1989a, 1989b; Liebert & Schuster, 1989; Martinerie et al., 1992]. Since bad values for t_l may destroy the embedding, the continuity tests for estimating d may also be used to determine t_l [Buzug et al., 1990; Buzug & Pfister, 1992; Gao & Zheng, 1993, 1994; Liebert et al., 1991].

Until now, no method for determining optimal embedding parameters seems to be accepted as being the most useful. The reason for this diversity is probably the fact that there exists no pair (d, t_l) that is optimal for all possible subsequent investigations and applications. Furthermore, many of the algorithms proposed need as much computation time as a direct computation of a quantity of interest, e.g. the dimension. In this case, one may repeat the computation for several values of d and t_l and thus obtain not only "optimal" reconstruction parameters but also a test for the robustness of the method used and a first (quantitative) information about the data. As a starting point one may use t_l-values for which the autocorrelation function drops to zero or has a local minimum or, where neither of them exists, where it has dropped to $1/e$.

Finally, we want to mention the Broomhead–King-coordinates [Broomhead & King, 1986; Landa & Rosenblum, 1991] that result in a new coordinate system where the origin is shifted to the center of mass of the reconstructed states and the axes are given by the (dominant) principal components of the distribution of points. This new coordinate system is based on a Karhunen–Loève transformation that may be computed by a singular-value decomposition. A discussion of the advantages (e.g. noise reduction) and disadvantages of this "post-processing" of the reconstructed states may, for example, be found in Palus and Dvorak [1992].

3.3. Modeling and prediction

Once the states of the underlying dynamical system have been reconstructed it is possible to study their temporal evolution. Via the delay coordinates map F the flow ϕ^t in the original state space induces a flow $\psi^t = F^{-1} \cdot \phi^t \cdot F$ in the reconstructed state space. To model this induced flow different methods have been devised. An important class are local approximations where only states in a neighborhood of a given reference point are used to compute the parameters of a locally valid model. Possible models are, for example, polynomials (including locally linear approximations) [Farmer & Sidorowich, 1987] or radial basis functions [Casdagli, 1989]. Of course, all methods for modeling the induced flow locally can in principle also be used for a global ansatz. In general, local methods are more flexible and often yield better results when used for prediction of the time series. Global methods, on the other hand,

provide model equations in "closed form" that may yield a better understanding of the underlying dynamics. When the time series originates from a continuous dynamical system and a global model is desired one should try to fit a vector field (differential equation) to the dynamics given by the reconstructed states. More details concerning different modeling techniques, typical problems and an interesting method for "verifying" a derived model using synchronization may be found in [Brown *et al.*, 1994]. The problem of parameter estimation from time series in the case that the model equations are formulated not in the reconstructed state space but in the "original state space", will be addressed in Sec. 3.8. In any case, a successful approximation of the induced flow may not only be used for prediction but also provides a strong evidence for the deterministic nature of the data. Furthermore, good (local) models are a prerequisite for many other methods of nonlinear time series analysis and nonlinear dynamics.

3.4. *Fractal dimensions*

The (fractal) dimension of an attractor characterizes its complexity and gives a lower bound for the number of equations or variables needed for modeling the underlying dynamical process. From the infinite family of (generalized) dimensions D_q the *correlation dimension D_2* introduced by Grassberger and Procaccia [1983], Grassberger *et al.* [1991] is mostly used in nonlinear time series analysis. This dimension is given by the scaling

$$C_d(r) \propto r^{D_2} \qquad (3)$$

of the *correlation sum*

$$C_d(r) = \frac{2}{(N - c)(N - c + 1)}$$

$$\times \sum_{i=1}^{N} \sum_{j=1}^{i-c} H(r - \|\mathbf{y}^i - \mathbf{y}^j\|) \qquad (4)$$

that counts the relative number of neighboring points closer than r (H is the Heaviside function with $H(x) = 1$ for $x > 0$ and zero elsewhere). The constant c is some correlation length and is used to omit points that are close neighbors in time [Theiler, 1986], and d is the embedding dimension. The correlation sum approach is a *fixed size* method, because for a given radius r the number of pairs with distance smaller than r is counted. As an alternative one may also use a *fixed mass* method to estimate the dimension of the attractor as was pointed

out by Badii and Politi [1984, 1985] and Grassberger [1985]. In this case the m nearest neighbors of each reference point $\mathbf{y}^n$ and the radius $r^n = r(m)$ of this cloud of m points are determined. For the limit $m/N \to 0$ one obtains an approximation of the

capacity dimension $\quad D_0 \approx - \dfrac{\log N}{\log \dfrac{1}{N} \sum\limits_{n=1}^{N} r^n}$ (5)

and an estimate of the

information dimension $\quad D_1 = - \dfrac{\log N}{\dfrac{1}{N} \sum\limits_{n=1}^{N} \log r^n}$ (6)

To investigate the scaling in the limit $m/N \to 0$ one can decrease the number of neighbors m *or* increase the number of data points N.

Figure 10 gives an example of the fixed size method, where the correlation sum $C_3(r)$ of (4) is calculated and plotted versus the radius r of the ball of neighboring points on a doubly logarithmic scale for the Duffing data of Fig. 9. The slope in a suitable intermediate range of r then gives the fractal dimension of the chaotic Duffing attractor (in the Poincaré section plane). The dashed line in Fig. 10 is a fit to the slope (shown with an offset) and gives the correlation dimension $D_2 = 1.1$ in the Poincaré plane, i.e. a fractal dimension of 2.1 for the continuous system. To detect whether the data are distinct from noise, surrogate data have been derived from the original data set according to the method of random Fourier phases as discussed
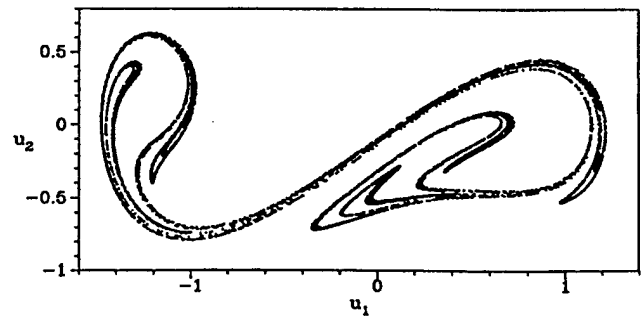

Fig. 9. A strange attractor of the double-well Duffing oscillator $\ddot{x} + d\dot{x} - x + x^3 = a \sin \omega t$ for $d = 0.2$, $a = 0.3$, and $\omega = 1.24$ in a Poincaré section plane.
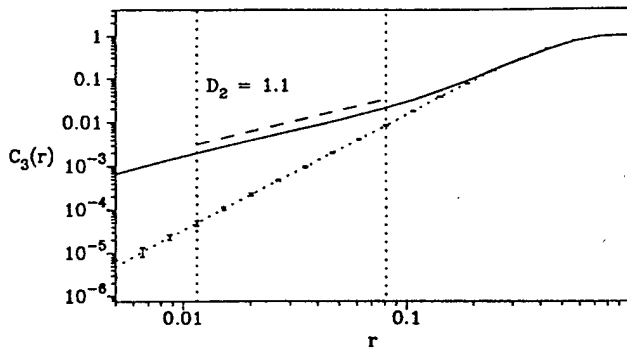
Fig. 10. Correlation sum $C_3(r)$ versus the distance r of neighboring points for the Duffing data of Fig. 7. The dashed curve is a fit to the slope giving $D_2 = 1.1$. The dotted curve is from surrogate data to detect determinism.

below in Sec. 3.7. Their correlation sum $C_3(r)$ is included in the figure as dotted curve. It is strongly distinct from the curve for the original data indicating a clear difference between the chaotic data and colored noise. The error bars correspond to ±5 standard deviations.

A very suitable method is the determination of the *pointwise dimensions* and their average [Holzfuss & Mayer-Kress, 1986; Mayer-Kress, 1986]. The pointwise dimension D at a point of an attractor can be calculated by exploring the scaling behavior of the "mass" of the attractor in the vicinity of the point, i.e. the number of points within a certain distance r. The number $N(r)$ of points in this ball of radius r should scale like $N(r) \propto r^D$ for $r \to 0$. When this is done for all points of the attractor, a histogram of pointwise dimensions is obtained. The average dimension and the standard deviation then are numbers to characterize the attractor. Of course, the histogram itself is a characteristic of the attractor.

When the attractor is sufficiently low-dimensional, its (fractal) dimension starting from low-dimensional embedding spaces ($d = 1, 2, 3, \ldots$) will level off at some value of d and does not increase further. This value then is said to determine the relevant number of (nonlinear) degrees of freedom, i.e. the number of dependent variables, of the dynamical system investigated. More details about dimension estimation methods, their possible pitfalls, extensions, and many references are given in the review articles by Grassberger *et al.* [1991], Theiler [1990], and Broggi [1988].

3.5. *Lyapunov spectra*

The Lyapunov spectrum is a set of real numbers $\{\lambda_1, \lambda_2, \ldots, \lambda_m\}$, the Lyapunov exponents. Lya-

punov exponents describe the mean exponential increase or decrease of small perturbations on an attractor and are invariant with respect to diffeomorphic changes of the coordinate system. When the largest Lyapunov exponent is positive, the system is said to be chaotic and to show sensitive dependence on initial conditions. That way, the notion of chaos can be quantified. For an exact definition and computational details see, for example, [Oseledec, 1968; Benettin *et al.*, 1980; Shimada & Nagashima, 1979; Eckmann & Ruelle, 1985; Geist *et al.*, 1990; Wolf *et al.*, 1985].

The different methods for computing Lyapunov exponents from time series that have been proposed so far can be divided into two classes: *Jacobian-based methods* and *direct methods*. In the *Jacobian methods*, a model is first fitted to the data and the Jacobian matrices of the model equations are then used to compute the Lyapunov exponents using standard algorithms which have been developed for the case when the equations of the dynamical system are known [Benettin *et al.*, 1980; Shimada & Nagashima, 1979; Eckmann & Ruelle, 1985; Geist *et al.*, 1990]. Usually local linear approximations are used [Sano & Sawada, 1985; Eckmann *et al.*, 1986; Stoop & Meier, 1988; Holzfuss & Lauterborn, 1989; Stoop & Parisi, 1991; Zeng *et al.*, 1992; Parlitz, 1993]. Modifications and improvements of the basic algorithm include:

- using neighboring states from a shell ($r_{min} < r < r_{max}$) instead of a ball ($r < r_{max}$) to reduce the influence of noise (see, e.g. [Eckmann *et al.*, 1986; Zeng *et al.*, 1991, 1992])
- using nonlinear approximations of the flow by polynomials [Briggs, 1990; Bryant *et al.*, 1990; Brown *et al.*, 1991; Abarbanel *et al.*, 1991], radial basis functions [Holzfuss & Parlitz, 1991; Parlitz, 1992; Kadtke *et al.*, 1993] or neural networks [Gencay & Dechert, 1992]
- using projections onto linear subspaces based on local Broomhead–King coordinates [Stoop & Parisi, 1991]
- using the embedding dimension d for selecting the neighbors and a small *local dimension* d_L for performing the fit and computing the resulting d_L Lyapunov exponents (see, e.g. [Eckmann *et al.*, 1986; Bryant *et al.*, 1990; Brown *et al.*, 1991]).

The latter two methods have been devised to avoid so-called *spurious Lyapunov exponents*, that occur when the dimension d of the reconstructed state space is larger than the dimension m of the original state space. Spurious Lyapunov exponents do

not depend on the dynamics but on the approximation scheme, the noise contained in the data and in some cases on the local curvature of the attractor. In principle they can take any values. (see [Čenys, 1993] for some nice examples). To identify spurious Lyapunov exponents one can estimate the local thickness of the attractor along the directions associated with the different Lyapunov exponents [Bryant *et al.*, 1990; Brown *et al.*, 1991] or, for high quality data, compare the exponents obtained with those computed for the time reversed series [Parlitz, 1992], because spurious exponents correspond to directions where the attractor is very thin and because they do not change their signs upon time reversal (in contrast to the true exponents). An investigation of the data requirements for Jacobian-based methods may be found in [Eckmann & Ruelle, 1992; Ellner *et al.*, 1991].

Direct methods directly estimate the divergent motion of the reconstructed states without fitting a model to the data. The best known and widely used method was introduced by Wolf *et al.* [1985]. There, a reference orbit $\{\mathbf{y}^n\}$ is considered and a neighboring orbit $\{\mathbf{z}^n\}$ starting from the nearest neighbor $\mathbf{z}^0$ of $\mathbf{y}^0$. After some transient the difference vector $\mathbf{u}^k = \mathbf{z}^{n+k} - \mathbf{y}^{n+k}$ points into the (time dependent) direction corresponding to the largest Lyapunov exponent λ_1. Then $\|\mathbf{u}^k\|$ grows, on the average, exponentially with $\exp(\lambda_1 t_s k)$ until it exceeds the range where a linear approximation of the flow at $\mathbf{y}^n$ holds. Now the neighboring orbit $\{\mathbf{z}^n\}$ has to be replaced by another neighboring orbit that is closer to the reference orbit *and* whose initial value lies on or near the line from the current reference state to the last point of the previous neighboring orbit in order to preserve the direction corresponding to the largest Lyapunov exponent. Criteria for the replacement threshold and other details of the algorithm are given in [Wolf *et al.*, 1985], including a FORTRAN program. In principle, it is possible to use this strategy also for computing the second largest Lyapunov exponent, but this turns out to be difficult. Thus it is quite fortunate that in many cases it suffices to establish the existence of at least one positive Lyapunov exponent. Due to its robustness the Wolf algorithm is often used for the analysis of experimental data (see, for example, [Fell *et al.*, 1993; Fell & Beckmann, 1994]).

Another direct method, which is even simpler because of fewer free parameters, was proposed by Sato *et al.* [1987] and Kurths and Herzel [1987].

Very similar to the Wolf algorithm, the average exponential growth of the distance of neighboring orbits is studied on a logarithmic scale, this time via the prediction error

$$p(k) = \frac{1}{Nt_s} \sum_{n=1}^{N} \log_2 \left(\frac{\|\mathbf{y}^{n+k} - \mathbf{y}^{nn+k}\|}{\|\mathbf{y}^n - \mathbf{y}^{nn}\|} \right), \quad (7)$$

where $\mathbf{y}^{nn}$ is the nearest neighbor of $\mathbf{y}^n$. The dependence of the prediction error $p(k)$ on the number of time steps k may be divided into three phases. Phase I is the transient where the neighboring orbit converges to the direction corresponding to the largest Lyapunov exponent. During phase II the distance growths exponentially with $\exp(\lambda_1 t_s k)$ until it exceeds the range of validity of the linear approximation of the flow around the reference orbit $\{\mathbf{y}^{n+k}\}$. Then phase III begins where the distance increases slower than exponentially until it decreases again due to foldings in the state space. If phase II is sufficiently long, a linear segment with slope λ_1 appears in the $p(k)$ versus k diagram. This not only allows an estimation of the largest Lyapunov exponent λ_1 but also provides a direct verification of exponential growth of distances to distinguish deterministic chaos from stochastic processes where a *non*-exponential separation of trajectories occurs [Dämmig & Mitschke, 1993]. Figure 11 gives an example for the determination of the largest Lyapunov exponent λ_1 by this method. The dotted curve is from surrogate data to detect determinism. The error bars are for ± 5 standard deviations.

Several authors have investigated modifications of this basic algorithm [Gao & Zheng, 1994; Rosenstein *et al.*, 1993; Kantz, 1994]. In particular, those states must be excluded from the search
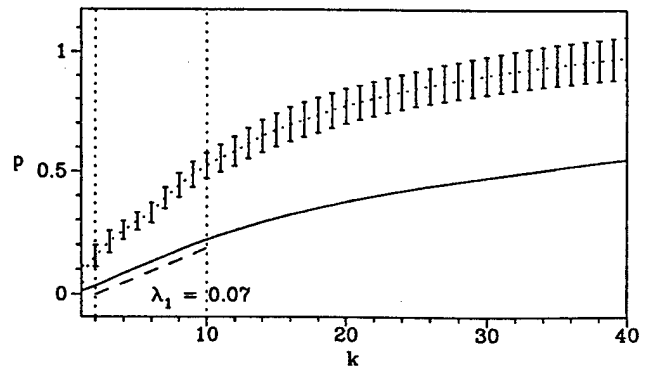

Fig. 11. Prediction error p for experimental pendulum data in a Poincaré section plane versus the time step number k. The slope of the solid line in the intermediate range of k gives the largest Lyapunov exponent λ_1.

for the nearest neighbor that are close together in time. Such states belong to the same segment of the trajectory and thus *cannot* converge to the most expanding direction. A possible criterion for a sufficiently large separation in time is the first minimum or zero of the autocorrelation function of the data set.

3.6. *Noise reduction*

Experimental data usually are contaminated with noise. To improve noisy data two possibilities are available: Linear filters and special nonlinear noise reduction methods that make use of the deterministic origin of the data. The influence of linear filters on the reconstruction of states was investigated by Sauer *et al.* [1991], Sauer and Yorke [1993], and by Broomhead *et al.* [1992]. Infinite impulse response (IIR) filters in general destroy the diffeomorphic equivalence between the original state space and the reconstructed state space. For finite impulse response (FIR) filters one can show that those filters which destroy the embedding constitute linear subspaces in the space of all FIR-filters. Therefore, with probability one, FIR-filters preserve the embedding. Nevertheless, one may find examples where the application of a FIR-filter makes subsequent estimations of Lyapunov exponents or dimensions more difficult as compared with the original "noisy" time series. Therefore, before filtering the data, one should always try to investigate the raw data. If it is known that the data originate from a deterministic system, methods for nonlinear noise reduction may be applied that have been developed only recently. The basic idea is to exploit the underlying determinism by correcting the states and the values of the time series using a (local or global) model that has previously been fitted to the (raw) data. More details about these applications of nonlinear time series analysis for noise reduction can be found in [Grassberger *et al.*, 1993; Kantz *et al.*, 1993; Kostelich & Schreiber, 1993].

3.7. *Detecting determinism*

If the time series originates from a yet unknown process it is important to investigate the question of whether the data contain some nonlinear deterministic dependencies or whether it is a purely stochastic signal, for instance, colored noise. Several methods have been devised to answer this question [Theiler *et al.*, 1992a, 1992b; Provenzale *et al.*, 1992; Smith, 1992; Takens, 1993; Wayland *et al.*, 1993; Palus *et al.*, 1993; Kaplan, 1994; Salvino

et al., 1994; Savit & Green, 1991]. Here, we want to focus on the *method of surrogate data* [Theiler *et al.*, 1992a, 1992b] that provides a rather general and flexible framework for investigating unknown data. The basic idea is to make some hypothesis about the data and then to try to falsify this hypothesis. As an example we will use the widely used hypothesis: "The data are (just) colored noise". In the first step, the data are modified in such a way that all structure except for the assumed properties will be destroyed or "randomized". This may be done, for example, by Fourier transforming the data and substituting the phases of the complex Fourier transform by random numbers. The power spectrum (or equivalently, the autocorrelation function) is not affected by this modification. After transforming back into the time domain we thus obtain a new time series with the same power spectrum. If the original data are just colored noise we have not destroyed any underlying structure, and subsequent estimations of dimensions, Lyapunov exponents, prediction errors, etc. should give the same results for the original time series and the surrogate data. If, however, the analysis yields significant differences, then our original time series was more than "just noise". To improve the statistical robustness of this test, one usually generates several surrogate data sets and compares the mean values of the quantities obtained in the course of the subsequent data analysis with the corresponding value from the original time series. This has been done in the above examples, where the fractal dimension and the largest Lyapunov exponent were calculated for chaotic Duffing oscillator data and for experimental data from a driven pendulum, respectively. The surrogate data show strong deviations, indicating that the time series investigated come from a deterministic, chaotic system. Furthermore, it is necessary to take into account a possibly performed static nonlinear transformation of the data that would distort the Gaussian distribution of the assumed colored noise (Algorithm II in [Theiler *et al.*, 1992a], see also [Rapp *et al.*, 1994]).

3.8. *Synchronization of chaotic dynamics*

Synchronization of periodic signals is a well-known phenomenon in physics, engineering and many other scientific disciplines. However, even chaotic systems may be linked in a way such that their chaotic oscillations are synchronized. In particular,

the case of one-directional coupling has been investigated intensely during the last years [Fujisaka & Yamada, 1983; Afraimovich *et al.*, 1986; Pecora & Carroll, 1990; Carroll & Pecora, 1993; Kocarev & Parlitz, 1995; Lai & Grebogi, 1994; Kocarev & Parlitz, 1996a] because of its potential application in communication systems [Cuomo & Oppenheim, 1993; Kocarev *et al.*, 1992; Parlitz *et al.*, 1992; Wu & Chua, 1993; Halle *et al.*, 1993; Schweizer & Kennedy, 1995]. There, an information signal containing a message is transmitted using a chaotic signal as a broadband carrier, and the synchronization is necessary to recover the information at the receiver. Different implementations of this basic idea have been suggested, and are currently investigated for their practical usefulnes (see, for example, [Feldmann *et al.*, 1996] or [Parlitz *et al.*, 1996a] and references cited therein).

Another application of synchronization consists in parameter estimations from time series. Assume that a (chaotic) experimental time series has been measured and the structure but not the parameter values of a model describing the underlying process are known. The goal is to find these unknown parameters and perhaps also the time evolution of the variables that have not been measured. This problem can for example be tackled using multiple shooting methods [Baake *et al.*, 1992] or similar approaches [Breeden & Hübler, 1990]. However, with all these methods, not only the parameters occur in the algorithm as unknown quantities, but also the initial values of the trajectory segments between the sampling times. One therefore has to solve a high dimensional minimization or fixed-point problem. This can be simplified considerably if a synchronization mechanism is used that automatically yields the right state variables. Hence, using synchronization, the dimension of the approximation problem is given just by the number of unknown model parameters.

The basic strategy consists in finding the right parameter values by minimizing the synchronization error of a numerical model driven by the given data [Brown *et al.*, 1994]. As an example, we consider Chua's circuit, which can be described by the following three-dimensional dynamical system:

$$C_1 \frac{dV_{C1}}{dt} = \frac{V_{C2} - V_{C1}}{R - g(V_{C1})}$$

$$C_2 \frac{dV_{C2}}{dt} = \frac{V_{C1} - V_{C2}}{R + I_L} \qquad (8)$$

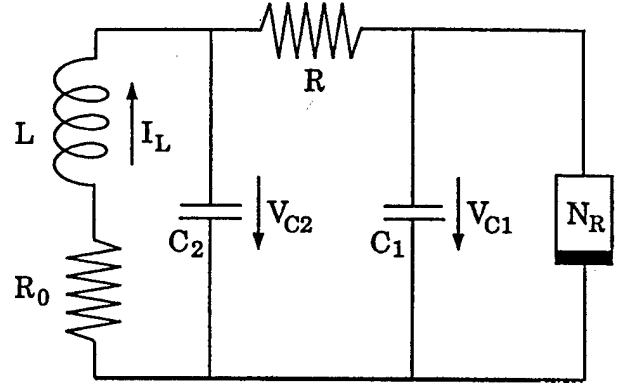$$L \frac{dI_L}{dt} = -V_{C2} - R_0 I_L,$$

Fig. 12. Circuit diagram of the Chua's circuit (8). The nonlinear resistor (Chua's diode) is given by (9).

where g is a piecewise-linear function defined by

$$g(V) = m_0 V$$
$$+ \frac{1}{2}(m_1 - m_0) \left[|V + B_p| - |V - B_p| \right] . \quad (9)$$

The circuit diagram is given in Fig. 12. Our goal is to estimate the parameter values of a hardware implementation of the circuit from a time series of the voltage V_{C1} sampled with 40 kHz. This time series (length 10 000 samples) was used to drive a numerically simulated model of Chua's circuit. The driving consisted in a repeated replacement of the variable V_{C1} in the model by the times series values at times $t_n = nT_s$, where $T_s = 0.025$ ms is the sampling time. Note that the coupling is not applied continuously, but only at discrete times t_n. Between this *sporadic coupling* the numerically simulated circuit oscillates freely and independently from the driving. This type of time discrete coupling between continuous systems may lead to synchronization if the coupling time T_s is suitably chosen [Amritkar & Gupte, 1993; Stojanovski *et al.*, 1996]. If the parameter values of the model coincide with those of the driving hardware circuit the synchronization error vanishes. In practice it possesses a clear minimum larger than zero due to the unevitable presence of noise. To locate this minimum in parameter space we used the routine POWELL from [Press *et al.*, 1986]. For the search of the smallest synchronization error all eight parameters in (8) and (9) have been varied. Additionally a possible offset b upon measurement has also been taken into account in the minimization process. The unknown amplification factor a of the A/D-converter that we used is contained in the breakpoint voltage B_p. Since this parametrization is redundant, the following normal-

54

ized parameters have been estimated:

$$p_1 = \frac{C_2}{C_1} \approx 10 \qquad p_2 = \frac{R^2 C_2}{L} \approx 16$$

$$p_3 = \frac{R R_0 C_2}{L} \approx 0.18 \quad p_4 = m_0 R \approx -0.69 \qquad (10)$$

$$p_5 = m_1 R \approx -1.28 \quad p_6 = \frac{1}{RC_2} \approx 5800$$

$$p_7 = aB_p \qquad p_8 = b,$$

where p_6 and p_7 give the time scaling and the amplitude scaling, respectively. The numerical values given are roughly estimated from the components used and are known up to a certain tolerance only. Figure 13 shows the convergence of the normalized parameters during the minimization process. The resulting saturation values are given in the diagrams and nicely match the values given in (10). Delay reconstructions using the experimental time series and the corresponding time series generated numerically with the estimated parameter values yield two almost indistinguishable attractors. These diagrams and further details of the algorithm may be found in [Parlitz *et al.*, 1996b]. There also numerically generated time series were used and yielded exact estimates of the parameters of the drive system. A possible application of this parameter estimation method are, for example, VLSI implementations of electronic systems like Chua's circuit where the determination of the actual values of the individual components is very difficult. In those cases where a continuous coupling between the experiment and a computer model is possible, similar methods may be applied [Parlitz, 1996; Parlitz & Kocarev, 1996].

The synchronization of two systems in the sense that both state vectors converge to the same value is not the only possibility. If, for example, a well-defined phase variable can be identified in both systems, the phenomenon of *phase synchronization* may occur [Rosenblum *et al.*, 1996; Parlitz *et al.*, 1996c]. In this case the difference between both phases is bounded and the average values of the resulting angular rotation frequencies are equal after some synchronization transient, while the corresponding amplitudes of both systems remain uncorrelated. This phenomenon may be used in technical or experimental applications where a coherent superposition of several output channels is desired.

Another concept of synchronization is the notion of *generalized synchronization* where the existence of a functional relationship between the
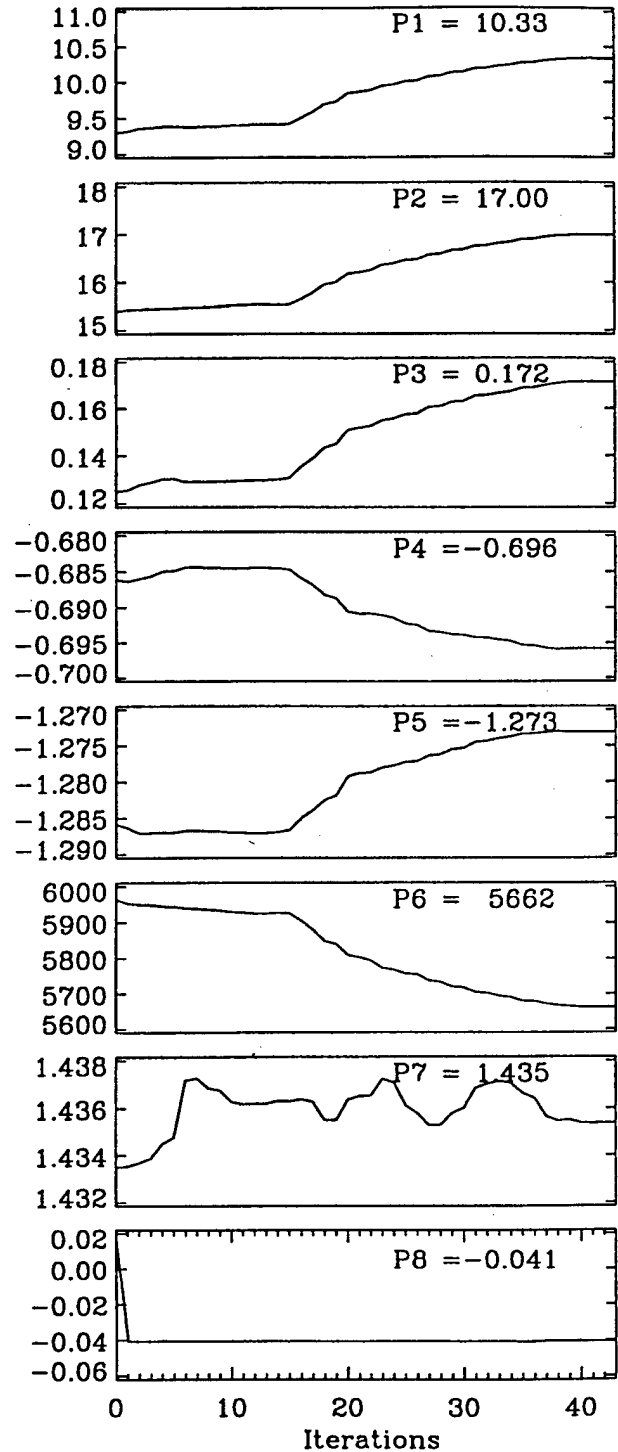

Fig. 13. Convergence of the normalized parameters in (10) during the minimization of the synchronization error.

states of the coupled systems is required (in the limit $t \to \infty$) [Afraimovich *et al.*, 1986; Rulkov *et al.*, 1995; Kocarev & Parlitz, 1996b]. This type of synchronization occurs for uni-directionally coupled systems when the driven system is asymptotically

stable [Kocarev & Parlitz, 1996b] and is therefore very robust with respect to parameter mismatch.

3.9. *Spatio-temporal data*

Methods for analyzing and predicting spatio-temporal time series (STTS) are currently one of the most important challenges in nonlinear time series analysis. In general a STTS consists of a sequence of (spatial) arrays $\mathbf{x}^n \in \mathbb{R}^M$ that are taken from a spatially extended dynamical system at discrete times $t_n = n\Delta t$ ($n = 1, \ldots, N$), where Δt is the sampling time and N is the length of the time series, i.e. the total number of patterns. Each frame $\mathbf{x}^n$ consists of M spatial samples x_m^n or "pixels" and may thus (at a fixed time n) formally be considered as a vector in $\mathbb{R}^M$.

The most successful approach developed until now to analyze a STTS $\{\mathbf{x}^n\}$ is based on a linear decomposition of the dynamics into spatial modes $\mathbf{b}^k$. The goal of this approach is to find a few dominating modes $\mathbf{b}^1, \ldots, \mathbf{b}^K$ ($K \ll M$) so that the patterns $\mathbf{y}^n = \mathbf{a} + \sum_{k=1}^K c_k^n \mathbf{b}^k$ provide a good approximation of the sequence $\{\mathbf{x}^n\}$. The vectors $\mathbf{a}$ and $\mathbf{b}^k$ are constant and may be computed using the *Karhunen–Loève transformation* (also known as *Proper Orthogonal Decomposition*, *Empirical Orthogonal Functions*, or *Singular Value Decomposition* (see, for example, [Lumley, 1970; Sirovich, 1987, 1989; Aubry et al., 1991]). The information about the temporal evolution is contained in the coefficients c_k^n. If such a decomposition into a few dominating modes is possible this method turns out to be very efficient. From the temporal evolution of the (few) coefficients $\{c_k^n\}$ ($k = 1, \ldots, K$), low dimensional models can be derived and may be used for analyzing and predicting the STTS [Ciliberto & Nicolaenko, 1991; Rico-Martinez et al., 1992; Chauve & Le Gal, 1992; Berry et al., 1994].

However, even for low dimensional systems, the number of dominating modes of a Karhunen–Loève transformation may be arbitrarily large. The dynamics of a delta pulse, for example, moving periodically back and forth on a one-dimensional axis, can only be covered by as many basis vectors as different positions of the pulse occur. Its closed orbit in state space is one-dimensional, but a large number of modes $\mathbf{b}^k$ is necessary to describe the STTS. This problem occurs in general when moving structures (e.g. solitons) appear in a spatially extended system. Another reason why the Karhunen–Loève transformation may not be the optimal choice

for processing spatio-temporal data from a dynamical system, is its purely probabilistic character. A low energy mode that is discarded when using this transformation may on the other hand be very important for the dynamical evolution. Therefore, alternatives are currently developed that try to avoid these problems [Kwasniok, 1996; Stone & Cutler, 1996].

For STTS whose dynamics originates from a low-dimensional chaotic attractor, but cannot be decomposed into a small number of spatial modes, other methods have been suggested that are based on a time-dependent activation of spatial patterns (or modes) using a low-dimensional state space reconstruction from a suitably chosen global observable of the system [Parlitz & Mayer-Kress, 1994].

Another interesting question when dealing with spatio-temporal systems concerns the (decay of the) correlation of probes taken at different spatial locations, where the word "correlation" may mean linear cross correlation or dynamical [Čenys et al., 1991] or probabilistic measures [Buzug et al., 1994; Mayer-Kress & Kurz, 1987].

Most of the analysis tools for STTS are still in their infancy and will have to prove their general usefulness. The investigation of STTS thus provides one of the main challenges of the near future, hopefully providing new tools that can be applied to large data sets with very complex dynamics.

4. Example From Acoustics

As mentioned before, acoustics together with hydrodynamics has been at the forefront of nonlinear dynamics and has also supplied one of the first experimental examples of nonlinear and chaotic dynamics via acoustic cavitation [Lauterborny & Cramer, 1981]. Moreover, it is a spatiotemporal, complex system far from being understood. The bubbles produced are highly nonlinear oscillators with their shock wave formation and light emission during oscillation, and their complicated interaction properties leading to three-dimensional filament formation.

4.1. *Acoustic cavitation noise*

Acoustic cavitation is accompanied by strong sound emission called acoustic cavitation noise, and a cloud of bubbles appears in the liquid forming dancing filaments. To investigate these phenomena an
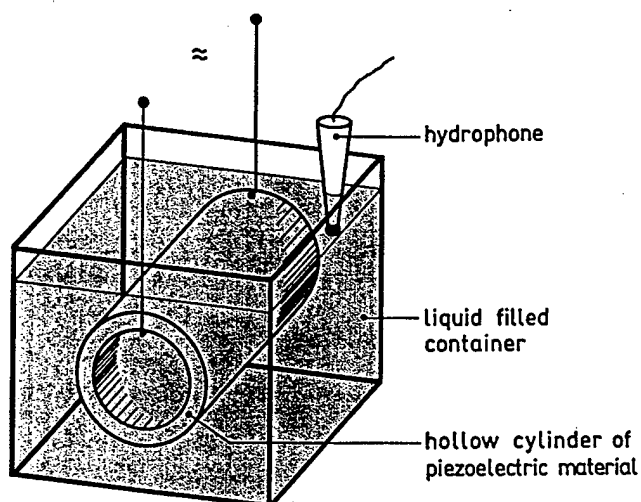
Fig. 14. Basic elements of the acoustic cavitation experiment: A water filled container with a piezoelectric cylindrical transducer and a hydrophone.

experimental arrangement has been set up as depicted in Fig. 14. A cylindrical transducer of piezoelectric material is used to generate a sound field in water. A typical transducer was of 76 mm length, 76 mm inner diameter, and 5 mm wall thickness submerged in a water filled glass container. Cylinders of different size were available driven at different frequencies between 7 kHz and 23 kHz. A standing acoustic wave is set up inside the cylinder having its maximum pressure amplitude along the axis of the transducer. A hydrophone monitors the sound output of the liquid. Some electronics and a computer surround the experiment to drive the cylinder and to store the sampled pressure data from the hydrophone.

A spectral analysis of fully developed cavitation appearing at high driving levels reveals a broadband noise spectrum with lines superimposed at the driving frequency and its harmonics, but also at the subharmonics, notably the first subharmonic at one half the driving frequency and its harmonics. To learn about the onset of this spectrum, experiments with increasing sound pressure amplitude of the driving sound field at, e.g. 23 kHz have been done. Conditions have been found, where a period-doubling cascade is observed leading to a low-dimensional strange attractor [Lauterborn & Cramer, 1981; Lauterborn & Holzfuss, 1991; Lauterborn *et al.*, 1993a]. Figure 15 shows an example of the spectrum of cavitation noise where two period doublings have taken place. The spectrum extends to high harmonics with almost equal amplitudes. The chaotic attractor (actually there seem

to be different types) has been visualized by embedding the pressure samples into reconstruction spaces [Lauterborn & Holzfuss, 1986; Lauterborn *et al.*, 1993a] and quantified by determining the
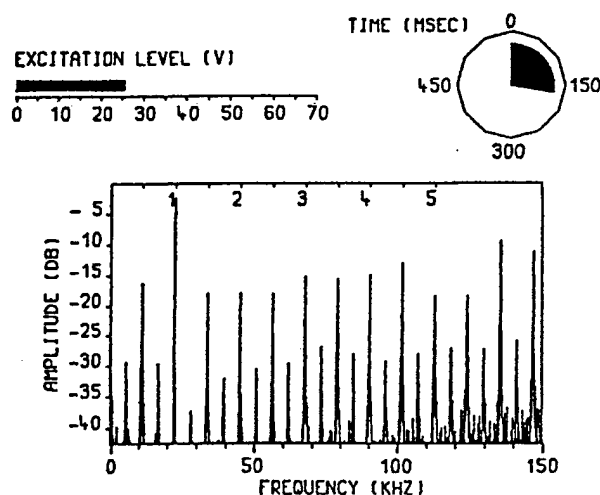

Fig. 15. Power spectrum of acoustic cavitation noise after two period-doubling bifurcations.
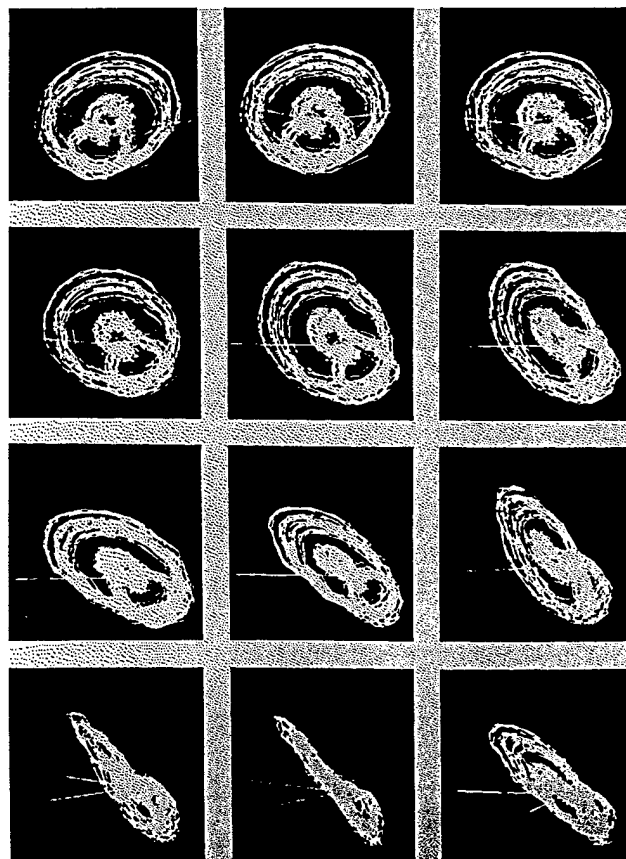

Fig. 16. Example of a chaotic sound attractor embedded in a three-dimensional space. Different projections illustrate its three-dimensional structure.

fractal dimension [Lauterborn & Holzfuss, 1986] and the Lyapunov spectrum [Holzfuss & Lauterborn, 1989]. An astonishingly low-dimensional strange attractor (Fig. 16) is obtained, having a dimension between two and three and just one positive Lyapunov exponent. This fact is so astonishing, because the sound output is generated by hundreds of bubbles of different size each being a nonlinear oscillator and each being capable of period doubling to chaotic oscillations [Lauterborn, 1976; Parlitz *et al.*, 1990]. Therefore some cooperative effects must take place synchronizing the bubble motions. A first approach to model bubble agglomeration in a sound field has been investigated [Parlitz *et al.*, 1995; Akhatov *et al.*, 1994]. We now turn to the measurement of the spatial dynamics.

4.2. Filamentary bubble pattern

The acoustic cavitation noise is produced by a cloud of bubbles oscillating in the driving sound field. They arrange themselves into filaments resembling Lichtenberg figures [Lichtenberg, 1777]. Figure 17 shows an example of the dendritic filament pattern seen along the axis of the transducer. Ten short exposure photographs (15 ns from a copper vapour laser) have been superimposed taken at 375 Hz from acoustic Lichtenberg figures produced from a 15 kHz sound field. As this mode of photographing enhances the structure, the bubbles obviously move along the branches. To clarify the dynamics of the spatial patterns and of the bubbles forming them time-resolved investigations have been initiated as described below.

4.2.1. High speed holographic cinematography

The observation and the visualization of the dynamics of three-dimensional motion is a difficult problem. The visualization is approached by virtual reality with quickly improving results. The observation of three-dimensional physical systems by measurement is intrinsically difficult to perform in real time. Should the optical measurements be done with similar time resolution as the acoustic ones, for instance, at 1 MHz sampling rate, at least a few million GByte per second data rate would be needed, as a single three-dimensional image contains about 1 GByte of voxels (volume picture elements). While this cannot yet be done in real time, high speed holographic cinematography comes close
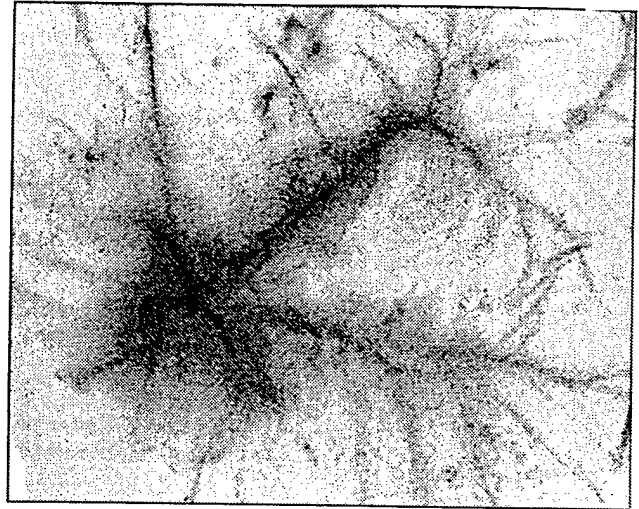

Fig. 17. Example of the filamentary bubble pattern (acoustic Lichtenberg figures) inside the cylinder in the projection along the axis of the transducer (Courtesy of A. Billo).
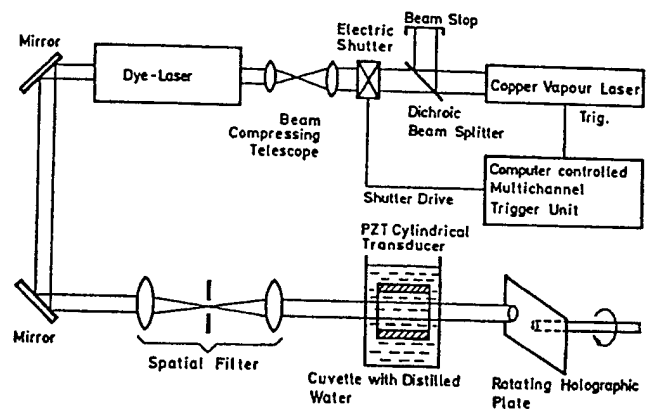

Fig. 18. Arrangement for high speed holographic cinematography with a copper vapour laser pumped dye laser in an in-line geometry.

to it [Hentschel & Lauterborn, 1985]. By taking holograms at a rate of up to 69300 holograms per second it could be shown that the whole filamentary structure is undergoing period doubling simultaneously with period doubling of the sound output [Lauterborn & Koch, 1987]. The acoustic source has thus been traced back to bubble oscillations.

An arrangement for taking in-line holograms with a rotating holographic plate for spatial separation of the individual holograms is depicted in Fig. 18. A copper vapor laser pumped dye laser for producing a series of high power coherent light pulses [Lauterborn *et al.*, 1993b] is used. The device is capable of taking holograms at rates up to 20 000 holograms per second with an area per hologram of up to one centimeter square and a total

capacity of about 40 holograms or more depending on the actual size chosen for a hologram. The copper vapor laser is capable of emitting a quasi infinite series of short light pulses with a duration of about 15 ns and an energy of about 1 mJ each. The light consists of two spectral lines, green and yellow. The dichroic beam splitter selects the green line to pump the dye laser and dumps the yellow line into a beam stop. The electric shutter opens for the time of one revolution of the rotating holographic plate on which the series of holograms is recorded. The dye laser serves as a coherence transformer boosting the coherence length from the few millimeters of the copper vapor laser light pulses up to about 7 cm. Simultaneously the wavelength is changed from $\lambda = 510.6$ nm to $\lambda = 571$ nm (Rhodamine 6G dye). The spatial filter improves the beam quality by removing interference fringes collected by diffraction on dust particles on the mirrors. An in-line holographic arrangement is used. In this geometry the light passing undisturbed through the liquid serves as the reference beam for the light scattered at the bubbles. A series of 16 holographic images is shown in Fig. 19 taken photographically from a series of in-line holograms recorded at 5700 holograms per second. The driving sound field had a frequency of about 7 kHz in this case. The filaments to be seen are quite stable. They are mainly made up of bubbles with a typical size of 50 μm or less.

The images of the holograms have been subjected to three-dimensional image processing as described next.

4.2.2. *Digital holographic-image processing*

A series of holograms contains an enormous amount of information that cannot properly be analyzed manually. Therefore a digital image processing system for three-dimensional images from holograms has been developed to handle the large amount of data [Haussmann & Lauterborn, 1980]. The basic arrangement is depicted in Fig. 20. The hologram is illuminated with the phase conjugate reference beam for projecting the real image into space. An image dissector camera with 4096 × 4096 addressable pixels scans the three-dimensional image plane by plane, typically 50 to 100 planes depending on the depth of the investigated volume, and feeds the data into a computer system. It mainly consists of a transputer net of up to 45 transputers T800 attached to a workstation with 128 MB RAM. The
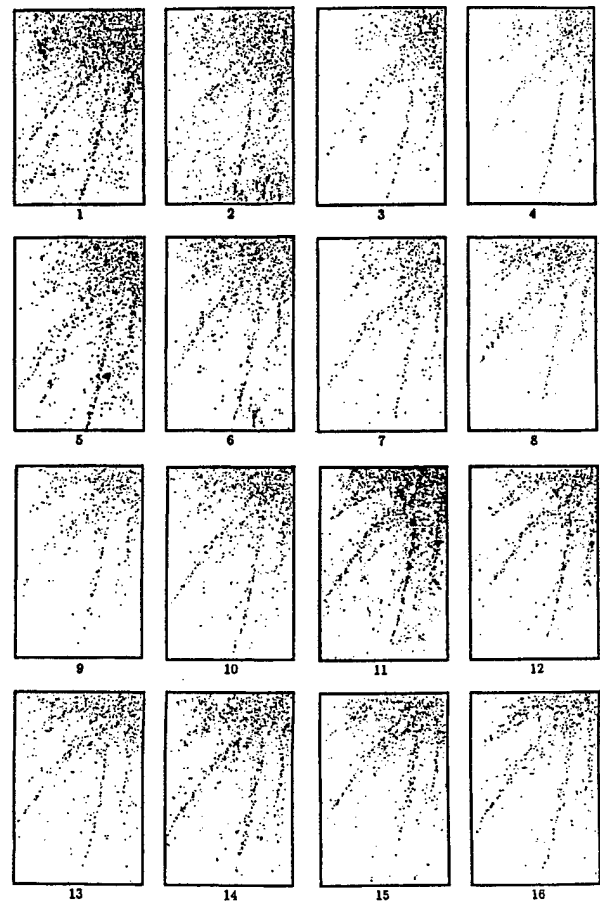

Fig. 19. Reconstructed holographic images from a series of holograms taken at a rate of 5700 holograms per second of a filamentary bubble pattern.
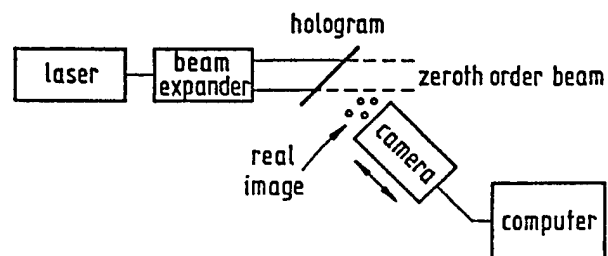

Fig. 20. Arrangement for digital image processing from (off-axis) holograms using the phase conjugate image, a high resolution image dissector camera and parallel computing.

transputers do the preprocessing of filtering and condensing the data leaving a data set per hologram of about 100 MB down from the about 1 GB of raw data read in. A large software package has been written for detecting and locating bubbles in the three-dimensional real image.

It is notoriously difficult to visualize three-dimensional point distributions. The solution, that,

Fig. 21. Color coded three-dimensional plot of a bubble distribution (Courtesy of H. Chodura).

however, cannot be printed, is virtual reality where the observer can look around in the distribution. As a first step to implement a virtual reality system for this purpose we have stored a data base of three-dimensional images of bubble distributions obtained from a holographic series. One of the images is given in Fig. 21 with color coded depth information and bubbles plotted as spheres according to their size. They can be viewed on a computer monitor with stereo capabilities and shutter eyewear. The holographic-image processing system is continuously improved for better performance with the ultimate goal to be able to study space-time chaotic systems.

5. Outlook

During the last two decades, nonlinearity has become a new paradigm of science (and not only of physics), because nonlinear phenomena are present in virtually all areas of our physical existence. Not accidentally, the development of nonlinear dynamics took place in an interdisciplinary fashion.

Nonlinearity manifests itself in the distribution of galaxies observed far out in the universe, in the distribution of matter in our milky way and the motion of celestial bodies, in the turbulent flows within the atmosphere and the oceans, in the overwhelming diversity of biological life on our planet, and, further down in size, to the motion of individual atoms. Nonlinear theoretical models of the physical world abound, and, in fact, fundamental theories of physics such as Einstein's theory of gravitation and other field theories are of nonlinear nature.

In view of this complexity that we are embedded in, how could man proceed in the unveiling of physical truth? The key to success was empiricism and reductionism, as Descartes taught us: In performing physical experiments the essential phenomenon has to be isolated, other, distracting influences have to be suppressed, all relevant quantities have to be controlled. In this sense, experiments are idealized — we reduce the complexity of a problem until we are able to identify relations that unravel the underlying physical laws. This line of attack has proved to be very successful, and

many basic laws of nature have been found and formulated.

To know the laws, however, does not mean that we may be able to predict how a given physical system will behave in the long run. This is the basic lesson that we have learned from nonlinear dynamics; the principle of limited predictability of nonlinear, classically deterministic, but chaotic systems. In this context, it imposes a limit on our ability to know that is comparable to the fundamental consequences of the uncertainty principle for quantum systems.

In setting out to discuss the future of nonlinear experimental physics we must be aware that its evolution might also be "sensitively dependent" on unforeseeable events (it probably is). There are enough examples where predictions were tried and failed thoroughly. However, educated guesses and some speculations about the future directions of nonlinear experimental science are possible, and we should nevertheless make plans for the next few years, each individual to his or her best knowledge. In that, diversity and freedom of basic research are the guarantees for new fields of investigation to emerge and for the promotion of man's knowledge.

The evolution of experimental physics depends mainly on two closely interacting developments: The progress in experimental techniques and instrumentation, and the shift of focus on the investigated physical phenomena, which is related to the advancement of theory. That both areas are intimately connected has been shown by the progress of nonlinear dynamics during the last decade. The realization that deterministic systems with just a few degree of freedoms can generate erratic and complex behavior has launched experimental activity which was partly reviewed in Sec. 2. The understanding of such systems, on the other hand, has given experimenters new, computer-aided tools to better analyze their experimental data. Thus, nonlinear time series analysis has become an indispensable tool for the analysis of chaotic dynamics and may be said to lie at the heart of experimental nonlinear physics. Up to now, and certainly for some time to come, time series of scalar or vector observables (with a few components only) will be the typically available outcomes of experiments. Thus there exists a strong motivation to improve on the by now quite elaborate methods of nonlinear time series analysis.

The science of nonlinear dynamics of low-dimensional systems has matured. Transition to chaos and the properties of low-dimensional attractors are quite well-understood now. The focus of attention will therefore shift to systems with a higher number of degrees of freedom, such as coupled oscillator systems and spatially extended systems in hydrodynamics, optics or acoustics. This implies that data taking will increasingly become more-dimensional to capture the spatio-temporal evolution. First experiments that rely on the processing of picture sequences are now undertaken. Technology has reached a stage where two-dimensional picture streams (video data, but also pictures of higher resolution), i.e. three-dimensional data with two spatial and one time coordinate, can be handled or at least stored continuously for off-line processing. It can be imagined that even four-dimensional data streams can and will be taken (three space, one time coordinate) in the form of holographic series as reported above. As in the past, progress will strongly depend on available technology, in particular computational speed and large-data-handling capability.

For such type of data (and the huge information streams associated with it) new methods of time series analysis for multidimensional data have to be developed. The methods should be designed to extract the most relevant information from the data in terms of some underlying theoretical model. It becomes quite clear that such modern experiments in nonlinear physics are only conceivable in connection with powerful computers. Indeed, the border between the real experiment and its analysis by simulation will become fuzzy. Computing power competes with experiment in sophisticated simulations, but it will also be necessary for analyzing experiments with complex systems.

As computing technology advances, the on-line processing power of experimental equipment will allow not only real-time evaluation of experimental outcomes but also the in-time reaction and manipulation of the system under investigation. Hopefully, this will open up new ways for the controlling of complex dynamics, with many interesting consequences for applications. To speculate a bit, this would also make possible a totally new experimental approach to complex systems that might help to reduce the (perhaps unnecessarily large) flood of data from high-dimensional systems: Automated interactive experimentalization. The computer, controlling all relevant experimental parameters (and there could be many!), could be programmed as an adaptive system interacting with the experiment by "asking questions" that depend on the course of the

measurement. A simple and straightforward application of such an approach would be the analysis of the instabilities and of the bifurcation set of a hydrodynamic experiment, say.

However, experiments are not made for the delight of a computer (however smart it may be) but for the curious human being that turns on the knob. When the systems under investigation and their dynamics are getting more complex, the experimental outcome has to be presented in the most intelligible way. To this end, virtual reality and 3-D visualization techniques have already been introduced into experiments and, in the future, will become more important. And in view of the above discussion, it is quite imaginable that the experimenter will take part in interactive experimentalization by acting in an artificial, knowledge- or model-based environment, with the computer as the interface between an abstract, conceptual representation and the real physical situation. This would be, to the authors' opinion, the most desirable form of a *virtual laboratory*, where physical experiments have not been replaced by simulations but are supported by them.

At this point it has to be remarked that, if these speculations become true, some well-established distinctions between scientific disciplines and their objects of study will have disappeared. For example, in a computer-based virtual lab, methods of artificial intelligence or complex dynamics (e.g. adaptive neural nets) might be employed to aid in the analysis. Regarding the object of study, a distinction between natural objects and human-made and thus artificial objects, e.g. self-organizing computing systems, might become irrelevant.

This touches on one of the central themes of 21st century physics: The problem of complexity. The word being often and widely used, a satisfactory definition of this notion is still lacking. Intuitively, we relate complexity with structural features or dynamical behavior that "go beyond chaos", that are describable neither by a small set of characteristic numbers nor by simple algorithms. The traces of complexity, like those of chaos, are all around us in the natural world and human society where nonlinearity and interaction of many elements come into play. Indeed, to tackle complexity experimentally, quite complicated, interacting systems have to be studied. Whether such endeavour can be reconciled with the successful reductionistic approach is still an open question.

Looking at the current state of experimental nonlinear physics, we recognize that classical non-linear dynamical systems are an important but not the only class of experimental systems. Nonlinear effects turn up, e.g. in a variety of transient phenomena that cannot be described by attractors in a given phase space. For example, the interaction of ultrashort light pulses with matter, the emission of shock waves and light in the process of cavitation, dielectric breakdown with plasma formation, the collapse of a supernova or laser fusion are intrinsically nonlinear processes that are not considered in the context of dynamical systems. In such processes, many physical phenomena take place and interact, which makes numerical simulations difficult. Here, experiments are still ahead of the simulation capabilities. Also, new experimental techniques, in particular the development of high-power femtosecond lasers, have become available that will push the frontiers of experimentally achieved time resolution and power. And there is the thrill of experimental physics: Not to know which new phenomena and new laws might expect us; now within reach, but still to be discovered.

Acknowledgments

References

Abarbanel, H. D. I., Brown, R. & Kennel, M. B. [1991] "Lyapunov exponents in chaotic systems: Their importance and their evaluation using observed data," *Int. J. Mod. Phys.* **B5**, 1347–1375.

Abarbanel, H. D. I., Brown, R., Sidorovich, J. J. & Tsimring, L. SH. [1993] "The analysis of observed chaotic data in physical systems," *Rev. Mod. Phys.* **65**, 1331–1392.

Abraham, N. B., Arecchi, F. T. & Lugiato, L. A. (ed.) [1987] *Instabilities and Chaos in Quantum Optics II* (Plenum, New York).

Afraimovich, V. S., Verichev, N. N. & Rabinovich, M. I. [1986] "Stochastic synchronization of oscillations in dissipative systems," *Radiophys. Quantum Electron.* **29**, 795–803.

Ahlers, G. & Behringer, R. L. [1978] "Evolution of turbulence from the Rayleigh–Bénard instability," *Phys. Rev. Lett.* **40**, 712.

Akhatov, I., Parlitz, U. & Lauterborn, W. [1994] "Pattern formation in acoustic cavitation," *J. Acoust. Soc. Am.* **96**, 3627–3635.

Alecsić, Z. [1991] "Estimating the embedding dimension," *Physica* **D52**, 362–368.

Amritkar, R. E. & Gupte, N. [1993] "Synchronization of chaotic orbits: The effect of a finite time step," *Phys. Rev.* **E47**, 3889–3895.

Arecchi, F. T., Meucci, R., Puccioni, G. & Tredicce, J. [1982] "Experimental evidence of subharmonic bifurcations, multistability, and turbulence in a Q-switched gas laser," *Phys. Rev. Lett.* **49**, 1217.

Aubry, N., Guyonnet, R. & Lima, R. [1991] "Spatiotemporal Analysis of Complex Signals: Theory and Applications," *J. Stat. Phys.* **64**, 683–739.

Baake, E., Baake, M., Bock, H. G. & Briggs, K. M. [1992] "Fitting ordinary differential equations to chaotic data," *Phys. Rev.* **A45**, 5524–5529.

Badii, R. & Politi, A. [1984] "Hausdorff dimension and uniformity factor of strange attractors," *Phys. Rev. Lett.* **52**, 1661–1664.

Badii, R. & Politi, A. [1985] "Statistical description of chaotic attractors," *J. Stat. Phys.* **40**, 725–750.

Barber, B. P. & Putterman, S. J. [1991] "Observation of synchronous picosecond sonoluminescence," *Nature* **352**, 318–320.

Barber, B. P., Hiller, R., Arisaka, K., Fetterman, H. & Putterman, S. [1992] "Resolving the picosecond characteristics of synchronous sonoluminescence," *J. Acoust. Soc. Am.* **91**, 3061–3063.

Bechhoefer, J., Ego, V., Manneville, S. & Johnson, B. [1995] "An experimental study of the onset of parametrically pumped surface waves in viscous fluids," *J. Fluid Mech.* **288**, 325–350.

Behringer, R. P. [1985] "Rayleigh–Bénard convection and turbulence in liquid helium," *Rev. Mod. Phys.* **57**, 657–687.

Bénard, H. [1901] "Les tourbillons cellulaires dans une nappe liquide transportant de la chaleur par convection en régime permanent," *Annales de Chimie et de Physique*, 7e série, t. XXIII, 62–144.

Benettin, G., Galgani, L., Giorgilli, A. & Strelcyn, J.-M. [1980] "Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems; a method for computing all of them. Part II: Numerical application," *Meccanica* **15**, 21–30.

Bergé, P., Pomeau, Y. & Vidal, C. [1984] *Order within Chaos: Towards a Deterministic Approach to Turbulence* (John Wiley & Sons, New York).

Berry, D. A., Titze, H., Herzel, H. & Krischer, K. [1994] "Interpretation of biomedical simulations of normal and chaotic vocal fold oscillations," *J. Acoust. Soc. Am.* **95**, 3595–3604.

Beyer, R. T. [1975] *Nonlinear Acoustics* (U. S. Government Printing Office).

Bowden, C. M. (ed.) [1981] *Optical Bistability* (Plenum, New York).

Breeden, J. L. & Huebler, A. [1990] "Reconstructing equations of motion from experimental data with unobserved variables," *Phys. Rev.* **A42**, 5817–5826.

Briggs, K. [1990] "An improved method for estimating Liapunov exponents of chaotic time series," *Phys. Lett.* **A151**, 27–32.

Broggi, G. [1988] "Evaluation of dimensions and entropies of chaotic systems," *J. Opt. Soc. Am.* **B5**, 1020–1028.

Broomhead, D. S. & King, G. P. [1986] "Extracting qualitative dynamics from experimental data," *Physica* **D20**, 217–236.

Broomhead, D. S., Huke, J. P. & Muldoon, M. R. [1992] "Linear filters and nonlinear systems," *J. Roy. Stat. Soc.* **B54**, 373–382.

Brown, R., Bryant, P. & Abarbanel, H. D. I. [1991] "Computing the Lyapunov spectrum of a dynamical system from an observed time series," *Phys. Rev.* **A43**, 2787–2806.

Brown, R., Rulkov, N. F. & Tracy, E. R. [1994] "Modeling and synchronizing chaotic systems from time-series data," *Phys. Rev.* **E49**, 3784–3800.

Bryant, P., Brown, R. & Abarbanel, H. D. I. [1990] "Lyapunov exponents from observed time series," *Phys. Rev. Lett.* **65**, 1523–1526.

Buzug, Th., Reimers, T. & Pfister, G. [1990] "Optimal reconstruction of strange attractors from purely geometrical arguments," *Europhys. Lett.* **13**, 605–610.

Buzug, Th. & Pfister, G. [1992] "Optimal delay time and embedding dimension for delay-time coordinates by analysis of the global static and local dynamical behavior of strange attractors," *Phys. Rev.* **A45**, 7073–7084.

Buzug, T., Pawelzik, K., von Stamm, J. & Pfister, G. [1993] "Mutual information and global strange attractors in Taylor–Couette flow," *Physica* **D72**, 343–350.

Carnell, M. T., Fiadeiro, P. T. & Emmony, D. C. [1995] "Cavitation phenomena generated by a lithotripter shock wave," *J. Acoust. Soc. Am.* **97**, 677–679.

Carroll, T. L. & Pecora, L. M. [1993] "Cascading synchronized chaotic systems," *Physica* **D67**, 126–140.

Casdagli, M. [1989] "Nonlinear prediction of chaotic time series," *Physica* **D35**, 335–356.

Casdagli, M., Eubank, S., Farmer, J. D. & Gibson, J. [1991] "State space reconstruction in the presence of noise," *Physica* **D51**, 52–98.

Čenys, A. [1993] "Lyapunov spectrum of the maps generating identical attractors," *Europhys. Lett.* **21**, 407–411.

Čenys, A. & Pyragas, K. [1988] "Estimation of the number of degrees of freedom from chaotic time series," *Phys. Lett.* **A129**, 227–230.

Čenys, A., Lasiene, G. & Pyragas, K. [1991] "Estimation of interrelation between chaotic observables," *Physica* **D52**, 332–337.

Chauve, M. P. & Le Gal, P. [1992] "Complex biorthogonal decomposition of a chain of coupled wakes," *Physica* **D38**, 407–413.

Ciliberto, S. & Gollub, J. P. [1985] "Chaotic mode competition in parametrically forced surface waves," *J. Fluid Mech.* **158**, 381–398.

Ciliberto, S. & Nicolaenko, B. [1991] "Estimating the number of degrees of freedom in spatially extended systems," *Europhys. Lett.* **14**, 303–308.

Coleman, A. J., Saunders, J. E., Crum, L. A. & Dyson, M. [1987] "Acoustic cavitation generated by an extracorporeal shockwave lithotripter," *Ultrasound Med. Biol.* **13**, 69–76.

Crum, L. A. [1994] "Sonoluminescence," *Physics Today* Sept. **47**, 22–27.

Cuomo, K. M. & Oppenheim, A. V. [1993] "Circuit implementation of synchronized chaos with applications to communications," *Phys. Rev. Lett.* **71**, 65–68.

Dämmig, M. & Mitschke, F. [1993] "Estimation of Lyapunov exponents from time series: The stochastic case," *Phys. Lett.* **A178**, 385–394.

Diacu, F. & Holmes, P. [1996] *Celestial Encounters — The Origin of Chaos and Stability* (Princeton University Press, New Jersey).

Ditto, W., Pecora, L., Shlesinger, M., Spano, M. & Vohra, S. (eds.) [1994] *Proceedings of the 2nd Experimental Chaos Conference* (World Scientific, Singapore).

Eckmann, J.-P. & Ruelle, D. [1985] "Ergodic theory of chaos and strange attractors," *Rev. Mod. Phys.* **57**, 617–656.

Eckmann, J.-P., Kamphorst, S. O., Ruelle, D. & Ciliberto, S. [1986] "Lyapunov exponents from time series," *Phys. Rev.* **A34**, 4971–4979.

Eckmann, J.-P. & Ruelle, D. [1992] "Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems," *Physica* **D56**, 185–187.

Ellner, S., Gallant, A. R., McCaffrey, D. & Nychka, D. [1991] "Convergence rates and data requirements for Jacobian-based estimates of Lyapunov exponents from data," *Phys. Lett.* **A153**, 357–363.

Faraday, M. [1831] "On a peculiar class of acoustical figures, and on certain forms assumed by groups of particles upon vibrating elastic surfaces," *Phil. Trans. Roy. Soc. London* **121**, 299–340. See Appendix: "On the forms and states assumed by fluids in contact with vibrating elastic surfaces," pp. 319–340.

Farmer, J. D. & Sidorowich, J. J. [1987] "Predicting chaotic time series," *Phys. Rev. Lett.* **59**, 845–848.

Feldmann, K. T. [1968a] "Review of the literature on Sondhauss thermoacoustic phenomena," *J. Sound Vib.* **7**, 71–82.

Feldmann, K. T. [1968b] "Review of the literature on Rijke thermoacoustic phenomena," *J. Sound Vib.* **7**, 83–89.

Feldmann, U., Hasler, M. & Schwarz, W. [1996] "Communication by chaotic signals: The inverse systems approach," *Int. J. Circuit Theory Appl.* **24**, 551–579.

Fell, J., Röschke, J. & Beckmann, P. [1993] "Deterministic chaos and the first positive Lyapunov exponent: A nonlinear analysis of the human electroencephalogram during sleep," *Biol. Cybern.* **69**, 139–146.

Fell, J. & Beckmann, P. [1994] "Resonance-like phenomena in Lyapunov calculations from data reconstructed by the time-delay method," *Phys. Lett.* **A190**, 172–176.

Frazer, A. M. & Swinney, H. L. [1986] "Independent coordinates in strange attractors from mutual information," *Phys. Rev.* **A33**, 1134–1140.

Frazer, A. M. [1989a] "Reconstructing attractors from scalar time series: A comparison of singular system and redundancy criteria," *Physica* **D34**, 391–404.

Frazer, A. M. [1989b] "Information and entropy in strange attractors," *IEEE Trans. Inf. Theory* **35**, 245–262.

Fujisaka, H. & Yamada, T. [1983] "Stability theory of synchronized motion in coupled-oscillator systems," *Prog. Theor. Phys.* **69**, 32–46.

Gaitan, D. F. & Crum, L. A. [1990] "Observation of sonoluminescence from a single, stable cavitation bubble in a water/glycerine mixture," in *Frontiers of Nonlinear Acoustics*, Proc. 12th Int. Symp. on Nonlinear Acoustics, eds. Hamilton, M. & Blackstock, D. T. (Elsevier, New York), pp. 459–463.

Gao, J. & Zheng, Z. [1993] "Local exponential divergence plot and optimal embedding of a chaotic time series," *Phys. Lett.* **A181**, 153–158.

Gao, J. & Zheng, Z. [1994] "Direct dynamical test for deterministic chaos and optimal embedding of a chaotic time series," *Phys. Rev.* **E49**, 3807–3814.

Gaponov-Grekhov, A. V. & Rabinovich, M. I. [1992] *Nonlinearities in Action — Oscillations, Chaos, Order, Fractals* (Springer, Berlin).

Geist, K., Parlitz, U. & Lauterborn, W. [1990] "Comparison of different methods for computing Lyapunov exponents," *Prog. Theor. Phys.* **83**, 875–893.

Gencay, R. & Dechert, W. D. [1992] "An algorithm for the n Lyapunov exponents of an n-dimensional unknown dynamical system," *Physica* **D59**, 142–157.

Gibbs, H. M., Hopf, F. A., Kaplan, D. L. & Shoemaker, R. L. [1981] "Observation of chaos in optical bistability," *Phys. Rev. Lett.* **46**, 474.

Gibbs, H. M. [1986] *Optical Bistability: Controlling Light by Light* (Academic Press, New York).

Gibiat, V. [1988] "Phase space representation of acoustical musical signals," *J. Sound Vib.* **123**, 529–536.

Gibson, J. F., Farmer, J. D., Casdagli, M. & Eubank, S. [1992] "An analytic approach to practical state space reconstruction," *Physica* **D57**, 1–30.

Gioggia, R. S. & Abraham, N. B. [1983] "Routes to chaotic output from a single-mode, dc-excited laser," *Phys. Rev. Lett.* **51**, 650.

Glazier, J. A. & Libchaber, A. [1988] "Quasi-periodicity and dynamical systems: An experimentalist's view," *IEEE Trans. Circuits Syst.* **35**, 790–809.

Gollub, J. P. & Swinney, H. L. [1975] "Onset of turbulence in a rotating fluid," *Phys. Rev. Lett.* **35**, 927–930.

Gollub, J. P. & Benson, S. V. [1980] "Many routes to turbulent convection," *J. Fluid Mech.* **100**, 449–470.

Grassberger, P. & Procaccia, I. [1983] "On the characterization of of strange attractors," *Phys. Rev. Lett.* **50**, 346–349.

Grassberger, P. [1985] "Generalizations of the Hausdorff dimension of fractal measures," *Phys. Lett.* **A107**, 101–105.

Grassberger, P., Schreiber, T. & Schaffrath, C. [1991] "Nonlinear time sequence analysis," *Int. J. Bifurcation and Chaos* **1**, 521–547.

Grassberger, P., Hegger, R., Kantz, H., Schaffrath, C. & Schreiber, T. [1993] "On noise reduction methods for chaotic data," *Chaos* **3**, 127–141.

Guckenheimer, J. & Holmes, P. [1983] *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer, New York).

Haken, H. [1975] "Analogy between higher instabilities in fluids and lasers," *Phys. Lett.* **A53**, 77–78.

Halle, K. S., Wu, C. W., Itoh, M. & Chua L. O. [1993] "Spread spectrum communication through modulation of chaos," *Int. J. Bifurcation and Chaos* **3**, 469–477.

Hasegawa, A. & Tappert, F. [1973] "Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. I. Anomalous dispersion," *Appl. Phys. Lett.* **23**, 142–144.

Haussmann, G. & Lauterborn, W. [1980] "Determination of size and position of fast moving gas bubbles in liquids by digital 3-D image processing of hologram reconstructions," *Appl. Opt.* **19**, 3529–3535.

Heisenberg, W. [1967] "Nonlinear problems in physics," *Physics Today* May, 27–33.

Helmholtz, H. [1870] *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (Vieweg, Braunschweig).

Hentschel, W. & Lauterborn, W. [1985] "High speed holographic movie camera," *Opt. Eng.* **24**, 687–691.

Herzel, H. [1993] "Bifurcations and chaos in voice signals," *Appl. Mech. Rev.* **46**, 399–413.

Holden, A. V. [1986] *Chaos* (Manchester University Press, Manchester).

Holt, R. G., Gaitan, D. F., Atchley, A. A. & Holzfuss, J. [1994] "Chaotic sonoluminescence," *Phys. Rev. Lett.* **72**, 1376–1379.

Holzfuss, J. & Mayer-Kress, G. [1986] "An approach to error-estimation in the application of dimension algorithms," in *Dimensions and Eutropies in Chaotic Systems — Quantification fo complex behavior* ed. Mayer-Kress, G. (Springer, Berlin), pp. 114–122.

Holzfuss, J. & Lauterborn, W. [1989] "Liapunov exponents from a time series of acoustic chaos," *Phys. Rev.* **A39**, 2146–2152.

Holzfuss, J & Parlitz, U. [1991] "Lyapunov exponents from time series," *Proc. Conf. Lyapunov Exponents* (Oberwolfach 1990) eds. L. Arnold, Crauel, H. & Eckmann, J.-P. in *Lecture Notes in Mathematics* (Springer, Berlin), 263–270.

Huerta, R., Santa Cruz, C., Dorronsore, J. R. & Lòpez, V. [1995] "Local state-space reconstruction using averaged scalar products of dynamical-system flow vectors," *Europhys. Lett.* **29**, 13–18.

Idogawa, T., Kobata, T., Komuro, K. & Iwaki, M. [1993] "Nonlinear vibrations in the air column of a clarinet artificially blown," *J. Acoust. Soc. Am.* **93**, 540–551.

Ikeda, K. [1979] "Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system," *Opt. Comm.* **30**, 257–261.

Kadtke, J. B., Brush, J. & Holzfuss, J. [1993] "Global dynamical equations and Lyapunov exponents from noisy chaotic time series," *Int. J. Bifurcation and Chaos* **3**, 607–616.

Kantz, H., Schreiber, T., Hoffmann, I., Buzug, T., Pfister, G., Flepp, C. G., Simonet, J., Badii, R. & Brun, E. [1993] "Nonlinear noise reduction: A case study on experimental data," *Phys. Rev.* **E48**, 1529–1538.

Kantz, H. [1994] "A robust method to estimate the maximal Lyapunov exponent of a time series," *Phys. Lett.* **A185**, 77–87.

Kaplan, D. [1994] "Exceptional events as evidence for determinism," *Physica* **D73**, 38–48.

Kember, G. & Fowler, A. C. [1993] "A correlation function for choosing time delays in phase portrait reconstructions," *Phys. Lett.* **A179**, 72–80.

Kennel, M. B., Brown, R. & Abarbanel, H. D. I. [1992] "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev.* **A45**, 3403–3411.

Kocarev, L., Halle, K. S., Eckert, K., Chua, L. O. & Parlitz, U. [1992] "Experimental demonstration of secure communications via chaotic synchronization," *Int. J. Bifurcation and Chaos* **2**, 709–713.

Kocarev, L. & Parlitz, U. [1995] "General approach for chaotic synchronization with applications to communcation," *Phys. Rev. Lett.* **74**, 5028–5031.

Kocarev, L. & Parlitz, U. [1996a], "Synchronizing spatiotemporal chaos in coupled nonlinear oscillators," *Phys. Rev. Lett.* **77**, 2206–2209.

Kocarev, L. & Parlitz, U. [1996b], "Generalized synchronization, predictability and equivalence of unidirectionally coupled dynamical systems," *Phys. Rev. Lett.* **76**, 1816–1819.

Korteweg, D. J. & deVries, G. [1895] "On the change of form of long waves advancing in a rectangular channel, and on a new type of long stationary wave," *Phil. Mag.* **39**, 422–443.

65

Kostelich, E. J. & Schreiber, T. [1993] "Noise reduction in chaotic time-series data: A survey of common methods," *Phys. Rev.* **E48**, 1752–1763.

Kwasniok, F. [1996] "The reduction of complex dynamical systems using principal interaction patterns," *Physica* **D92**, 28–60.

Kurths, J. & Herzel, H. [1987] "An attractor in solar time series," *Physica* **D25**, 165–172.

Lai, Y.-C. & Grebogi, C. [1994] "Synchronization of spatiotemporal chaotic systems by feedback control," *Phys. Rev.* **E50**, 1894–1899.

Landa, P. S. & Rosenblum, M. G. [1991] "Time series analysis for system identification and diagnostics," *Physica* **D48**, 232–254.

Lauterborn, W. [1976] "Numerical investigation of nonlinear oscillations of gas bubbles in liquids," *J. Acoust. Soc. Am.* **59**, 283–293.

Lauterborn, W. & Cramer, E. [1981] "Subharmonic route to chaos observed in acoustics," *Phys. Rev. Lett.* **47**, 1445–1448.

Lauterborn, W. & Holzfuss, J. [1986] "Evidence for a low dimensional strange attractor in acoustic turbulence," *Phys. Lett.* **A115**, 369–372.

Lauterborn, W. & Koch, A. [1987] "Holographic observation of period-doubled and chaotic bubble oscillations in acoustic cavitation," *Phys. Rev.* **A35**, 1774–1976.

Lauterborn, W. & Parlitz, U. [1988] "Methods of chaos physics and their application to acoustics," *J. Acoust. Soc. Am.* **84**, 1975–1993.

Lauterborn, W. & Holzfuss, J. [1991] "Acoustic chaos," *Int. J. Bifurcation and Chaos* **1**, 13–26.

Lauterborn, W., Schmitz, E. & Judt, A. [1993a] "Experimental approach to a complex acoustic system," *Int. J. Bifurcation and Chaos* **3**, 635–642.

Lauterborn, W., Judt, A. & Schmitz, E. [1993b] "High-speed off-axis holographic cinematography with a copper-vapor-pumped dye laser," *Opt. Lett.* **18**, 4–6.

Lauterborn, W., Kurz, T. & Wiesenfeldt, M. [1995] *Coherent Optics* (Springer, Berlin).

Lauterborn, W. [1996] "Nonlinear dynamics in acoustics," *Acustica* **82**, S46–S55.

Legge, K. A. & Fletcher, N. H. [1989] "Nonlinearity, chaos, and the sound of shallow gongs," *J. Acoust. Soc. Am.* **86**, 2439–2443.

Leighton, T. G. [1994] *The Acoustic Bubble* (Academic, London).

Libchaber, A. & Maurer, J. [1980] "Une expérience de Rayleigh–Bénard de géometrie réduite: Multiplication, accrochage et démultiplication de fréquences," *J. Phys. Coll.* **41**, 51–56.

Libchaber, A. [1982] "A Rayleigh–Bénard experiment: Helium in a small box," in *Nonlinear Phenomena at Phase Transitions and Instabilities*, ed. Riste, T. (Plenum Press, New York), pp. 259–286.

Lichtenberg, G. C. [1777] "De nova methodo naturam ac motum fluidi electrici investigandi," *Novi Commentarii Societatis Regiae Scientiarum Gottingensis*, tomus VIII, ad annum 1777, Commentationes physicae et mathematicae classis, Gottingae: Apud Joann. Christian Dieterich, pp. 168–180.

Liebert, W. & Schuster, H. G. [1989] "Proper choice of the time delay for the analysis of chaotic time series," *Phys. Lett.* **A142**, 107–111.

Liebert, W., Pawelzik, K. & Schuster, H. G. [1991] "Optimal embeddings of chaotic attractors from topological considerations," *Europhys. Lett.* **14**, 521–526.

Lorenz, E. N. [1963] "Deterministic nonperiodic flow," *J. Atmos. Sci.* **20**, 130–141.

Lumley, J. [1970] *Stochastic Tools in Turbulence* (Academic Press, New York).

Madan, R. N. (ed.) [1993] *Chua's Circuit: A Paradigm for Chaos* (World Scientific, Singapore).

Maganza, C., Caussé, R. & Laloë, F. [1986] "Bifurcations, period doublings and chaos in clarinet–like systems," *Europhys. Lett.* **1**, 295–302.

Martinerie, J. M., Albano, A. M., Mees, A. I. & Rapp, P. E. [1992] "Mutual information, strange attractors, and the optimal estimation of dimension," *Phys. Rev.* **A45**, 7058–7064.

Mayer-Kress, G. (ed.) [1986] *Dimensions and Entropies in Chaotic Systems — Quantification of Complex Behavior* (Springer, Berlin).

Mayer-Kress, G. & Kurz, T. [1987] "Dimension densities for turbulent systems with spatially decaying correlation functions," *Complex Systems* **1**, 820–829.

Mayer-Kress, G., Choi, I., Weber, N., Barger, R. & Hübler, A. [1993] "Musical signals from Chua's circuit," *IEEE Trans. Circuits Syst.* **40**, 688–695.

McIntyre, M. E., Schumacher, R. T. & Woodhouse, J. [1983] "On the oscillations of musical instruments," *J. Acoust. Soc. Am.* **74**, 1325–1345.

Melde, F. [1860] "Über die Erregung stehender Wellen eines fadenförmigen Körpers," *Ann. Phys. Chem.* **109**, 193–215.

Mende, W., Herzel, H. & Wermke, K. [1990] "Bifurcation and chaos in newborn infant cries," *Phys. Lett.* **A145**, 418–424.

Mettin, R., Parlitz, U. & Lauterborn, W. [1993] "Bifurcation structure of the driven van der Pol oscillator," *Int. J. Bifurcation and Chaos* **3**, 1529–1555.

Mollenauer, L. F., Stolen, R. H. & Gordon, J. P. [1980] "Experimental observation of picosecond pulse narrowing and solitons in optical fibers," *Phys. Rev. Lett.* **45**, 1095–1098.

Mollenauer, L. F. & Gordon, J. P. [1984] "The soliton laser," *Opt. Lett.* **9**, 13–15.

Moon, F. C. [1992] *Chaotic and Fractal Dynamics* (John Wiley & Sons, New York).

Mork, J., Mark, J. & Tromborg, B. [1990] "Route to chaos and competition between relaxation oscillations for a semiconductor laser with optical feedback," *Phys. Rev. Lett.* **65**, 1999–2002.

Müller, G. & Lauterborn, W. [1995] "Experiments with the thermoacoustic oscillator," *Proc. Int. Symp. Musical Acoustics* July 2–6, Le Normont, Dourdan, France, pp. 178–183.

Müller, G. & Lauterborn, W. [1996] "The bowed string as a nonlinear dynamical system," *Acustica* **82**, 657-664.

Mullin, T. [1993] *The Nature of Chaos* (Clarendon Press, Oxford).

Oseledec, V. I. [1968] "A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems," *Trans. Moscow Math. Soc.* **19**, 197–231.

Ott, E. [1993] *Chaos in Dynamical Systems* (Cambridge University Press, Cambridge).

Packard, N. H., Crutchfield, J. P., Farmer, J. D. & Shaw, R. S. [1980] "Geometry from a time series," *Phys. Rev. Lett.* **45**, 712–716.

Palus, M. & Dvorak, I. [1992] "Singular-value decomposition in attractor reconstruction: Pitfalls and precautions," *Physica* **D55**, 221–234.

Palus, M., Albrecht, V. & Dvorak, I. [1993] "Information theoretic test for nonlinearity in time series," *Phys. Lett.* **A175**, 203–209.

Parlitz, U. & Lauterborn, W. [1987] "Period-doubling cascades and devil's staircases of the driven van der Pol oscillator," *Phys. Rev.* **A36**, 1428–1434.

Parlitz, U., Englisch, V., Scheffczyk, C. & Lauterborn, W. [1990] "Bifurcation structure of bubble oscillators," *J. Acoust. Soc. Am.* **88**, 1061–1077.

Parlitz, U. [1992] "Identification of true and spurios Lyapunov exponents from time series," *Int. J. Bifurcation and Chaos* **2**, 155–165.

Parlitz, U., Chua, L. O., Kocarev, L., Halle, K. S. & Shang, A. [1992] "Transmission of digital signals by chaotic synchronization," *Int. J. Bifurcation and Chaos* **2**, 973–977.

Parlitz, U. [1993] "Lyapunov exponents from Chua's circuit," *J. Circ. Syst. Comput.* **3**, 507–523.

Parlitz, U. & Mayer-Kress, G. [1995] "Predicting low-dimensional spatiotemporal dynamics using discrete wavelet transforms," *Phys. Rev.* **E51**, R2709–R2711.

Parlitz, U., Scheffczyk, C., Akhatov, I. & Lauterborn, W. [1995] "Structure formation in cavitation bubble fields," *Chaos, Solitons & Fractals* **5**, 1881–1891.

Parlitz, U. [1996] "Estimating model parameters from time series by auto-synchronization," *Phys. Rev. Lett.* **76**, 1232–1235.

Parlitz, U. & Kocarev, L. [1996] "Multichannel communication using auto-synchronization," *Int. J. Bifurcation and Chaos* **6**, 581–588.

Parlitz, U., Kocarev, L., Stojanovski, T. & Preckel, H. [1996a] "Encoding messages using chaotic synchronization," *Phys. Rev.* **E53**, 4351–4361.

Parlitz, U., Junge, L. & Kocarev, L. [1996b] "Synchronization based parameter estimation from time series," *Phys. Rev.* **E54**, 6253–6259.

Parlitz, U., Junge, L., Lauterborn, W. & Kocarev, L. [1996c] "Experimental observation of phase synchronization," *Phys. Rev.* **E54**, 2115–2117.

Pecora, L. M. & Carroll, T. L. [1990] "Synchronization in chaotic systems," *Phys. Rev. Lett.* **64**, 821–824.

Philipp, A., Delius, M., Scheffczyk, C., Vogel, A. & Lauterborn, W. [1993] "Interaction of lithotripter-generated shock waves with air bubbles," *J. Acoust. Soc. Am.* **93**, 2496–2509.

Poincaré, H. [1903] *La science et l'hypothèse.* (Bibliothèque de Philosophie scientifique, Paris). English translation: *Science and Hypothesis* (Dover, 1952).

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. [1986] *Numerical Recipes* (Cambrige University Press, Cambridge).

Provenzale, A., Smith, L. A., Vio, R. & Murante, G. [1992] "Distiguishing between low-dimensional dynamics and randomness in measured time series," *Physica* **D58**, 31–49.

Rapp, P. E., Albano, A. M. Zimmerman, I. D. & Jiménez-Moltaño, M. A. [1994] "Phase-randomized surrogates can produce spurious identifications of non-random structure," *Phys. Lett.* **A192**, 27–33.

Rayleigh, J. W. S. [1883a] "On the crispations of fluid resting upon a vibrating support," *Phil. Mag. Ser. 5,* **16**, 50–58.

Rayleigh, J. W. S. [1883b] "On maintained vibrations," *Phil. Mag., Ser. 5,* **15**, 229–235.

Rayleigh, J. W. S. [1887] "On the maintenance of vibrations by forces of double frequency, and on the propagation of waves through a medium endowed with a periodic structure," *Phil. Mag., Ser. 5,* **24**, 145–159.

Remoissenet, M. [1996] *Waves called Solitons*, 2nd edition (Springer, Berlin).

Rico-Martinez, R., Krischer, K., Kevrekidis, I. G., Kube, M. C. & Hudson, J. L. [1992] "Discrete-versus continuous-time nonlinear signal processing of Cu electrodissolution data," *Chem. Eng. Comm.* **118**, 25–48.

Riess, P. [1859] "Das Anblasen offener Rohre durch eine Flamme," *Ann. Phys. Chem.* **108**, 653–656.

Rijke, P. L. [1859] "Notiz über eine neue Art, die in einer an beiden Enden offenen Röhre enthaltene Luft in Schwingungen zu versetzen," *Ann. Phys. Chem.* **107**, 339–343.

Rosenblum, M. G., Pikovsky, A. & Kurths, J. [1996], "Phase synchronization of chaotic oscillators," *Phys. Rev. Lett.* **76**, 1804–1807.

Rosenstein, M. T., Collins, J. J. & de Luca, C. J. [1993] "A practical method for calculating largest Lyapunov exponents from small data sets," *Physica* **D65**, 117.

Rosenstein, M. T., Collins, J. J. & De Luca, C. J. [1994] "Reconstruction expansion as a geometry-based framework for choosing proper delay times," *Physica* **D73**, 82–98.

Roy, R., Murphy, T. W., Jr., Maier, T. D., Gills, Z. & Hurt, E. R. [1992] "Dynamical control of a chaotic

laser: Experimental stabilization of a globally coupled system," Phys. Rev. Lett. **68**, 1259.

Ruelle, D. & Takens, F. [1971] "On the nature of turbulence," *Commun. Math. Phys.* **20**, 167–192.

Rulkov, N. F., Sushchik, K. M., Tsimring, L. S. & Abarbanel, H. D. I. [1995] "Generalized synchronization of chaos in directionally coupled chaotic systems," *Phys. Rev.* **E51**, 980–994.

Russell, J. S. [1844] "Report on waves," *14th meeting of the British Association Report, York*, 311–390.

Russell, J. S. [1895] *The Wave of Translation in the Oceans of Water, Air and Ether* (Trubner, London).

Salvino, L. W. & Cawley, R. [1994] "Smoothness implies determinism: A method to detect it in time series," *Phys. Rev. Lett.* **73**, 1091–1094.

Sano, M. & Sawada, Y. [1985] "Measurement of the Lyapunov spectrum from a chaotic time series," *Phys. Rev. Lett.* **55**, 1082–1085.

Sato, S., Sano, M. & Sawada, Y. [1987] "Practical methods of measuring the generalized dimension and largest Lyapunov exponent in high dimensional chaotic systems," *Prog. Theor. Phys.* **77**, 1–5.

Sauer, T., Yorke, Y. & Casdagli, M. [1991] "Embedology," *J. Stat. Phys.* **65**, 579–616.

Sauer, T. & Yorke, J. A. [1993] "How many delay coordinates do you need?" *Int. J. Bifurcation and Chaos* **3**, 737–744.

Sauer, T. [1994] "Reconstruction of dynamical systems from interspike intervals," *Phys. Rev. Lett.* **72**, 3811–3814.

Savit, R. & Green, M. [1991] "Time series and dependent variables," *Physica* **D50**, 95–116.

Schuster, H. G. [1988] *Deterministic Chaos*, 2nd ed. (VCH Publishers, Weinheim).

Schweizer, J. & Kennedy, M. P. [1995] "Predictive Poincaré control," *Phys. Rev.* **E52**, 4865–4876.

Shimada, I. & Nagashima, T. [1979] "A numerical approach to ergodic problems of dissipative dynamical systems," *Prog. Theor. Phys.* **61**, 1605–1616.

Sirovich, L. [1987] "Turbulence and the dynamics of coherent structures," *Quart. Appl. Math* **45**, 561–590.

Sirovich, L. [1989] "Chaotic dynamics of coherent structures," *Physica* **D37**, 126–145.

Smith, C. W., Tejwani, M. J. & Farris, D. A. [1982] "Bifurcation universality for first-sound subharmonic generation in superfluid helium-4," *Phys. Rev. Lett.* **48**, 492–494.

Smith, L. [1992] "Identification and prediction of low dimensional dynamics," *Physica* **D58**, 50–76.

Sondhauss, C. [1850] "Über die Schallschwingungen der Luft in erhitzten Glasröhren und in gedeckten Pfeifen von ungleicher Weite," *Ann. Phys. Chem.* **79**, 1–34.

Steinecke, I. & Herzel, H. [1995] "Bifurcations in an asymmetric vocal fold model," *J. Acoust. Soc. Am.* **97**, 1874–1884.

Stojanovski, T., Kocarev, L. & Parlitz, U. [1996] "Driving and synchronising by chaotic impulses," *Phys. Rev.* **E54**, 2128–2131.

Stone, E. & Cutler, A. [1996] "Archetypal analysis of spatio-temporal dynamics," *Physica* **D90**, 209–224.

Stoop, R. & Meier, P. F. [1988] "Evaluation of Lyapunov exponents and scaling functions from time series," *J. Opt. Soc. Am.* **B5**, 1037–1045.

Stoop, R. & Parisi, J. [1991] "Calculation of Lyapunov exponents avoiding spurious elements," *Physica* **D50**, 89–94.

Swift, G. W. [1988] "Thermoacoustic engines," *J. Acoust. Soc. Am.* **84**, 1145–1180.

Swift, G. W. [1995] "Thermoacoustic engines and refrigerators," *Physics Today* **48**, July, 22–28.

Swinney, H. L. & Gollub, J. P. [1978], "The transition to turbulence," *Physics Today* **31**, 41–49.

Taconis, K. W., Beenakker, J. J. M., Nier, A. O. C. & Aldrich, L. T. [1949] "Measurements concerning the vapour liquid equilibrium of solutions of He^3 in He^4 below 2.19° K," *Physica* **15**, 733–739.

Takens, F. [1981] "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*, eds. Rand, D. A. & Young, L.-S. (Springer, Berlin), pp. 366–381.

Takens, F. [1993] "Detecting nonlinearities in stationary time series," *Int. J. Bifurcation and Chaos* **3**, 241–256.

Theiler, J. [1986] "Spurious dimension from correlation algorithms applied to limited time-series data," *Phys. Rev.* **A34**, 2427–2431.

Theiler, J. [1990] "Estimating fractal dimension," *J. Opt. Soc. Am.* **A7**, 1055–1073.

Theiler, J., Galdrikian, B., Longtin, A., Eubank, S. & Farmer, J. D. [1992a] "Using surrogate data to detect nonlinearity in time series," in *Nonlinear Modeling and Forecasting*, eds. Casdagli, M. & Eubank, S., SFI Studies in the Sciences of Complexity, Vol. XII (Addison-Wesley,Reading, MA), pp. 163–188.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Farmer, J. D. [1992b] "Testing for nonlinearity in time series: The method of surrogate data," *Physica* **D58**, 77–94.

Thompson, J. M. T. & Stewart, H. B. [1986] *Nonlinear Dynamics and Chaos* (Wiley, Chichester).

Thuras, A. L., Jenkins, R. T. & O'Neil, H. T. [1935] "Extraneous frequencies generated in air carrying intense sound waves," *J. Acoust. Soc. Am.* **6**, 173–180.

Tuffilaro, N. B. [1989] "Nonlinear and chaotic string vibrations," *Am. J. Phys.* **57**, 408–414.

Vohra, S., Spano, M., Shlesinger, M., Pecora, L. & Ditto, W. (eds.) [1992] *Proceedings of the 1st Experimental Chaos Conference* (World Scientific, Singapore).

Wayland, R., Bromley, D., Pickett, D. & Passamante, A. [1993] "Recognizing determinism in a time series," *Phys. Rev. Lett.* **70**, 580–582.

Weiss, C. O. & King, H. [1982] "Oscillation period doubling chaos in a laser," *Opt. Comm.* **44**, 59–61.

Weiss, C. O., Godone, A. & Olafsson, A. [1983] "Routes to chaotic emission in a cw He-Ne laser," *Phys. Rev.* **A28**, 892.

Weiss, C. O., Klische, W., Ering, P. S. & Cooper, M. [1985] "Instabilities and chaos of a single mode NH_3 ring laser," *Opt. Comm.* **52**, 405.

Weiss, C. O. & Vilaseca, R. [1991] *Dynamics of Lasers* (VCH, Weinheim).

Wheatley, J. & Cox, A. [1985] "Natural engines," *Phys. Today* **38**(8), 50–58.

Wolf, A., Swift, J. B., Swinney, L. & Vastano, J. A. [1985] "Determining Lyapunov exponents from a time series," *Physica* **D16**, 285–317.

Wu, C. W. & Chua, L. O. [1993] "A simple way to synchronize chaotic systems with applications to secure communication systems," *Int. J. Bifurcation and Chaos* **3**, 1619–1627.

Yazaki, T., Takashima, S. & Mizutami, F. [1987] "Complex quasiperiodic and chaotic states observed in thermally induced oscillations of gas columns," *Phys. Rev. Lett.* **58**, 1108–1111.

Yazaki, T., Sugioka, S., Mizutami, F. & Mamada, H. [1990] "Nonlinear dynamics of a forced thermoacoustic oscillation," *Phys. Rev. Lett.* **64**, 2515–2518.

Yazaki, T. [1993] "Experimental observations of thermoacoustic turbulence and universal properties of the quasiperiodic transition to chaos," *Phys. Rev.* **E48**, 1806–1816.

Young, F. R. [1989] *Cavitation* (McGraw-Hill, London).

Zabusky, N. J. & Kruskal, M. D. [1965] "Interaction of 'solitons' in a collisionless plasma and the recurrence of initial states," *Phys. Rev. Lett.* **15**, 240–243.

Zeng, X., Eykholt, R. & Pielke, R. A. [1991] "Estimating the Lyapunov-exponent spectrum from short time series of low precision," *Phys. Rev. Lett.* **66**, 3229–3232.

Zeng, X., Pielke, R. A. & Eykholt, R. [1992] "Extracting Lyapunov exponents from short time series of low precision," *Modern Phys. Lett.* **B6**, 55–75.

Subharmonic Route to Chaos Observed in Acoustics

Werner Lauterborn and Eckehart Cramer[a]

Third Physical Institute, University of Göttingen, D-3400 Göttingen, Federal Republic of Germany

(Received 14 September 1981)

A subharmonic route to chaos including period-doubling bifurcations up to $f/8$ has been observed in experiments on acoustical turbulence (acoustic cavitation noise). The system also shows signs of reverse bifurcation with increasing control parameter (acoustic driving pressure amplitude). In view of the large variety of phenomena observed and yet to be expected the system investigated may well serve as a further experimental paradigm of nonlinear dynamical systems besides Rayleigh-Bénard and circular couette flow.

There is increasing evidence that period-doubling bifurcations[1] and strange attractors[2] are common phenomena for a large class of nonlinear dynamical systems. Most of this evidence stems from relatively simple mathematical models like the three-variable differential systems of Lorenz[3] and Rössler[4] and one-dimensional iterated maps on the unit interval[5] which show links to dynamical systems via the Poincaré return map. The discovery of universal properties in period-doubling bifurcations of iterated maps by Feigenbaum[1] could be confirmed for the Lorenz model[6] and a five-variable model of the Navier-Stokes equation,[7] and has stimulated the search for additional universal features of nonlinear dynamical systems.[8-11]

Compared with the large body of theoretical work, experiments are rather sparse. Up to now there are only two physical systems where the onset of chaos is studied systematically and which show some analogy to the behavior of the above mathematical models. These are the Rayleigh-Bénard experiment on the flow in a flat convective layer of liquid heated from below[12-14] and experiments on circular couette flow (flow between two cylinders, the inner one rotating).[15,16] Experiments in other fields are just emerging, such as in optics for optically bistable cavities[17] or proposed similar experiments with noisy Josephson junctions,[18] charge-density waves in anisotropic solids and superionic conductors,[19] or pinned dislocation lines.[20]

This paper presents experiments in acoustics which in view of the results to be reported and in analogy to the newly coined terms of "optical turbulence"[17] and "solid-state turbulence"[19,20] may be called experiments on "acoustical turbulence." The experiment consists in irradiating a liquid (water) with sound of high intensity (control parameter is the sound pressure amplitude) and looking for the sound output of the liquid,

called acoustic cavitation noise. The physical situation is a somehow fundamental one: The transport of acoustical energy through a liquid is considered. It bears much resemblance with the Rayleigh-Bénard problem where the transport of heat through a liquid is investigated.

To irradiate the liquid a piezoelectric cylinder of 76-mm length, 76-mm inner diameter, and 5-mm wall thickness is used. When driven at its main resonance at 21.56 kHz a high-intensity acoustic field is generated in the interior, and cavitation is easily achieved. The noise is picked up by a broadband microphone[21] and digitized at rates up to 2 MHz after suitable low-pass filtering (to avoid aliasing in the subsequent Fourier analysis) and strong filtering of the driving frequency (to be able to store the noise with just an 8-bit storage). Sound pressure power spectra are calculated via the fast-Fourier-transform algorithm from usually 4K samples out of the 128K storage available. More details of the experimental setup are given elsewhere.[22]

Power spectra of acoustic cavitation noise usually consist of instrumentally sharp lines on a noise background. The lines are related to the driving frequency f_0 and lie at $(n/m)f_0$ ($n, m = 1, 2, 3, \ldots$). Of special interest are the lines at $m \geq 2$, $n < m$, i.e., in the subharmonic region $f < f_0$ of the spectrum. In early experiments the occurrence of lines at $f_0/2$, $f_0/3$, and $f_0/4$ has already been found,[23,24] but no convincing explanation could be given. The explanation that bubbles in water driven at twice their natural resonance are responsible for the $f_0/2$ line[25] had to be abandoned since bubbles of the necessary size could not be found and are unlikely to be present in the experimental situation.[26] Instead, after an extensive numerical investigation of bubble oscillations, it has been argued that special ultraharmonic resonances of bubbles smaller than resonant size (especially the $\frac{3}{2}$ resonance where two

oscillations of the driving sound field match three oscillations of the natural oscillation frequency of the bubble) are responsible for the subharmonic line at $f_0/2$ in the power spectrum,[26] but a direct verification of this hypothesis has not yet been possible.

The present experiments were undertaken to add to our understanding of the subharmonic line problem. They differ from previous experiments in that they are fully computer controlled to realize almost any desired control parameter history.

Figure 1 gives just one example of a pressure power spectrum obtained at a driving voltage of 15 V. The history in this case was to linearly increase the driving voltage to 15 V and then to stay there for some time. These precautions are necessary to arrive at the third period-doubling bifurcation with lines at $nf_0/8$. But even when staying at a constant voltage large fluctuations are observed, and usually only spectra with subharmonics as low as $f_0/4$ can be observed. Figure 1 suggests that our nonlinear acoustical system may follow the period-doubling route to chaos and may belong to the universal class of Feigenbaum systems. Unfortunately this could not yet be proved as the next bifurcation with lines at $nf_0/16$ could not be reached. There may be principal difficulties in achieving this aim because our system can be expected to be a noisy one and this has been shown to limit the bifurcation sequence.[8]

As a result of the total computerization of the experiment we are able to do complex measurements on acoustical turbulence and to gather enormous amounts of data. The question therefore arises of how to present the results. One way that we found very appealing is in the form of grey-scale pictures analogous to "visible speech" where the power spectrum is plotted versus time with the amplitude encoded as grey scale. We have adopted this kind of presentation for our studies of how the system reaches chaos and have plotted the pressure power spectrum as a function of the voltage at the driving transducer. Figure 2 gives an example for the case where the voltage is increased linearly from 0 to 60 V in about 250 ms. During this time 128K samples of the pressure in the liquid are taken at a rate of 500 kHz. The total experiment thus lasts just a quarter of a second. From these data about 1000 overlapping short-time spectra are calculated with 4K data each and a shift of 128 samples from one spectrum to the next. In Fig. 2 three successive spectra are combined to give a total of 370 spectra. The grey level is encoded with the aid of a 3×3 matrix so that a binary plotter can be used.

FIG. 1. Example of a pressure power spectrum of acoustic cavitation noise with subharmonic lines as low as $f_0/8$ ($f_0 = 22.56$ kHz), i.e., three period-doubling bifurcations have taken place. The driving frequency is strongly filtered.
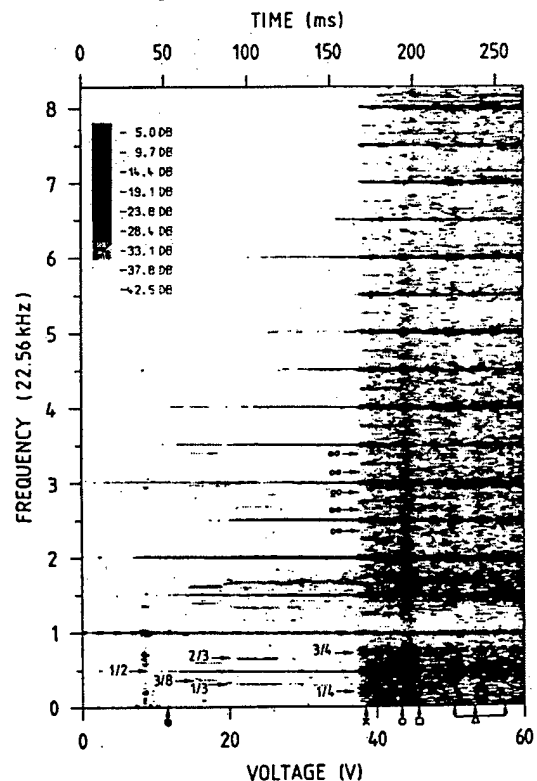

FIG. 2. Sequence of power spectra displayed as a grey-scale picture. The voltage at the driving piezoelectric cylinder is increased linearly from 0 to 60 V in 262 ms.

Many interesting features are immediately visible in Fig. 2:

(1) The first period-doubling bifurcation sets in at about 12 V (see closed circle in Fig. 2).

(2) Further subharmonic lines appear in a process that does not look like period doubling. These lines seem to detach from the $f_0/2$ (and $3f_0/2$) line with stops at $3f_0/8$ ($5f_0/8$) and $f_0/3$ ($2f_0/3$) until suddenly and with large amplitude the $f_0/4$ line and its odd harmonics set in (see 1/2, 3/8, 1/3, 2/3, 1/4, 3/4, and cross in Fig. 2).

(3) The $f_0/4$ bifurcation occurs together with a marked increase in broadband noise and a broadening of the otherwise sharp line spectrum (from cross to open circle in Fig. 2; the white gap around f_0 is due to the band reject filter needed to suppress the otherwise dominating driving frequency).

(4) At about 43 V a further increase in the broadband noise level is observed with the line spectrum still detectable and additional lines at certain odd harmonics of $f_0/8$ (see from open circle to open square and ∞ in Fig. 2). This state of the system ranges from 43 to 46 V only and is the most chaotic state encountered (chaos defined in terms of broadband noise at high level in the power spectrum).

(5) In a process looking like reverse bifurcation,[27] the system returns to a line spectrum with lines only at $f_0/2$ and its harmonics and with much less broadband noise (open square in Fig. 2).

(6) Satellite lines appear around $f_0/2$ and $3f_0/2$ (and also some other lines) which show some periodicity with the control parameter (open triangle in Fig. 2).

A general observation from other experiments is that the state of total chaos seems to be unstable since it cannot be sustained by the driving sound field for a longer period of time. Instead, oscillations are observed between the background noise and the line strength. Also single lines may visibly oscillate. This is best seen in films produced from sequences of up to 2000 power spectra plotted on a graphic display (cathode-ray tube) and filmed with a 16-mm film camera.

The picture of Fig. 2 is a rather condensed form of looking at the properties of a dynamical system. It would be very interesting to see other dynamical systems like those of Lorenz[3] and Rössler[4] as well as the experimental ones of Rayleigh-Bénard and circular couette flow displayed in this way.

Some effort has been spent in modeling the experiments theoretically. As a first step single spherical bubbles have been taken and subjected to a sinusoidal driving pressure of increasing amplitude. The mathematical model is a highly nonlinear ordinary differential equation of second order for the radius of the bubble as a function of time and includes surface tension, viscosity, and compressibility of the liquid (water).[28] Radius-time curves have been calculated numerically for different bubble sizes, sound pressure amplitudes, and frequencies to get response curves for this nonlinear system.[28] To simulate the present experiments bubble wall oscillation power spectra have been calculated and plotted in the manner of Fig. 2 as a function of the driving pressure. Qualitatively similar behavior is observed but strongly depends on the bubble size. Of most importance seems to be the observation that bifurcations are obtained, mostly from f_0 to $f_0/2$ to $f_0/3$ (or also $f_0/4$) to (quickly) $f_0/4$ to chaos and then back to f_0 or $f_0/2$ (directly). Thus both successive bifurcations as well as reverse bifurcations are observed in this simple mathematical model, like those encountered in the experiments.

More sophisticated models must take into account that many bubbles are present in the liquid which all couple via their sound radiation. A fuller discussion of these questions as well as of the physical processes responsible for the observed phenomena is given in Ref. 22.

We arrive at the conclusion that in our acoustic system there is a subharmonic route to turbulence, but one which does not solely proceed via successive period-doubling bifurcations although this route is strongly involved. We therefore propose to make a distinction between a subharmonic and a period-doubling route to chaos. Moreover, our system shows signs of reverse bifurcations which may be worthwhile to study in greater detail.

---

[1]M. J. Feigenbaum, J. Stat. Phys. 19, 25 (1978).
[2]D. Ruelle and F. Takens, Commun. Math. Phys. 20,

167 (1971).

[3]E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).

[4]O. E. Rössler, Phys. Lett. **57A**, 397 (1976).

[5]P. Collet and J. P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems* (Birkhäuser, Basel, 1980).

[6]V. Franceschini, J. Stat. Phys. **22**, 397 (1980).

[7]V. Franceschini and C. Tebaldi, J. Stat. Phys. **21**, 707 (1979).

[8]J. P. Crutchfield and B. A. Huberman, Phys. Lett. **77A**, 407 (1980).

[9]B. A. Huberman and A. B. Zisook, Phys. Rev. Lett. **46**, 626 (1981).

[10]J. P. Crutchfield, M. Nauenburg, and J. Rudnick, Phys. Rev. Lett. **46**, 933 (1981).

[11]B. Shralman, C. E. Wayne, and P. C. Martin, Phys. Rev. Lett. **46**, 935 (1981).

[12]A. Libchaber and J. Maurer, J. Phys. (Paris), Colloq. **41**, C3-51 (1980).

[13]J. P. Gollub and S. V. Benson, J. Fluid Mech. **100**, 449 (1980).

[14]M. Giglio, S. Musazzi, and U. Perini, Phys. Rev. Lett. **47**, 243 (1981).

[15]Yu. N. Belyąev, A. A. Monakhov, S. A. Shcherbakov, and I. M. Yavorskaya, Pis'ma Zh. Eksp. Teor. Fiz. **29**, 329 (1979) [JETP Lett. **29**, 295 (1979)].

[16]P. R. Fenstermacher, H. L. Swinney, and J. P. Gollub, J. Fluid Mech. **94**, 103 (1979).

[17]H. M. Gibbs, F. A. Hopf, D. L. Kaplan, and R. L. Shoemaker, Phys. Rev. Lett. **46**, 474 (1981).

[18]B. A. Huberman, J. P. Crutchfield, and N. H. Packard, Appl. Phys. Lett. **37**, 750 (1980).

[19]B. A. Huberman and J. P. Crutchfield, Phys. Rev. Lett. **43**, 1743 (1979).

[20]C. Herring and B. A. Huberman, Appl. Phys. Lett. **36**, 975 (1980).

[21]Our own construction, built in our laboratory by G. Heinrich. It is made up of a thin disk of PZT-4 material bonded on a brass rod, with cutoff frequency of about 1 MHz.

[22]W. Lauterborn and E. Cramer, to be published.

[23]R. Esche, Acustica, Akust. Belhe. **2**, 208 (1952).

[24]L. Bohn, Acustica **7**, 201 (1957).

[25]W. Güth, Acustica **6**, 532 (1956).

[26]W. Lauterborn, J. Acoust. Soc. Am. **59**, 283 (1976).

[27]J. Crutchfield, D. Farmer, N. Packard, R. Shaw, G. Jones, and R. J. Donelly, Phys. Lett. **76A**, 1 (1980).

[28]E. Cramer, in *Cavitation and Inhomogenities in Underwater Acoustics*, edited by W. Lauterborn (Springer-Verlag, New York, 1980).

The Role of Surface Tension in Stable Single-Bubble Sonoluminescence

I. Akhatov* and N. Gumerov

Department of Continuous Media Mechanics, Bashkir University, 32 Frunze Str., Ufa 450074, Russian Federation

C. D. Ohl, U. Parlitz,† and W. Lauterborn

Drittes Physikalisches Institut, Universität Göttingen, Bürgerstraße 42-44, D-37073 Göttingen, Germany

A theory for stable bubble oscillations in high pressure sound fields is presented. It is based on the strong influence of the surface tension on the dynamics of small bubbles and takes into account rectified diffusion and the resonancelike response of small bubbles to very strong acoustic pressure amplitudes. This theory provides an explanation for the existence of small, stably oscillating bubbles that have been observed in experiments on sonoluminescence. [S0031-9007(96)02169-2]

PACS numbers: 43.25.+y, 43.35.+d, 47.55.Bx, 78.60.Mq

Bubbles in a liquid that are subject to an external sound field not only oscillate strongly nonlinearly but may also emit light. This phenomenon is called *sonoluminescence* (SL) and was discovered by Marinesco and Trillat [1] in 1933. Since then it has been investigated experimentally as well as theoretically by many authors [2–12]. The interest in SL was restimulated by the elaborate experiments of Gaitan *et al.* [6], who investigated SL of a single bubble in water trapped by a strong periodical acoustic field. This phenomenon is called *single-bubble sonoluminescence* (SBSL) and was investigated in a number of papers [7–11].

One of the most striking results of SBSL experiments was the observation that bubbles can oscillate permanently for several days without dissolution and without changing their size. A detailed investigation of the underlying physical mechanism shows that there are many effects and phenomena that have to be taken into account. This list includes rectified diffusion, surface tension, dissolved gas, thermoconductivity, acoustic radiation, viscosity, microstreaming around a bubble (maybe generated by surface waves on the bubble [13]), the nonlinear character of bubble oscillations, and the fragmentation and coalescence of bubbles [14–17]. The analysis of experimental and theoretical results shows that the main effects that are important for the generation of stable cavitation bubbles are rectified diffusion, surface tension, and shape oscillations. The mechanisms of how gas diffusion may lead to stability of the bubble size were discussed in [18] in order to analyze the possibility of multiple stable equilibrium radii of the bubble. The stability of shape oscillations was investigated in [19,20] where it was shown that large bubbles have a tendency to disintegrate due to unstable surface oscillations. In this Letter we focus on the influence of rectified diffusion and surface tension on the stability of small bubbles in a sound field.

Without external sound field, bubbles of any size are unstable because the pressure inside the bubble is larger than in the liquid, and therefore the bubble will dissolve slowly due to a continuous mass flux from the interior of the bubble into the liquid. In the presence of a periodic acoustic field the bubble starts to oscillate. During the expansion period gas diffuses from the liquid into the bubble, and during the contraction cycle the diffusion process takes place in the opposite direction. There is a net flow of gas into the bubble because the area of the bubble wall is greater during the expansion period and therefore more gas will enter during the expansion than will leave during the contraction cycle. This phenomenon is called *rectified diffusion* and leads to a growth of the bubble [21,22]. For small amplitudes it was shown [23,24] that the growth rate depends on the sound field amplitude P_a, the resonance radius R_r, and the equilibrium radius R_0 of the bubble. *This growth rate is closely correlated with the response curve that describes the dependence of the maximum bubble radius on the equilibrium radius.* The theory for weakly nonlinear oscillations gives a good description of the growth and dissolution processes of sufficiently large bubbles and small pressure amplitudes. However, it provides no answer to the questions pertaining to (stable) sonoluminescence: *Why are small gas bubbles in a liquid stable in the presence of a strong sound field?*

For single-bubble oscillations under medium and large pressure amplitudes a complicated scenario of bifurcations and coexisting (chaotic) attractors exists [25,26]. Our numerical simulations show, however, that for very small bubbles in very strong sound fields the dynamics becomes more regular and a new type of strong resonance with a thresholdlike increase in oscillation amplitude occurs [12,27]. The physical reason for this phenomenon is the fact that for very small bubbles the surface tension pressure $P_\sigma = 2\sigma/R_0$ is very high and the bubbles behave like flexible solid particles even for large driving pressures. One cycle of a typical bubble oscillation for this case is shown in Fig. 1(b) for the normalized bubble radius $R(t)/R_0$. Figure 1(a) shows the driving pressure of the external sound field $p_a(t) = -P_a \sin(\omega t)$. When
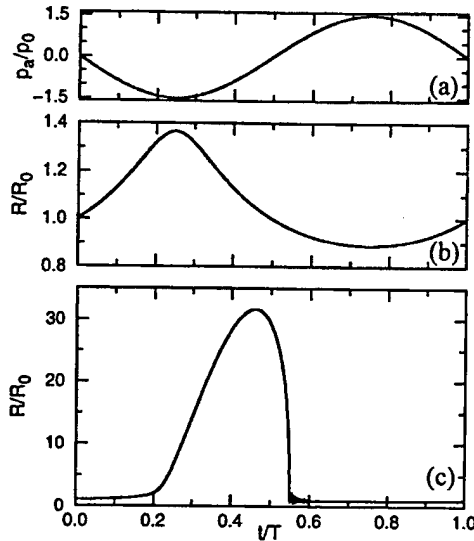
FIG. 1. The influence of surface tension on bubble oscillations. Plotted is one period of oscillation (frequency $\nu = \omega/2\pi = 20$ kHz) of the normalized sound field pressure $p_a(t)/p_0$ (a) and of the normalized bubble radius $R(t)/R_0$ vs time t (b), (c) for a pressure amplitude of $P_a = 1.5$ bar. (b) A bubble with equilibrium radius $R_0 = 1$ μm oscillates with small amplitude due to the relatively high surface tension pressure $P_\sigma = 2\sigma/R_0$. (c) For $R_0 = 1.5$ μm strongly nonlinear oscillations with high amplitude occur.

FIG. 2. (a) Response curves showing the normalized maximum bubble radius R_m/R_0 vs the equilibrium radius R_0 for different pressure amplitudes $P_a = 1.1$–1.5 bar. (b) Illustration of the cavitation threshold showing the nonlinearly averaged concentration near the bubble wall $\langle \bar{c} \rangle_\tau$ vs the equilibrium radius R_0 for different pressure amplitudes $P_a = 1.1$–1.5 bar. The horizontal dashed lines denote different levels of gas concentrations $\bar{c}_\infty$ in the liquid. The open circles denote the cavitation threshold value R_{th} of the radius and the filled circle the stable bubble radius R_s.

we increase the size of the bubble R_0 it starts to oscillate differently. During the expansion period the influence of the surface tension decreases rapidly, and therefore the amplitude of the expansion grows enormously leading to a strong collapse. The transition point may be called *nonstatic Blake threshold* [3,28]. The kind of oscillation present beyond this threshold is shown in Fig. 1(c). The results given in Fig. 1 and in all following figures have been computed using the Keller-Miksis model [26,29]:

$$\left(1 - \frac{\dot{R}}{C_l}\right) R \ddot{R} + \frac{3}{2} \dot{R}^2 \left(1 - \frac{\dot{R}}{3C_l}\right) = \left(1 + \frac{\dot{R}}{C_l}\right) \frac{P}{\rho} + \frac{R}{\rho C_l} \frac{dP}{dt},$$

with

$$P = \left(p_0 + \frac{2\sigma}{R_0}\right)\left(\frac{R_0}{R}\right)^{3\kappa} - p_0 - \frac{2\sigma}{R} - \frac{4\mu}{R} \dot{R} - p_a(t),$$

for air bubbles in water at 20°C with $\kappa = 1.4$, $\sigma = 0.0725$ N/m, $p_0 = 1$ bar, $C_l = 1500$ m/s, and a driving frequency of $\omega = 2\pi \cdot 20$ kHz. Qualitatively the same results have been obtained for the Gilmore model [30].

When the equilibrium radius of the bubble is increased further the influence of the surface tension pressure P_σ becomes smaller and a nonmonotonous resonance curve for the normalized radius R_m/R_0 occurs. These response

curves are shown in Fig. 2(a) for different values of the pressure P_a. One can see that the nonmonotonous behavior starts for $P_a > 1.2$ bar.

Now we investigate the rectified diffusion in the strong resonance region for small bubbles. The theoretical formulation for mass transport across the dynamic interface associated with a spherical bubble undergoing volume oscillations was derived in Ref. [31]. The equations governing the convection and diffusion of dissolved gas in the liquid outside a spherical bubble can been written in the following form:

$$\frac{\partial c}{\partial t} + \frac{R^2(t)\dot{R}(t)}{r^2} \frac{\partial c}{\partial r} = \frac{D}{r^2} \frac{\partial}{\partial r}\left(r^2 \frac{\partial c}{\partial r}\right), \quad (1)$$

$$c|_{r=R(t)} = H\left(p_0 + \frac{2\sigma}{R_0}\right)\left(\frac{R(t)}{R_0}\right)^{-3\kappa}, \quad c|_{r=\infty} = c_\infty, \quad (2)$$

$$\frac{dm}{dt} = 4\pi R^2(t) D \frac{\partial c}{\partial r}\bigg|_{r=R(t)}. \quad (3)$$

Here Eq. (1) describes the convective diffusion, where $R(t)$ is the bubble radius governed by some dynamical equation for bubble oscillations, c is the mass concentration of gas dissolved in the liquid, and D is the diffusivity of the gas in the liquid. Equation (2) gives the boundary

conditions at the bubble surface and far from it provided by Henry's law, which relates the concentration of gas in a liquid to the partial pressure of the gas above the liquid. The symbol H denotes Henry's constant and c_∞ is the initial uniform gas concentration in the fluid where the bubble is assumed to be created. Equation (3) describes the rate of the gas transport across the bubble interface.

This mass transport problem was solved approximately for large Peclet numbers [31] (Pe $= R_0^2 \omega/D \gg 1$). In this case Eqs. (1)–(3) can be simplified by a transformation of the problem into normalized Lagrangian coordinates $\eta = [r^3 - R^3(t)]/3R_0^3$ to avoid difficulties because of the moving boundary conditions. Another analytical difficulty—oscillatory behavior of the concentration close to the bubble surface and slow diffusion behavior farther away from the bubble—was solved by splitting the problem into two parts. Finally, it was shown that the time averaged rate of mass transport in the case of any non-linear periodic bubble oscillation may be approximated as follows:

$$\frac{d\bar{m}}{d\tau} = \frac{\bar{c}_\infty - \langle\bar{c}\rangle_\tau}{T_{rd}}, \quad T_{rd} = \int_0^\infty \frac{d\eta}{\langle[3\eta + \bar{R}^3(t)]^{4/3}\rangle},$$

$$(4)$$

$$\bar{c} = \frac{c(R(t),t)}{T_{rd}}, \quad \bar{c}_\infty = \frac{c_\infty}{c_0}, \quad \bar{m} = \frac{m}{m_0}, \quad \bar{R} = \frac{R}{R_0}.$$

Here c_0 is the saturation concentration in the liquid separated from gas at pressure p_0 by a plane boundary and m_0 is the mass of the gas soluted in the liquid displaced by the undisturbed bubble. The variable $\tau = tD/R_0^2$ is the slow diffusion time scale, and T_{rd} is the dimensionless characteristic time of rectified diffusion mass growth rate of the bubble.

In this approach two different averaging procedures are used. First, ordinary averaging over the period T of the acoustic field

$$\langle f(t)\rangle = \frac{1}{T}\int_0^T f(t)\,dt;$$

and second, a "nonlinear averaging" procedure in a specific nonlinear time scale

$$\langle f(t)\rangle_\tau = \frac{1}{\int_0^T R^4(t)\,dt}\int_0^T R^4(t)f(t)\,dt,$$

where the radius evolution $R(t)$ is used for computing the gas concentration near the bubble wall $\langle\bar{c}\rangle_\tau$. This approach was also used by Brenner et al. [18] to investigate the multiple stable equilibrium radii of the bubble for medium pressure amplitudes. Here we use it to consider the stability problem in the case of very small bubbles in a strong acoustic field.

Figure 2(b) shows the averaged gas concentration $\langle\bar{c}\rangle_\tau$ near the bubble wall vs the equilibrium radius R_0 of the bubble for different amplitudes P_a of the acoustic field. For small and medium values of P_a the concentration decreases monotonically as a function of the equilibrium

radius. For sufficiently large amplitudes P_a, however, the corresponding concentration curves possess a global minimum for small bubble radii. This nonmonotonous dependence of $\langle\bar{c}\rangle_\tau$ on R_0 is a result of the strong resonance shown in Fig. 2(a).

Since the characteristic time T_{rd} in Eq. (4) is always positive the evolution of the mean bubble mass depends only on the difference between the concentration of gas in the liquid $\bar{c}_\infty$ and the nonlinearly averaged concentration near the bubble wall $\langle\bar{c}\rangle_\tau$. As can be seen in Fig. 2(b) there are two possible scenarios. If $\bar{c}_\infty$ is large [upper dashed line in Fig. 2(b)] a single equilibrium point $\langle\bar{c}\rangle_\tau = \bar{c}_\infty$ exists that is unstable. This case is denoted in Fig. 2(b) by the open circle at the point of intersection of the upper dashed line with the concentration curve for $P_a = 1.2$ bar. The unstable equilibrium provides a threshold value R_{th} for the bubble radius. Bubbles with radius $R_0 < R_{th}$ dissolve due to diffusion flux from the bubble into the liquid. On the other hand, bubbles with $R_0 > R_{th}$ grow permanently due to rectified diffusion until they become very large (and may disintegrate).

If $\bar{c}_\infty$ is small [lower dashed line in Fig. 2(b)] two equilibrium points $\langle\bar{c}\rangle_\tau = \bar{c}_\infty$ exist. The left fixed point denoted by the open circle in Fig. 2(b) is unstable and closely related to the previous case. It defines the cavitation threshold radius R_{th}. The equilibrium point plotted as a filled circle at the right hand side in Fig. 2(b) is stable and provides a stable radius R_s for single bubbles oscillating in the acoustic field. Bubbles with radius $R_{th} < R < R_s$ grow until they reach the stable radius R_s. If the bubble radius is larger than R_s the bubble shrinks until $R_0 = R_s$. A necessary condition for the existence of R_s is a nonmonotonous dependence of the nonlinearly averaged concentration $\langle\bar{c}\rangle_\tau$ on the equilibrium radius R_0 for small bubbles. Only in this case a range of $\bar{c}_\infty$ values exists such that stable bubble oscillations are possible. For given small values of the concentration of gas in the liquid $\bar{c}_\infty$ there exists a lower threshold value P_a^c for the pressure amplitudes P_a that leads to stable bubbles as can be seen in Fig. 3 showing the dependence of the threshold value R_{th} (dashed curve) and the stable bubble radius R_s (solid curve) on the pressure amplitude P_a. For small values of P_a bubbles of any size will dissolve. At some critical value P_a^c the stable bubble radius R_s occurs due to a saddle-node bifurcation. When the pressure P_a is increased further the value of R_s increases, the bubble becomes very large and will eventually be destroyed due to dynamical instabilities (e.g., surface oscillations). Such a finite pressure range for stable bubble oscillations has also been observed experimentally [6].

In this Letter a theory for stable bubble oscillations in high pressure sound fields has been presented that is based on the strong influence of the surface tension on the dynamics of small bubbles. The approach presented takes into account the interaction of two effects: rectified diffusion and the resonancelike response of small bubbles on
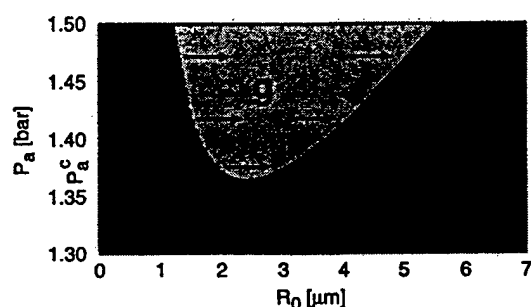
FIG. 3. Cavitation threshold radius R_{th} (dashed curve) and stable bubble radius R_s (solid curve) in dependence on the pressure amplitude P_a for fixed small value of the gas concentration $\bar{c}_\infty$ corresponding to the lower dashed line in Fig. 2(b). The islands of growth (light shading) in the sea of dissolution (dark shading) is a necessary prerequisite for stable bubble oscillations and thus SBSL to occur.

very strong acoustic pressure amplitudes due to the surface tension pressure. The results provide an explanation for the existence of small, stably oscillating bubbles that have been observed in experiments on sonoluminescence.

During the reviewing process a paper appeared where the occurrence of sonoluminescence in the parameter space (R_0, P_a) has been determined experimentally for just the case studied here, the small bubble, large acoustic pressure amplitude limit [32]. There one main finding is that sonoluminescence experimentally occurs along a *line* in the (R_0, P_a) parameter space, just as predicted by our Fig. 3 (solid line). We predict and with the measurement technique of Holt and Gaitan [32] it can be checked that this line starts at a bifurcation value where the stable and unstable branch of bubble oscillations equilibrium meet opening up an island of growth in the sea of dissolution. Both theory and experiment fall short in explaining the low gas content in the bubble. This is not at all astonishing as a wealth of other phenomena are not considered, for instance pertaining to the chemistry inside the bubble.

---

*Electronic address: iskander@ncan.bashkiria.su
†Electronic address: ulli@physik3.gwdg.de

[1] N. Marinesco and J. J. Trillat, Proc. R. Acad. Sci. **196**, 858 (1933).
[2] H. Kuttruff, Acustica **12**, 230 (1962).
[3] A. J. Walton and G. T. Reynolds, Adv. Phys. **33**, 595 (1984).

[4] K. S. Suslick, Science **247**, 1439 (1990).
[5] B. P. Barber and S. J. Putterman, Nature (London) **352**, 318 (1991).
[6] D. F. Gaitan, L. A. Crum, C. C. Church, and R. A. Roy, J. Acoust. Soc. Am. **91**, 3166 (1992).
[7] R. Löfstedt, B. P. Barber, and S. Putterman, Phys. Fluids A **5**, 2911 (1993).
[8] R. G. Holt, D. F. Gaitan, A. A. Atchley, and J. Holzfuss, Phys. Rev. Lett. **72**, 1376 (1994).
[9] L. A. Crum, Phys. Today **47**, 22 (1994); L. A. Crum and S. Cordry, in *Bubble Dynamics and Interface Phenomena*, edited by J. R. Blake *et al.* (Kluwer, Dordrecht, 1994).
[10] B. P. Barber, K. Weninger, R. Löfstedt, and S. Putterman, Phys. Rev. Lett. **74**, 5276 (1995).
[11] R. Löfstedt, K. Weninger, S. Putterman, and B. P. Barber, Phys. Rev. E **51**, 4400 (1995).
[12] C. D. Ohl, H.-C. Nägerl, R. Blatt, and W. Lauterborn, *DAGA—Fortschritte der Akustik*, edited by T. Portele and W. Hess (DEAG e.V., Oldenburg, 1996), pp. 418–419.
[13] C. C. Church, J. Acoust. Soc. Am. **83**, 2210 (1988).
[14] *Cavitation and Inhomogeneities in Underwater Acoustics*, edited by W. Lauterborn (Springer, Berlin, 1980).
[15] F. R. Young, *Cavitation* (McGraw-Hill Book Company, London, 1989).
[16] T. G. Leighton, *The Acoustic Bubble* (Academic Press, London, 1994).
[17] R. I. Nigmatulin, *Dynamics of Multiphase Media* (Hemisphere Publishing Corporation, New York, 1991), Vols. 1,2.
[18] M. P. Brenner, D. Lohse, D. Oxtoby, and T. F. Dupont, Phys. Rev. Lett. **76**, 1158 (1996).
[19] H. W. Strube, Acustica **25**, 289 (1971).
[20] M. P. Brenner, D. Lohse, and T. F. Dupont, Phys. Rev. Lett. **75**, 954 (1995).
[21] D.-Y. Hsieh and M. S. Plesset, J. Acoust. Soc. Am. **33**, 206 (1961).
[22] A. Eller and H. G. Flynn, J. Acoust. Soc. Am. **37**, 493 (1963).
[23] L. A. Crum and G. M. Hansen, J. Acoust. Soc. Am. **72**, 1586 (1982).
[24] N. A. Gumerov, Prikl. Mat. Mekh. **55**, 2256 (1991).
[25] W. Lauterborn, J. Acoust. Soc. Am. **59**, 283 (1976).
[26] U. Parlitz, V. Englisch, C. Scheffczyk, and W. Lauterborn, J. Acoust. Soc. Am. **88**, 1061 (1990).
[27] W. Lauterborn, Acustica **20**, 105 (1968).
[28] F. G. Blake, Technical Memo 12, Acoustics Research Laboratory, Harvard University, 1949.
[29] J. B. Keller and M. Miksis, J. Acoust. Soc. Am. **68**, 628 (1980).
[30] F. R. Gilmore, Hydrodynamics Laboratory, California Institute of Technology, Pasadena, California, Report No. 26-4, 1952.
[31] M. M. Fyrillas and A. J. Szeri, J. Fluid Mech. **277**, 381 (1994).
[32] R. G. Holt and D. F. Gaitan, Phys. Rev. Lett. **77**, 3791 (1996).

Bifurcation Superstructure in a Model of Acoustic Turbulence

Werner Lauterborn and Engelbert Suchla

Drittes Physikalisches Institut, Universität Göttingen, D-3400 Göttingen, Federal Republic of Germany

A model of acoustic turbulence is investigated for its bifurcation structure by the calculation of spectral as well as ordinary bifurcation diagrams and (subharmonic) attractor maps. A superstructure resulting from nonlinear resonances is found with period-doubling Feigenbaum direct cascades and Grossmann inverse cascades as fine structure. Connected with the superstructure is a new family of periodic chaos with a different type of chaos belonging to each basic subharmonic period of oscillation.

When a liquid is irradiated with sound of high intensity it may rupture to form bubbles or cavities. The phenomenon is called acoustic cavitation[1] and is accompanied by intense noise emission, the acoustic cavitation noise. The noise seems not to be due to a statistical rupture process in the liquid but to be of deterministic origin.[2] The evidence stems from the subharmonic route to chaos (broadband acoustic noise) which is observed when the sound pressure amplitude (taken as the control parameter) is raised. Thus an example of acoustic turbulence had been found experimentally.[2]

The question left is how to describe the experiment theoretically, in order to get a deeper understanding of the nature of the phenomenon. This paper investigates a theoretical model which, although relatively simple, displays diverse routes to and through chaos and complicated bifurcation structures. As is made evident below these may be explained as resulting from a superstructure of nonlinear hysteretic resonances with Feigenbaum[3] direct and Grossmann[4] inverse cascades as fine structure. In the chaotic regions the solutions of the model can also be transformed via subharmonic Poincaré maps into one-dimensional quadratic-looking maps. This connection is considered the main proof that the model belongs to the class of deterministically chaotic systems.[5]

The model is obtained by a set of simplifying assumptions to make it tractable. In the experiment, once the rupture process has started, thousands of tiny bubbles are generated, oscillating and moving in a complicated manner.[6] As a first approximation the mutual interaction of the bubbles is neglected. Then only the dynamics of a single bubble in a sound field needs to be considered. As a further approximation translational motions are neglected, and the bubble is taken as spherical. Even this system is far too complex to be written down easily as a result of heat and mass transfer across the boundary of the bubble. When these effects are neglected a reasonable model for an oscillating spherical bubble in a cold liquid can be formulated.[7] It is given by a highly nonlinear ordinary differential equation of second order for the radius R of the bubble as a function of time t:

$$R\left(1-\frac{U}{C}\right)\frac{d^2R}{dt^2}+\frac{3}{2}\left(1-\frac{U}{3C}\right)\left(\frac{dR}{dt}\right)^2-\left(1+\frac{U}{C}\right)H-\frac{U}{C}\left(1-\frac{U}{C}\right)R\frac{dH}{dR}=0, \quad \frac{dR}{dt}=U. \tag{1}$$

U is the bubble wall velocity, C is the sound velocity at the bubble wall, and H is the free enthalpy which for water is given by

$$H=\frac{n}{n-1}\frac{A^{1/n}}{\rho_0}\{[p(R)+B]^{(n-1)/n}-p_\infty+B)^{(n-1)/n}\}. \tag{2}$$

A, B, and n are constants ($A=3001$ bars, $B=3000$ bars, and $n=7$), and ρ_0 is the density at p_0 ($\rho_0=0.998$ g cm^{-3}, $p_0=1$ bar). The pressure $p(R)$ at the bubble wall is given by

$$p(R)=\left(p_0+\frac{2\sigma}{R}\right)\left(\frac{R_n}{R}\right)^{3\gamma}-\frac{2\sigma}{R}-\frac{4\mu U}{R}, \tag{3}$$

where p_0 is the static external pressure (1 bar), R_n the equilibrium radius of the bubble, σ the surface tension (72.5 dyn cm^{-1}), γ the ratio of the specific heats of the gas in the bubble (1.33), and μ the viscosity of the liquid (0.01 P). The pressure at infinity p_∞ is taken as $p_\infty=p_0-p_a\sin(2\pi f_a t)$, with p_a the acoustic pressure amplitude of the applied sound field and f_a its frequency. The sound veloci-

ty in the liquid at the bubble wall is given by

$$C = [C_0^2 + (n-1)H]^{1/2}, \text{ with } C_0 = 1482 \text{ m/s.}$$

The above model has been investigated for its sound radiation and resonance behavior[8] in the spirit of earlier work[9] and the first results connecting it with the experiments[2] have been given.[10] To follow as closely as possible the experimental procedure the bubble, described by Eqs. (1) to (3), is subject to an increasing or decreasing sound pressure amplitude p_a. The numerically obtained radius-time curve is transformed into a series of short-time spectra as were the pressure-time data of the experiment[2] and plotted in the same way as the "theoretical visible noise." Only the bubble radius is considered here and not the sound radiated, but this does not affect the subharmonic bifurcation structure as the radiated sound is directly coupled to the radial motion. We suggest calling this kind of plot a *spectral bifurcation diagram*. It may be used with any chaotic system and not just noise[2] or bubble radii.

Figure 1 gives an example of a calculated spectral bifurcation diagram for the following parameters: bubble radius at rest $R_n = 100$ μm; driving frequency $f_a = 23.56$ kHz (a value used in experiments); and sound pressure amplitude p_a raised from 0 to 14.8 bars in 40 ms. Each single spectrum has been calculated from 2K data points comprising 0.68 ms. By shifting the window of 2K data by 1K in the data base of calculated radii a new spectrum is obtained and plotted every 0.34 ms. The resolution in the frequency domain is $\frac{1}{16}$ of the driving frequency. The individual spectra are normalized to the strong-

est line in the spectrum, in this case always the line at the normalized driving frequency 1. The gray scale unfortunately does not give a good representation in print of the relative intensities of the spectral lines (compare Ref. 2). Thus we are now working with color graphics. These, however, are not reproducible here. A quite complicated behavior is observed with a period-doubling bifurcation sequence to chaos, a window in the chaotic region, and then, starting at about 20 ms (7.4 bars), a second period-doubling bifurcation sequence. The apparently chaotic bands near 1 and 7 ms are due to transients and are not of importance here. The second sequence we suggest calling a *subharmonic* period-doubling bifurcation sequence because it is not a window in the chaotic region connected with the first sequence but has its origin in a new attractor of *basic* period $2T_a = 2/f_a$ (called a *subharmonic resonance* in nonlinear oscillator theory[9]) which takes over stability from the chaos of basic period T_a and then undergoes itself period doubling to chaos of basic period $2T_a$. Our nomenclature infers that the seemingly similar spectra at about 7.5 and 25 ms are of different origin. This is indeed the case, as can be shown by recourse to the resonant properties of the system, i.e., by tracing back the oscillation in parameter space (f_a, p_a) to the corresponding resonance peak at lower driving pressure amplitudes p_a. It then becomes evident that the second sequence starts from a different resonance than the first one. A full projection of the traced-back bifurcation diagram in (R, f_a, p_a) space cannot be given because of computer-time limitations. As the resonance to which the oscillation at about 25 ms can be traced back is a subharmonic resonance of order $\frac{1}{2}$ (see Ref. 9 for the definition) this oscillation happens to be of doubled period. Depending on the order of the resonance, period tripling, quadrupling, and so on may and does occur (compare Fig. 2 below).

Many spectral bifurcation diagrams have been calculated which lead to the following picture of the bifurcation properties of our model. In the (f_a, p_a) parameter space many period-doubling Feigenbaum direct cascades and Grossmann inverse cascades are encountered. They appear as distinct entities and are therefore called *Feigenbaum-Grossmann objects*. The question of what governs the distribution of these objects in parameter space can be answered by the observation that they adhere to the resonances of the system. The resonances thus impose an ordering on the appearance of Feigenbaum-Grossmann objects, i.e., a distinct bifurcation superstructure.
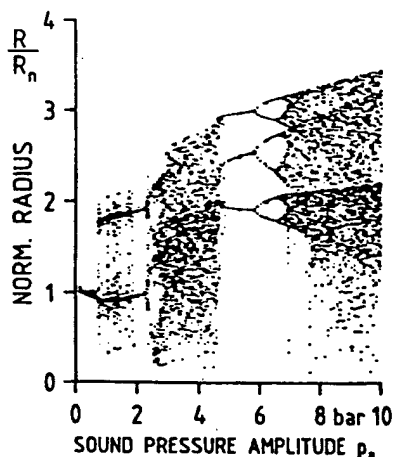
SOUND PRESSURE AMPLITUDE (bar)

FIG. 1. Spectral bifurcation diagram of bubble wall motion for a bubble with radius at rest of $R_n = 100$ μm driven at a frequency of $f_a = 23.56$ kHz. The sound pressure amplitude p_a is raised from 0 to 14.8 bars in 40 ms (photograph from a color graphics terminal).

FIG. 2. Bifurcation diagram of a bubble with a radius at rest of $R_n = 230$ μm. Frequency of the driving sound field is $f_a = 23.56$ kHz. The sound pressure amplitude is increased in steps of 0.1 bar. The scattered points that do not follow the obvious pattern are due to transients that have not yet died out (a result of computer-time limitations).

The scenario put forward above is not considered to be a specialty of the above model but to be a general (in the chaos language, universal) feature of most driven nonlinear oscillatory systems. This conjecture gets more support from experiments and calculations with a driven semiconductor oscillator[11] or Toda oscillator.[12] The so-called self-replicating attractor of Brorson, Dewey, and Linsay[11] is conjectured to belong to resonances of basic periods 1, 2, 3, 4, and 5, each with its own separate type of periodic chaos. The driven Toda oscillator[12]

$$\ddot{x} + r\dot{x} + (e^x - 1) = a \cos\omega t \qquad (4)$$

is an equation that we believe will become of fundamental importance in the area of nonlinear oscillations as a result of its simplicity. Recently Risken[13] encountered this equation in the context of quantum optics when simplifying a set of laser equations. Equation (4) as well as the experiments[12] show different basic periods and their period doubling to separate types of chaos. One result then is that the bifurcation structure of a driven dissipative nonlinear oscillatory system is composed of many Feigenbaum-Grossmann objects, distributed in parameter space as a kind of elementary fine structure according to the resonance properties of the system in question.

To test our ideas, also ordinary bifurcation diagrams have been calculated via the Poincaré return map. In these diagrams the radii of the bubble in the Poincaré plane of section are plotted versus the sound pressure amplitude. Figure 2 gives a bifurcation diagram for a bubble with a radius at rest of $R_n = 230$ μm in a sound field of frequency $f_a = 23.56$ kHz. First a period-1 oscillation (in units of $T_a = 1/f_a$) is stable; then a jump occurs to a period-2 oscillation. The first subharmonic resonance has taken over before the period-1 oscillation has found its way to chaos and even to the first period-doubling bifurcation. The subharmonic oscillation of basic period 2 then bifurcates by successive period doubling to chaos. Near 5 bars the second subharmonic resonance of basic period 3 takes over, which again bifurcates to chaos. This diagram strongly resembles those of Brorson, Dewey, and Linsay[11] and Klinker, Meyer-Ilse, and Lauterborn,[12] and fits into the resonance scenario developed above.

Next, what we suggest calling attractor maps have been calculated as a check of the chaotic properties of the oscillation. As an attractor map we define the map which is given by the approximate curve that may be obtained when one coordinate (say x) of a point on an attractor in a Poincaré plane of section is plotted versus the preceding one for a sufficient number of (usually consecutive) pairs, i.e., $x_{k+1} = \pi_1 P(x_k, y_k)$, where P denotes the Poincaré return map and π_1 the projection on the first entry of P. The first known example of an attractor map is the well-known Lorenz map given by a hatlike function.[14]

It is found that attractor maps may consist of several separate pieces. The number is connected with both the basic period of the attractor from which chaos originated and the inverse cascade of split chaotic regions of Grossmann.[4] In the fully developed chaotic region which occurs after period doubling and inverse cascade the number of separate pieces of the attractor map is the number of the basic period from which chaos originated. Thus in the case of the $R_n = 230$-μm bubble and a sound pressure amplitude of 7.4 bars (compare Fig. 2) we get a three-piece attractor map, whereas an $R_n = 100$-μm bubble at 11 bars yields a two-piece attractor map. By the definition of a subharmonic attractor map of order m through $x_{k+m} = \pi_1 P^m(x_k, y_k)$, where P^m is the mth iterate of the Poincaré map, and the plotting of x_{k+m} vs x_k, a one-piece map is obtained with a quadraticlike maximum. An example is given in Fig. 3, where r_{k+2} is plotted versus r_k for a 100-μm bubble at 11 bars sound pressure amplitude. Here r is the radius normalized to the maximum radius encountered in the sequence. We have several examples that show that when the inverse cascade is followed in the direc-
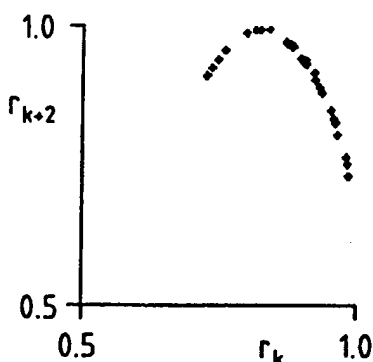
FIG. 3. Subharmonic attractor map of order 2 for the normalized radius $r = R/R_{max}$, $R_{max} = 546 \ \mu$m. $R_n = 100 \ \mu$m, $f_a = 23.56$ kHz, $p_a = 11$ bars, Poincaré plane of section at $\omega t = 155°$.

tion of the splitting of chaotic regions the attractor map also splits, which may be reconciled by using the $(m \times 2^n)$th iterate of the Poincaré map where n is the number of splittings that already occurred in the inverse cascade.

In this paper we have investigated a theoretical model of acoustic turbulence and found a connection between nonlinear resonances and the structure of bifurcation diagrams. This connection is conjectured to be a general (or universal) feature of driven dissipative passive nonlinear oscillatory systems. The nonlinear resonances impose a superstructure on the bifurcation behavior of the system. Feigenbaum direct and Grossman inverse cascades occur as fine structure which may be (partially) suppressed by the gross hysteresis features from the superstructure.

[1]Recent reviews are given by R. E. Apfel, in *Ultrasonics*, edited by P. D. Edmonds (Academic, New York, 1981), and by E. A. Neppiras, Phys. Rep. **61**, 159 (1980).

[2]W. Lauterborn and E. Cramer, Phys. Rev. Lett. **47**, 1445 (1981).

[3]M. J. Feigenbaum, J. Stat. Phys. **19**, 25 (1978).

[4]S. Grossman and S. Thomae, Z. Naturforsch. **32a**, 1353 (1977).

[5]See, e.g., *Dynamical Systems and Chaos*, edited by L. Garrido (Springer, Berlin, 1983).

[6]For examples of acoustic cavitation bubble fields, see *Cavitation and Inhomogeneities in Underwater Acoustics*, edited by W. Lauterborn (Springer, New York, 1980).

[7]F. R. Gilmore, California Institute of Technology Report No. 26-4, 1952 (unpublished). For a more accessible derivation, see *High Intensity Ultrasonic Fields*, edited by L. D. Rozenberg (Plenum, New York, 1971).

[8]E. Cramer and W. Lauterborn, Acustica **49**, 226 (1981).

[9]W. Lauterborn, J. Acoust. Soc. Am. **59**, 283 (1976).

[10]W. Lauterborn, Appl. Sci. Res. **38**, 165 (1982).

[11]S. D. Brorson, D. Dewey, and P. S. Linsay, Phys. Rev. A **28**, 1201 (1983).

[12]T. Klinker, W. Meyer-Ilse, and W. Lauterborn, Phys. Lett. **101A**, 371 (1984).

[13]H. Risken, private communication.

[14]E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).

FUNDAMENTALS AND APPLICATIONS
OF NONLINEAR ACOUSTICS

Mark F. Hamilton
Department of Mechanical Engineering
The University of Texas at Austin
Austin, Texas 78712-1063

The seeds of nonlinear acoustics were planted in the 18th and 19th centuries by the mathematicians and physicists who laid the theoretical foundations for fluid mechanics and wave motion. The principal contributions during this era are summarized in a benchmark article published by Lord Rayleigh at the beginning of the 20th century ["Aerial plane waves of finite amplitude," *Proc. Roy. Soc.* **A 84**, 247-284 (1910)]. However, not until the middle of the 20th century were reasonable approximations of these theories developed that provided access to relatively simple yet accurate quantitative descriptions of nonlinear effects in sound fields.

In recent years, sessions on nonlinear acoustics at meetings of the Acoustical Society of America have focused on topics that include sonic booms generated by supersonic aircraft, extracorporeal lithotripsy (the disintegration of kidney stones with focused shock waves) and a host of other nonlinear effects in biomedical acoustics, nonlinearity in thermoacoustic engines and acoustic pumps, 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 and other elastic solids. Each of these processes is associated with nonlinear terms that are discarded in the small-signal approximation (often called the acoustic approximation) of the equations of motion and state.

The purpose of this lecture is to provide an introduction to the basic theoretical models used in nonlinear acoustics, and to illustrate the principal nonlinear effects with results from a variety of experiments. Topics of discussion will include overviews of: 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; 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 pointed out.

A 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). The second review article (D. T. Blackstock, "Nonlinear Acoustics—Theoretical," AIP, 1972) presents greater theoretical detail and 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. A recent textbook on nonlinear acoustics that may be consulted for additional topics and references is by M. F. Hamilton and D. T. Blackstock, eds., *Nonlinear Acoustics* (Academic Press, 1998).

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.
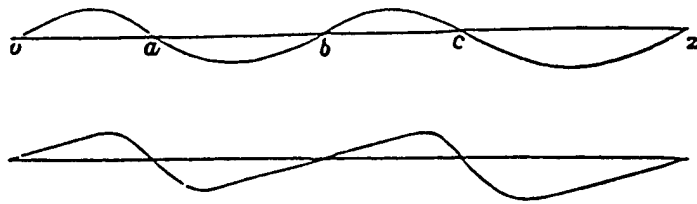
Fig. 1—Sketch of waveform distortion.

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]

$$\bar{x} = \frac{1}{\beta \epsilon k}, \tag{7}$$

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

$$\left(\frac{dx}{dt}\right)_{shock} = c_0 + \frac{\beta}{2}(u_a + u_b),$$ (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

$$u = u_0 \sum_{n=1}^{\infty} B_n \sin n\omega\tau.$$ (9)

Prior to shock formation the Fourier coefficients are given by

$$B_n = \frac{2J_n(n\sigma)}{n\sigma}, \qquad \sigma < 1,$$ (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:

$$B_n = \frac{2}{n(1+\sigma)}, \qquad \sigma > 3.$$ (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
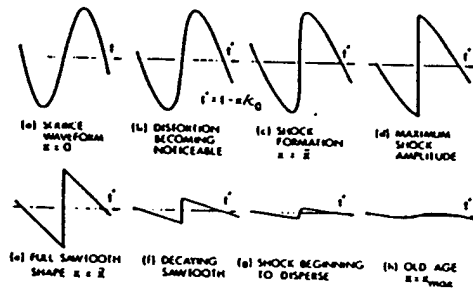
Fig. 2—Case history of a high intensity sound wave.
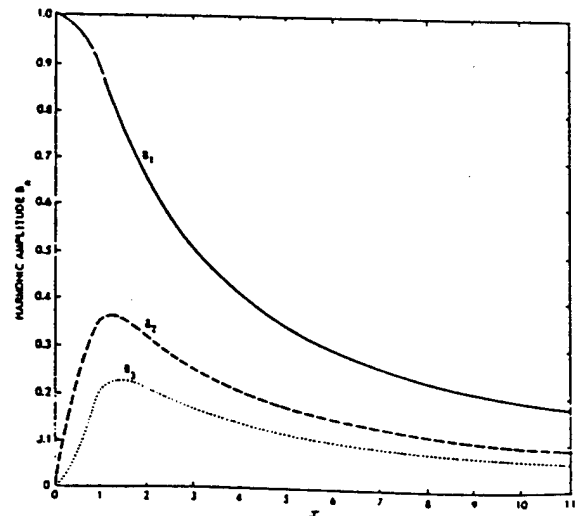
Fig. 3—Harmonic amplitudes of an originally sinusoidal wave.

in Fig. 2 [49], with the first three harmonic amplitudes graphed in Fig. 3 [48]. In Fig. 2a is shown one cycle of a sinusoidal source waveform. The subsequent distortion process leading up to shock formation (Fig. 2c) 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. 2e), 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. 2f). 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. 2g). Thermoviscous losses effectively low-pass filter the wave, which ultimately yields a wave dominated by the fundamental source frequency component (Fig. 2h). 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 $\bar{x}$ and either the source amplitude or frequency is increased. The shock formation distance $\bar{x} = (\beta \epsilon k)^{-1}$ 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
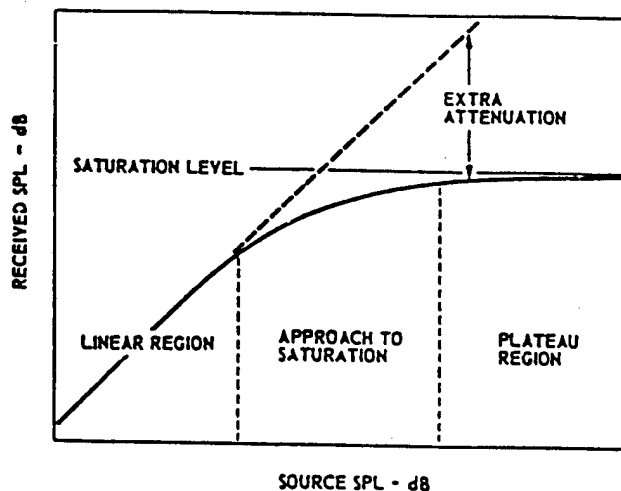
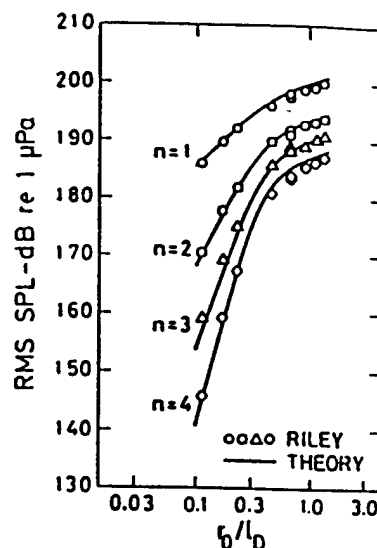Fig. 4—Amplitude response curve showing the development of saturation.

Fig. 5—Saturation of harmonic components in a directive sound beam.

$$u = \frac{2c_0}{\beta kx} \sum_{n=1}^{\infty} \frac{1}{n} \sin n\omega\tau, \qquad \sigma \gg 1. \tag{13}$$

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 $r_0/l_D = \beta\epsilon kr_0$. 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
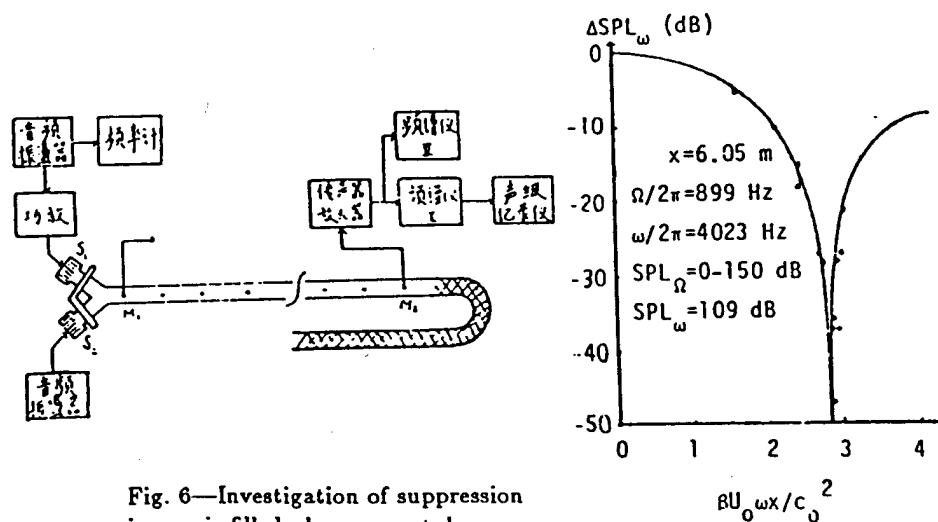
Fig. 6—Investigation of suppression in an air filled plane wave tube.

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 $b = \varsigma + (4/3)\eta + \kappa(1/c_v - 1/c_p)$, where ς 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

$$u = \frac{4c_0\alpha}{\beta k}\sum_{n=1}^{\infty}e^{-n\alpha x}\sin n\omega\tau, \qquad \sigma \gg \Gamma \gg 1. \tag{18}$$

92

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
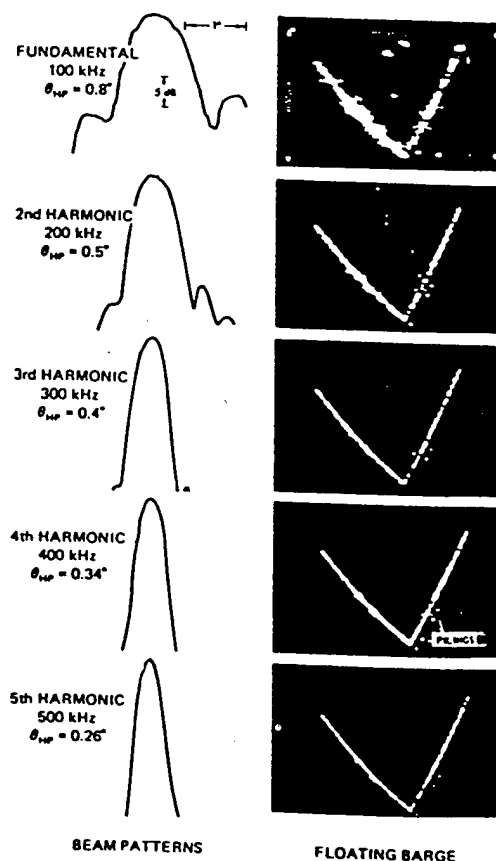
Fig. 7—Use of nonlinearly generated harmonics for acoustic imaging.

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

$$u(r,\tau) = \sum_{n=-\infty}^{\infty} u_n(r)e^{jn\omega\tau}, \quad \text{where} \quad u_n(r) = \frac{\omega}{2\pi}\int_{-\pi/\omega}^{\pi/\omega} u(r,\tau)e^{-jn\omega\tau}d\tau. \tag{26}$$

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.

95

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]

$$p_- = -\frac{j\beta k_-^2 a^2 p_{01} p_{02}}{4\alpha_T \rho_0 c_0^2} \frac{e^{-jx-r}}{r} D_W(\theta) D_A(\theta), \qquad r \gg \alpha_T^{-1}, k_- a^2/2, \qquad (28)$$

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 $\text{Im}(p_- e^{j\omega-t})$, 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

96

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
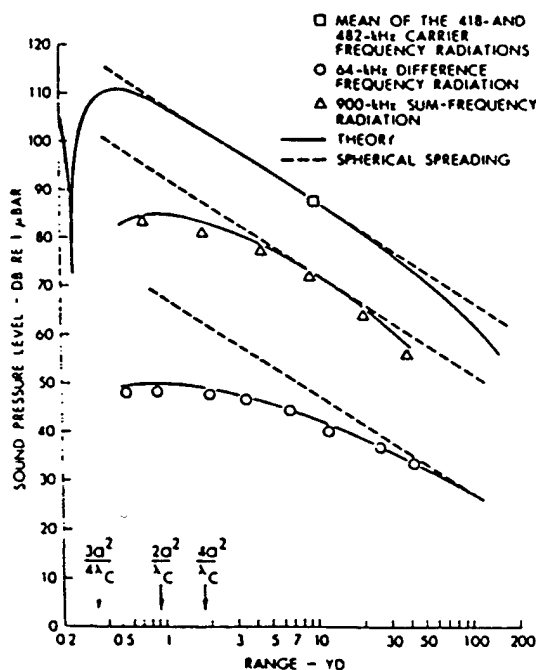
Fig. 8—Propagation curves for a parametric array.
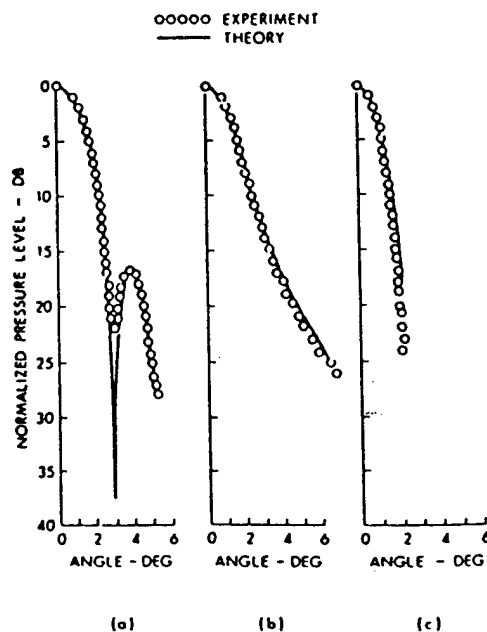
Fig. 9—Beam patterns for the (a) 482 kHz primary, (b) 64 kHz difference frequency, and (c) 900 kHz sum frequency waves.

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.

x = 4 5 cm 50 m V / div.	x = 19.0 cm. 0.9 m V / div
x = 7 5 cm. 10 m V / div.	x = 24.0 cm 0 4 m V / div
x = 10.5 cm 1 7 m V / div	x = 33.0 cm 0 4 m V / div
x = 14 5 cm 0 9 m V / div	x = 42.0 cm. 0.4 m V / div
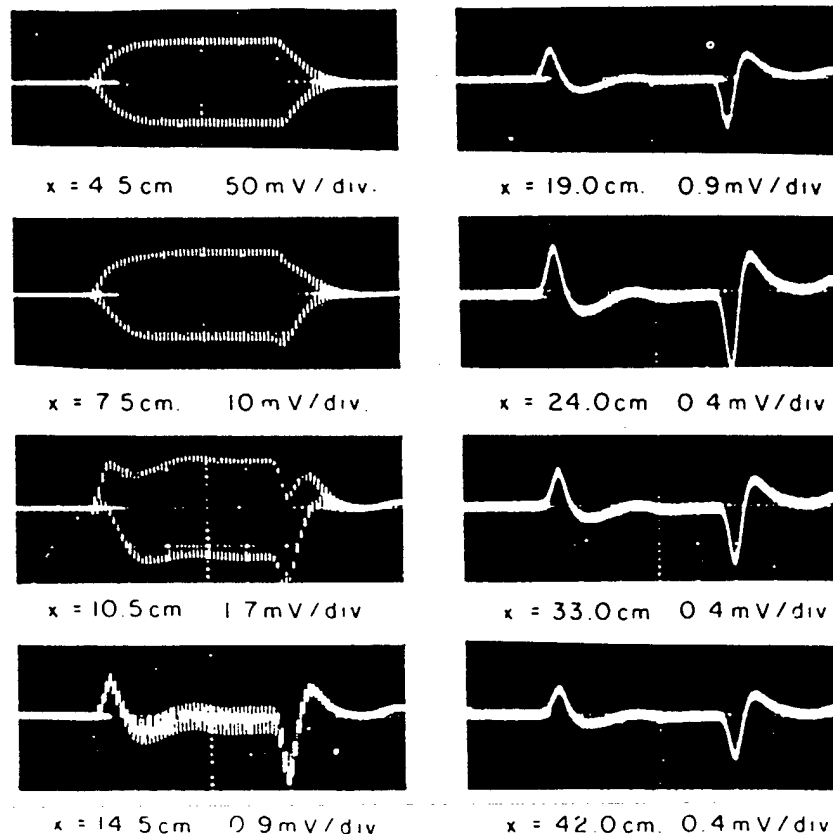
Fig. 10—Self-demodulation of a pulsed 10 MHz
carrier wave in carbon tetrachloride.

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
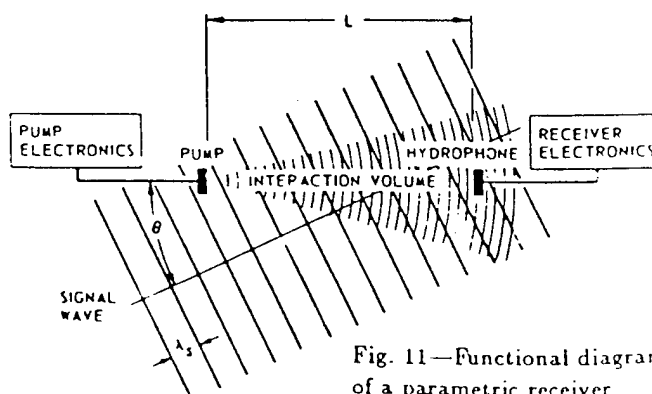
Fig. 11—Functional diagram
of a parametric receiver

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

$$p(z = 0) = p_0 \sin \omega t, \qquad r \leq a, \tag{33}$$
$$= 0, \qquad r > a,$$

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 $\mathrm{Im}(p_1 e^{j\omega r})$. 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 $p = \mathrm{Im}(p_1 e^{j\omega r})$ 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. 12a 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 12a 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. 12c, 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. 12c than in Fig. 12a. 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. 12a, the same lobe is down by only 11 dB in Fig. 12c (see Ref. 49 for related experimental work). However, the most striking feature of Fig. 12c is the
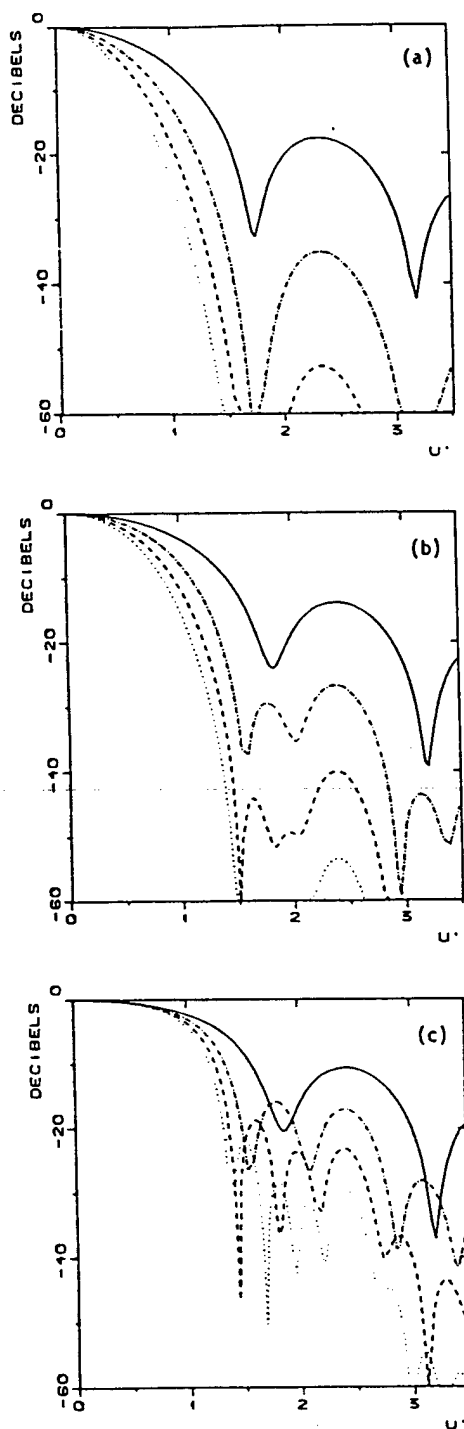
Fig. 12—Beam patterns calculated at $z = 10z_0$ for $n = 1$ (—), $n = 2$ (– –), $n = 3$ (— —), and $n = 4$ ($\cdots$), where $u' = 0.45ka\tan\theta$, and (a) $\alpha z_0 = 1.0$, (b) $\alpha z_0 = 0.1$, (c) $\alpha z_0 = 0.01$.

appearance of nearfield effects. In the second harmonic beam pattern there are twice as many sidelobes as in Fig. 12a, 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 12b 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.

103
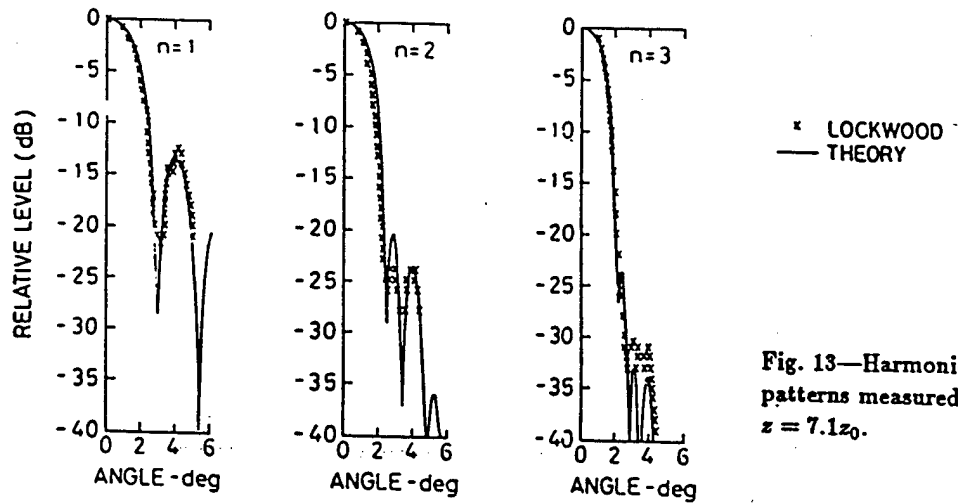
Fig. 13—Harmonic beam patterns measured at $z = 7.1z_0$.

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 $\text{Im}(p_1 e^{j\omega r})$. 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]:

$$p_2 = \frac{j\beta p_g^2 k^2 r_0^2}{4\rho_0 c_0^2(1 - jz/z_0)} \exp\left(-4\alpha z - \frac{2(r/r_0)^2}{1 - jz/z_0} - j2\alpha\right)\left(E_1[-j2\alpha] - E_1[-j2\alpha(1 - jz/z_0)]\right), \tag{39}$$

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 $\text{Im}(p_2 e^{j2\omega r})$. 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

104

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 $\bar{r}_0 = r_0(1 - jz_0/d)^{-1/2}$. Substitution of $\bar{r}_0$ 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

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," *Sov. 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).

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 loss-less 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:

$$\rho T \frac{DS}{Dt} = C_v \left[\rho \frac{D\Im}{Dt} - \frac{\gamma - 1}{\beta_e} \frac{D\rho}{Dt} \right] = \psi^{(\eta)} - \frac{\partial Q_i}{\partial x_i} \qquad (3n\text{-}3)$$

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, $\Im$ is absolute temperature, S is entropy per unit mass, $\gamma = C_p/C_v$ is the ratio of specific heats, $\beta_e = -\rho^{-1}(\partial \rho/\partial \Im)_p$ is the coefficient of thermal expansion, $\psi^{(\eta)} = 2\eta d_{ij}d_{ji} + \eta' d_{ii}d_{ii}$ 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,

$$Q_i = -\kappa \frac{\partial \Im}{\partial x_i} \qquad (3n\text{-}4)$$

where κ is the coefficient of thermal conduction. For heat radiation the relation between q and $\Im$ 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

$$\frac{\partial Q_i}{\partial x_i} = \rho C_v q(\Im - \Im_0) \qquad (3n\text{-}5)$$

where $\Im_0$ 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

$$p = R\rho\Im \qquad (3n\text{-}6)$$

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 $c_0^2 = \gamma R\Im_0 = \gamma p_0/\rho_0$, where p_0 and ρ_0 are the ambient values of p and ρ. Let $\Im = \beta_{e0}(1 + \theta)$, $p = p_0 + \rho_0 c_0^2 P$, and $\rho = \rho_0(1 + s)$, where β_{e0} is the ambient value of β_e (for perfect gases $\beta_{e0}\Im_0 = 1$). Assume that θ, P, and s are small quantities of first order. Expansion of Eq. (3n-6) to second order yields

$$\theta = \gamma P - s + s^2 - \gamma Ps \qquad (3n\text{-}7)$$

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

$$\theta = \gamma P - s - (\gamma - 1)s^2 \qquad (3n\text{-}8)$$

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 $\mathfrak{I} = \mathfrak{I}(p,\rho)$. Recognizing that $(\partial\mathfrak{I}/\partial p)_\rho = \gamma(\rho c^2 \beta_s)^{-1}$, 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

$$h = 1 + \frac{\gamma B}{2A} + \tfrac{1}{2}(\gamma - 1)\left(1 - \frac{B}{2A}\right) - (\gamma - 1)^2(4\beta_{s0}\mathfrak{I})^{-1} \tag{3n-16}$$

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 $\mathfrak{I}_0{}^{-1}$ 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 r and δ, called "Riemann invariants," may be defined by

$$2r = \lambda + u \qquad 2\delta = \lambda - u \tag{3n-24}$$

If Eqs. (3n-20) are first added and then subtracted, the results are respectively

$$r_t + (u + c)r_x = 0 \tag{3n-25a}$$
$$\delta_t + (u - c)\delta_x = 0 \tag{3n-25b}$$

115

Thus, as first found by Riemann (ref. 6),

$$\left(\frac{dx}{dt}\right)_{t-\text{const}} = u + c \qquad (3n\text{-}26a)$$

$$\left(\frac{dx}{dt}\right)_{t-\text{const}} = u - c \qquad (3n\text{-}26b)$$

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

$$c = c_0 \pm (\beta - 1)u \qquad (3n\text{-}27)$$

where $\beta = \frac{1}{2}(\gamma + 1)$. Combination of this equation with Eq. (3n-11) leads to

$$p - p_0 = p_0 \left\{ \left[1 \pm (\beta - 1)\frac{u}{c_0} \right]^{2\gamma/(\gamma-1)} - 1 \right\} \qquad (3n\text{-}28)$$

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,

$$p - p_0 = \pm \rho_0 c_0 u \qquad (3n\text{-}29)$$

The nonlinear differential equation for simple waves, Eq. (3n-22), becomes

$$u_t + (\beta u \pm c_0)u_x = 0 \qquad (3n\text{-}30)$$

If we restrict ourselves momentarily to outgoing waves, the propagation speed is

$$\left(\frac{dx}{dt}\right)_{u-\text{const}} = \beta u + c_0 \qquad (3n\text{-}31a)$$

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,

(a) X = 0 (b) X > 0 (c) X = $\overline{X}$ (d) X > $\overline{X}$
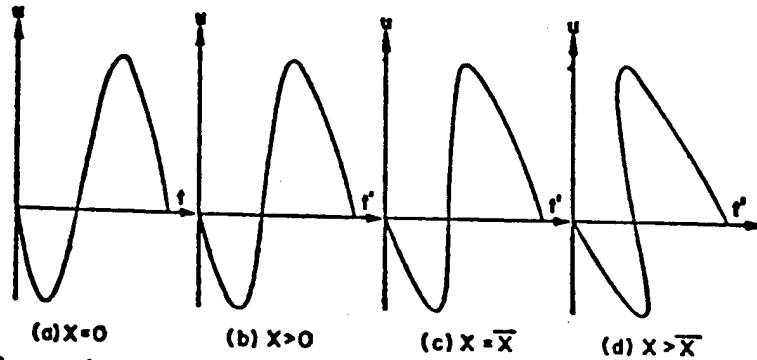
FIG. 3n-1. Progressive distortion of a finite-amplitude wave. Symbols are: u = particle velocity, x = spatial coordinate, t = time, $t' = t - x/c_0$ (delay time), $\bar{x}$ = point at which a shock begins to form.

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-1d. The interpretation of Fig. 3n-1d 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.

116

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

$$\phi = t - \frac{x - X(\phi)}{\beta X'(\phi) \pm c_0} \tag{3n-36b}$$

117

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-1d. 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-1c, 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-37b), 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-37b)] and (3n-37a). 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

$$\sigma = \frac{\beta \epsilon x}{x_\epsilon} \qquad \epsilon = \frac{u_0}{c_0} \tag{3n-43}$$

Here x_ϵ 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-42a).

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:

$$\left(\frac{dx}{dt}\right)_{\text{--const}} = \frac{\pm c_0}{1 \mp \beta u/c_0} \tag{3n-45}$$

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

$$u\Big|_{x=0} = g(t)H(t) = g(t')$$ (3n-48)

where it is assumed that $g(t) = 0$ for $t < 0$, the solution is

$$u = g(\phi)$$ (3n-49a)
$$\phi = t' + \beta c_0^{-2} x g(\phi)$$ (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):

$$\left(\frac{dx}{dt}\right)_{u=\text{const}} = u \pm c_0(1 + c_1 U + c_2 U^2 + \cdots)$$ (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

$$\phi = t - \frac{x - X(\phi)}{u \pm c_0(1 + c_1 U + c_2 U^2 \cdots)}$$ (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

$$\xi_{tt} = c_0^2 G(\xi_a) \xi_{aa}$$ (3n-52)

where

$$G(\xi_a) = 1 + \left(\frac{M_3}{M_2}\right) \xi_a + \left(\frac{M_4}{M_2}\right) \xi_{aa} \cdots$$ (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

$$\lambda = -c_0 \int_0^{\xi_a} [G(\xi_a')]^{\frac{1}{2}} d\xi_a'$$ (3n-54)
$$= -c_0[\xi_a - \tfrac{1}{4}m_3\xi_a^2 + (\tfrac{1}{8} - \tfrac{1}{8}m_4)m_3^2\xi_a^3 \cdots]$$ (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

$$\xi_a = \mp U + \tfrac{1}{4}m_3 U^2 \mp \tfrac{1}{8}m_4 m_3^2 U^3 \cdots$$ (3n-56)

The propagation speed for simple waves is

$$\left(\frac{da}{dt}\right)_{u=\text{const}} = \pm c_0 G^{\frac{1}{2}}$$ (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

$$\left(\frac{da}{dt}\right)_{u=\text{const}} = \pm c_0[1 \pm \tfrac{1}{2}m_3 U + \tfrac{1}{4}m_3^2(1 - 2m_4)U^2 \cdots]$$ (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

$$\beta = \frac{1}{2}(\gamma + 1) \qquad \text{for gases} \tag{3n-60a}$$

$$\beta = 1 + \frac{B}{2A} \qquad \text{for liquids} \tag{3n-60b}$$

$$\beta = \frac{-M_3}{2M_2} \qquad \text{for solids} \tag{3n-60c}$$

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-60a) or (3n-60b). 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

$$w(0,t') = g(t') \tag{3n-63}$$

121

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-65b) 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-67b).

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-2a 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-2b and c). After the shock is born (Fig. 3n-2c), 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-2c). In Fig. 3n-2f 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-2g) 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.
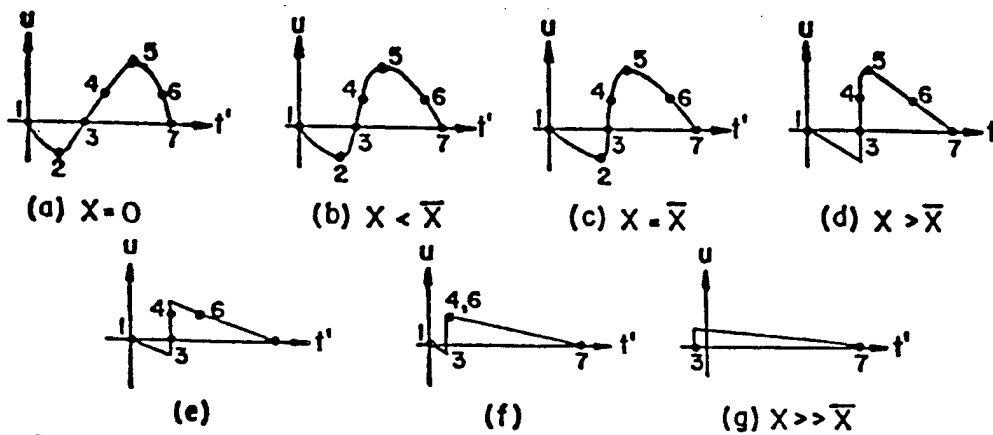
FIG. 3n-2. Development and decay of a finite-amplitude wave. Numbered points refer to initial phases (values of ϕ) of the wave. (*From ref. 27.*)

Suppose now that a shock begins to form at time $\bar{l}$ and distance $\bar{x}$. It will arrive at a subsequent point x at time t_s given by

$$t_s = \bar{l} + \int_{\bar{x}}^{x} v^{-1} d\mu \tag{3n-71}$$

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

$$t_s = \bar{l}' - \tfrac{1}{2}\beta c_0^{-2} \int_{\bar{x}}^{x} (u_a + u_b) d\mu \tag{3n-74}$$

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

$$t_s = \bar{l}' - \tfrac{1}{2}\beta c_0^{-2} \int_{\bar{x}}^{x} (w_a + w_b) d\mu \tag{3n-75}$$

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 ($u = u_a$, $\phi = \phi_a$, $t' = t'_a$) and just behind it ($u = u_b$, $\phi = \phi_b$, $t' = t'_a$). 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

$$t'_a = g^{-1}(w_a) - \beta c_0^{-2} z w_a \tag{3n-78a}$$

and just behind

$$t'_a = g^{-1}(w_b) - \beta c_0^{-2} z w_b \tag{3n-78b}$$

Equations (3n-78a), (3n-78b), and (3n-75) or (3n-76) are to be solved simultaneously for w_a, w_b, and t'_a.
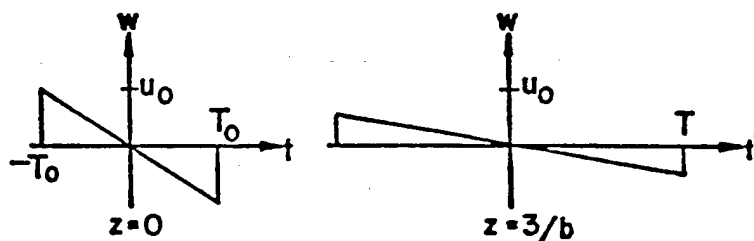
Fig. 3n-3. N wave.

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,

$$\frac{dt'_s}{dz} = -\tfrac{1}{2}\beta c_0^{-2}w_b = \frac{\tfrac{1}{2}bt'_s}{1 + bz}$$

Integration gives

$$-t'_a = T = T_0(1 + bz)^{\tfrac{1}{2}}$$

The amplitude of the wave is therefore given by

$$u_b = \left(\frac{A_0}{A}\right)^{\tfrac{1}{2}} \frac{u_0}{(1 + bz)^{\tfrac{1}{2}}}$$

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:

$$w = \mp \frac{u_0}{1 + b|z|} \frac{t'}{T_0} \qquad -T < t' < T \qquad (3n\text{-}79)$$

$$T = T_0(1 + b|z|)^{\frac{1}{2}} \qquad (3n\text{-}80)$$

$$|u_b| = \left(\frac{A_0}{A}\right)^{\frac{1}{2}} \frac{u_0}{(1 + b|z|)^{\frac{1}{2}}} \qquad (3n\text{-}81)$$

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^{-1}, and $r^{-1}(\ln r)^{-\frac{1}{2}}$, respectively.

Sawtooth Wave. Assume that the wave shown in Fig. 3n-3a 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

$$|w_b| = \frac{\pi u_0}{\pi + \beta \epsilon k |z|} \qquad (3n\text{-}82)$$

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

$$u_b = \frac{\pi u_0}{1 + \sigma} \qquad (3n\text{-}83)$$

where (see Sec. 3n-4) $\sigma = \beta \epsilon k x = x/\bar{x}$. 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 $\beta \epsilon k z$.

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

$$\frac{\tau}{T} = \frac{12\delta}{c_0 |u_b| T} = \frac{12\delta}{c_0 |w_b| T} \left(\frac{A}{A_0}\right)^{\frac{1}{2}} \qquad (3n\text{-}84)$$

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 ($= u_0 T_0$) so that τ/T is simply

126

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,

$$\frac{r_{max}}{r_0} \approx \frac{\beta u_0 c_0 T_0}{12\delta} \qquad (3\text{n-}85a)$$

The comparable result for cylindrical N waves is

$$\frac{r_{max}}{r_0} \approx \left(\frac{\beta u_0 c_0 T_0}{12\delta}\right)^2 \qquad (3\text{n-}85b)$$

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,

$$u_t + \beta u u_{x'} = \delta u_{x'x'} \qquad (3\text{n-}86a)$$

Here $x' = x - c_0 t$, $\delta = \frac{1}{2}\nu[\mathcal{U} + (\gamma - 1)/\mathrm{Pr}]$, $\nu = \eta/\rho_0$ is the kinematic viscosity, $\mathcal{U} = (\eta' + 2\eta)/\eta$ 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-60b). In certain cases it applies also to solids (ref. 33).

Equation (3n-86a) 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)

$$c_0^2 u_x - \beta u u_{t'} = \pm \delta c_0^{-1} u_{t't'} \qquad (3\text{n-}86b)$$

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

$$\pm c_0^2 \zeta_x - \delta \zeta_{t't'} = 0 \tag{3n-88}$$

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 $K = c_0^2/4\delta x$. 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 $\zeta_0 = \exp\left[\frac{1}{2}\Gamma(1 - \cos \omega t')\right]$ 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 ($\alpha \lambda = \omega \delta/c_0^2$). 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 constant 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

$$u_{\mathfrak{s}} = \frac{\pi u_0}{1 + \sigma}$$

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

$$u \cong 4\alpha\lambda c_0 \beta^{-1} e^{-\alpha z} \sin \omega t' \qquad (3n\text{-}95)$$

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)

$$u_{\mathfrak{s}} \cong \frac{\pi c_0^2}{\beta \omega x} \qquad (3n\text{-}96)$$

but this result is accurate only in the sawtooth region, which is defined roughly by $3\bar{x} < x < \alpha^{-1}$ (ref. 35).

Nonplanar Waves. For other one-dimensional waves the analog of Eq. (3n-86b) is

$$c_0^2 (u_z + uA_z/2A) - \beta c_0 uu_{t'} = \delta u_{t't'} \qquad (3n\text{-}97)$$

(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

$$c_0^2 w_s - \beta c_0 ww_{t'} = \delta \left(\frac{A}{A_0}\right)^{\frac{1}{2}} w_{t't'} \qquad (3n\text{-}98)$$

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

$$m \equiv \frac{b_\infty^2 - b_0^2}{b_0^2} \qquad (3n\text{-}99)$$

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:

$$\left(q + \frac{\partial}{\partial t'}\right) u_s - b_\infty^{-1}\left(\beta_i q + \beta_a \frac{\partial}{\partial t'}\right) u u'_t = \pm \frac{m}{2b_0} u_{t't'} \qquad (3n\text{-}100)$$

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 $(A/A_0)^{\frac{1}{2}} = r/r_0 = e^{t/r_0}$], and some initial attempts at solving this equation were described.

Boundary-layer Effects. Consider the propagation of a plane wave in a thermo-viscous 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:

$$u_x - \frac{\beta}{c_0^2} u u_{t'} = \mp \frac{1 + (\gamma - 1)/\sqrt{\text{Pr}}}{c_0 D/2} \left(\frac{\nu}{\pi}\right)^{\frac{1}{2}} \int_0^\infty u_{t'}(x, t' - \mu) \frac{d\mu}{\sqrt{\mu}} \quad (3n\text{-}101)$$

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

$$\mathfrak{r} = \frac{c}{\gamma - 1} + \frac{u}{2} \quad (3n\text{-}102a)$$

$$\mathfrak{s} = \frac{c}{\gamma - 1} + \frac{u}{2} \quad (3n\text{-}102b)$$

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:

$$t_{\mathfrak{r}\mathfrak{s}} + N(\mathfrak{r} + \mathfrak{s})^{-1}(t_{\mathfrak{r}} + t_{\mathfrak{s}}) = 0 \quad (3n\text{-}103)$$

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

$$\lambda_{\text{wall}} = 2\lambda_i$$

131

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

$$\left(\frac{p}{p_0}\right)_{wall} = \left[2\left(\frac{p_i}{p_0}\right)^{1/\mu} - 1\right]^{\mu} \tag{3n-105}$$

where $\mu = 2\gamma/(\gamma - 1)$. Now define a wall amplification factor α by

$$\alpha = \frac{p_{wall} - p_0}{p_i - p_0}$$

Substitution from Eq. (3n-105) gives

$$\alpha = \frac{[2(p_i/p_0)^{1/\mu} - 1]^{\mu} - 1}{p_i/p_0 - 1}. \tag{3n-106}$$

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 $(p_i - p_0 \ll p_0)$ $\alpha = 2$, in agreement with linear theory. The limiting value for very strong waves is $\alpha = 2^{\mu}$ $(= 2^7$ 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 α is entirely different; the limiting value for strong shocks is $\alpha = 2 + (\gamma + 1)/(\gamma - 1) = 8$ 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 ($2\tau = \lambda_{surface} + u_{surface} = u_{surface}$, since $\lambda = 0$ when $p = p_0, \rho = \rho_0$). The result is

$$u_{surface} = 2u_i$$

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-16. 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.

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. 360b.
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 Ingenicurw.* 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).

Wave Propagation in Periodic, Random, and Quasicrystalline Media

J. D. Maynard

The Pennsylvania State University, University Park, PA 16802

Abstract

One of the most interesting problems currently under study condensed matter physics involves the quantum wave nature of electrons in complex and novel situations. While the wave nature of electrons is certainly established, the behavior of electron waves in disordered samples, samples with inelastic scattering in boundary elements, disordered samples in magnetic fields, etc. has become understood only through recent discoveries. While interest in the behavior of waves propagating in complex scattering fields dates back to Lord Rayleigh, the discovery of Anderson localization, coherent backscatter, Furstenberg's rigorous theorem for one dimension, etc. are contemporary developments. A recent problem involves normal electron "persistent currents" involving electrons which maintain their phase coherence around a normal metal ring; theory is being taxed to explain some experimental results. In order to gain insight into the behavior of waves in complex systems, we have been using classical acoustic wave systems to model various quantum wave systems. Macroscopic acoustic systems can be fabricated which precisely duplicate some of the relevant features of the quantum system, and permit direct control of parameters and measurement of eigenvalues, eigenfunctions and other properties which are difficult if not impossible to measure in the electron systems. The effects which we have studied include the effects of inelastic scattering on Anderson localization (phonon assisted hopping), the effects of nonlinearity on Anderson localization (possibly a model for many-electron effects), the effects of a two-dimensional quasicrystalline (Penrose tile) symmetry, and the normal electron persistent current problem.

An early goal in our research was to study Anderson localization in two dimensions. However, before attempting this project it was felt that it would be advantageous to learn about Anderson localization initially in a one-dimensional experiment. In this experiment, the wave medium was a long (15 m) 0.178 mm diameter steel wire suspended vertically; a tension T_0 in the wire was maintained with a weight attached at the lower end. The wave field ψ consisted of transverse waves in the wire generated with an electromechanical actuator at one end of the wire. The periodic or nearly periodic potential field V for the wire was provided by small lead masses spaced along the wire with an average lattice constant $a = 15$ cm; a total of 50 masses was used. The masses were sufficiently small so that the potential V could be approximated as a series of delta functions with strength $m\omega^2/T_0$; the system was a precise analog of a Kronig-Penny potential field. A transducer measuring amplitude and phase of the transverse vibrations of the wire could be translated along the length of the wire; thus, complete eigenfunctions could be measured directly. For diagonal type disorder, the positions of the masses were varied. Several sets of measurements were made with the positions randomly varied within maximum displacements from lattice sites of 0.007a, 0.01a, 0.02a, and 0.05a. Results with static disorder configurations were in good

agreement with computer simulations.

Before studying the localization effects, we first verified Bloch wave behavior, as developed for electrons in solid state physics and realized in the band structure of crystals, by making measurements with the masses spaced periodically. The frequency response (band structure) of the system was measured by monitoring the transverse wave amplitude near one end of the series of masses while sweeping the frequency (analogous to electron Fermi energy) of the transverse actuator at the opposite end. The response shows distinct bands and gaps, with each band containing approximately 50 eigenfrequencies (corresponding to the eigenstates of the 50 site system), positioned as predicted by Bloch's theorem. By driving the static periodic system at one of the eigenfrequencies and translating the transducer along the wire, a Bloch wave eigenstate could be recorded. Some clear observations of Anderson localization were obtained with 2% disorder, i.e. with the masses displaced from the periodic lattice sites with a flat random distribution between $\pm 0.02a$. The eigenvalue spectrum showed dramatic departure from Bloch response. Some eigenstates appeared in what were the gaps of the periodic system.

While the one-dimensional system described above was built for pedagogical reasons, it became apparent that the system could be used for more serious research. That is, the system was used to study inelastic scattering and phonon assisted hopping. This was modeled in the "adiabatic approximation" with a time-dependent potential of the form $V[x(1 + \epsilon \cos \Omega t)]$. The time dependence is provided by a second electro-mechanical actuator which modulates the strain ϵ in the wire at a low frequency Ω.

In addition to studying the effects of inelastic scattering, we have also studied the effects of nonlinearity. It has been suggested that the nonlinear Schrodinger equation is a model for many electron effects; however, because of exchange effects, etc., this is probably of limited utility. At any rate, the nonlinear wave equation is of interest in its own right. While the study of waves in disordered media and the study of nonlinear systems, including effects such as soliton propagation, have each received considerable attention, the combined field, nonlinear wave propagation in disordered media, is relatively new. A fundamental question is: Does nonlinearity weaken Anderson localization? There are about eight theoretical papers (or sequences of papers) addressing this question in the literature, and roughly half of these papers predict that nonlinearity will weaken Anderson localization, and the others predict that it will not. While this statement seems to imply a controversy, the theories are in fact not contradictory, because as is common in nonlinear problems, the question does not have a unique answer. Some of the theory papers consider the existence of exponentially localized solutions of a Hamiltonian with a nonlinear term, and the result is that under general conditions Anderson localized eigenstates are still present when there is nonlinearity. An open question involved the possibility of resonant tunneling between localized states, and our experiment was particularly suited for addressing this question. A finite amplitude at one localization site changes the instantaneous arc length of the wire, which changes the tension in the wire. The change in tension may then parametrically excite a state which is localized at a different site, resulting in nonlinearly enhanced tunneling.

Our measurements of nonlinear effects in the mass-loaded wire system consisted of examining the spectral response (amplitude at a fixed site as a function of frequency), measured at a distance of about four localization lengths from the drive actuator and normalized by dividing by the drive amplitude, for different drive amplitudes. If the system were strictly linear, then the normalized response would not change. If the Anderson localization is weakened by the nonlinearity, then as the drive amplitude is increased, the normalized response at the distant site should increase. We found that the normalized response actually decreased by an average of 30%, indicating an enhanced localization. One might imagine that the nonlinearity causes an Anderson localized state not to parametrically excite a distant site, but rather to "dig a deeper hole" for itself.

A somewhat different behavior has been predicted for the propagation of a soliton through a disordered medium. We have studied this type of behavior using surface waves on a thin film of superfluid helium on a disordered substrate, and by driving the waves at finite amplitude, have found results consistent with the predictions.

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. Quasiperiodic (Penrose tile) patterns are a new symmetry, and it was felt that it would be important to understand how this new symmetry might effect the properties of materials. A fundamental question is: given a wave equation (for Schrodinger waves or sound waves) with a potential field having Penrose tile symmetry, what is the effect of the symmetry on the eigenvalues and eigenfunctions. Unlike the theorems for the quasiperiodic patterns in one dimension, "quasi-Bloch" theorems for two and three dimensions, if they exist, have not yet been discovered.

In the spirit of the experiments with the mass-loaded wire, we used an acoustic model to study the effects of Penrose tile symmetry on eigenvalues and eigenfunctions in two dimensions. A convenient model to use as a guide for designing classical analogs is the tight-binding model. In this model lattice sites are occupied by local oscillators which, if isolated, would have one or more sharp eigenfrequencies (such as the quantized energy levels of an isolated atom). These local oscillators are allowed to interact through some coupling mechanism to nearest- neighbor lattice sites; the local oscillator eigenfrequencies then broaden out into bands of eigenfrequencies. The symmetry of the coupling, i.e. the quasi-periodic Penrose pattern, should have some manifestation in the resulting eigenvalue spectrum and band structure.

For studying wave mechanical (i.e. interference) effects in a complex system, the system should have little or no damping, i.e. the local oscillators and the coupling mechanism should have a high quality factor (Q). For the local oscillators in the Penrose tile simulation ordinary commercial tuning forks (440 Hz) were used. These have the advantage that they can be mounted by the stem and still maintain a high Q oscillation. The tuning forks are epoxied into a heavy aluminum plate at the centers of rhombuses in a Penrose pattern. For the nearest neighbor coupling, arcs of ~1 mm diameter steel wire are spot-welded from one tine of a tuning fork to that of a nearest neighbor. Using the four sides of each rhombus,

137

four nearest neighbors are identified, and each tine of a tuning fork is coupled to the two nearest tines of the adjacent tuning forks.

In order to drive the oscillations of the coupled tuning fork system, an electromagnet is positioned near one tine of the array, and an AC current is passed through the electromagnet. The response of the system is monitored with four electrodynamic transducers (electric guitar pick-ups) positioned next to random tines in the array. By sweeping the frequency of the drive electromagnet, the resonant response of the system is detected with the pick-up transducers; the resonant frequencies, in bands near 440 Hz, correspond to the eigenvalues of the quasiperiodic system. The eigenvalue spectrum, determined as a composite of the resonant spectra (containing ~300 resonances) from twenty different positions in the Penrose pattern, shows gaps and bands whose widths are in the ratio of the Golden Mean, $\left(\sqrt{5}+1\right)/2$, a result not predicted by theory or computer simulations. By using small mirrors on the ends of each tine and a scanning laser system, the eigenfunctions could also be measured. Localized as well as extended eigenfunctions were found.

One of our most recent acoustic analog experiments probes an aspect of the discrepancy between theory and the normal electron persistent current experiments. These experiments, such as that of R. Webb, et al., involve a normal metal ring surrounded by a drive loop and squid pickup loop. The pickup loop monitors the current in the ring, through its mutual inductance, as the magnetic flux Φ in the system is changed with the drive loop. The current in the ring is found to oscillate with a period of $\Phi_0 = h/e$, the normal electron flux quantum, as predicted by theory. However, the magnitude of the current oscillations exceed the theoretically predicted value by more than an order of magnitude. Our acoustic analog measurements suggest a solution to this discrepancy.

References

1. Shanjin He and J. D. Maynard, "Detailed Measurements of Inelastic Scattering in Anderson Localization", Phys. Rev. Lett. **57**, 3171-3174 (1986).

2. Shanjin He and J. D. Maynard, "Eigenvalue spectrum, density of states, and eigenfunctions of a two-dimensional quasicrystal", Phys. Rev. Lett. **62**, 1888-1891, (1989).

3. M. J. McKenna, R. L. Stanley, and J. D. Maynard, "Effects of nonlinearity on Anderson localization", Phys. Rev. Lett. **69**, 1807-1810 (1992).

4. J. D. Maynard, "A possible explanation of the discrepancy in electron persistent current amplitudes: A superfluid persistent current analog", J. Low Temp. Phys. **89**, 155-159 (1992).

5. V. A. Hopkins, J. Keat, T. Zhang, and J. D. Maynard, "Observation of the predicted behavior of nonlinear pulse propagation in disordered media", Phys. Rev. Lett. **76**, 1102-1105 (1996).

POROUS MEDIA

James M. Sabatier
National Center for Physical Acoustics
University of Mississippi

Abstract

The physics of the interaction of airborne sound with the surface of the earth[1] will be presented. The ground is an example of a fluid filled porous material. When sound in the atmosphere is incident upon the ground surface vibrational energy is transferred to both the fluid and solid phases. This coupling has been referred to as acoustic-to-seismic coupling[2,3] and will be discussed in the light of the Biot model[4]. The relevant pore parameters, air-porosity, permeability and pore tortuosoity can be determined from acoustic measurements. Experimental techniques used to study porous soils include buried microphones, geophones and laser Doppler vibrometers. Specific measurements of the acoustic-to-seismic coupling ratio and the physics that can be extracted will be presented. The principle components of the Biot model, including Type I and II phase velocity and attenuation and boundary conditions at various interfaces, will be described and compared to measured data. Since air-filled soils approximate the light-fluid limit of Biot theory, rigid-framed porous media models can also be used to describe the fluid phase physics and these models will be described.

References

1. Sabatier, *et al.*, "The interaction of airborne sound with the porous ground: The theoretical formulation," J. Acoust. Soc. Am. **79**,1345-1352(1986).

2. Attenborough, *et al.*, "The acoustic transfer function at the surface of a layered poroelastic soil," J. Acoust. Soc. Am. **79**, 1353-1358 (1986).

3. Sabatier, *et al.*, "Acoustically induced seismic waves," J. Acoust. Soc. Am. **80**, 646-649 (1986).

4. M. A. Biot, "Theory of propagation of elastic waves in a fluid saturated porous solid. I. Low frequency range. II High frequency range," J. Acoust. Soc. Am. **28**, 168-191 (1956).

Tutorial on Quantum Mechanics and Applications in Acoustics

J. D. Maynard

The Pennsylvania State University, University Park, PA 16802

Abstract

The purpose of this talk will be to discuss the fundamentals of quantum mechanics, to work some sample problems relevant to understanding real applications in acoustics, and to review such applications. A tentative outline of the tutorial is as follows:

I. The historical development of quantum mechanics; there were many misinterpretations, but the final result was correct.

II. The correct formulation of quantum mechanics; the myths of wave-particle duality, paradoxes, and the uncertainity principle

III. Examples of problems in quantum mechanics relevant for acoustics:

    A. The hydrogen atom

    B. Angular momentum

    C. The interaction of atoms with electromagnetic fields; selection rules; perturbation theory and the Golden Rule

    D. The simple harmonic oscillator

IV. Applications of quantum mechanics in acoustics:

    A. Collisions of molecules (required for kinetic theory and dissipative processes in thermohydrodynamics and acoustics)

    B. Bloch's theorem and the band structure of solids

    C. Fermi and Bose-Einstein statistics (electrons in solids)

    D. Quantized sound waves: phonons and their interactions in solids, ultrasonic attenuation

    E. Superconductivity and superfluidity

        1. Ultrasonic attenuation at the superconducting transition

        2. New sound modes in superfluids

References

1. E. Merzbacher, *Quantum Mechanics* (John Wiley, New York, 1961).

2. A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1966).

3. W. H. Zurek, Phys. Today, October, 36 (1991). Decoherence and the transition from quantum to classical

4. M. Tinkham, *Group Theory and Quantum Mechanics* (McGraw-Hill, New York, 1964).

5. R. N. Schwartz, Z. I. Slawsky, and K. F. Herzfeld, J. Chem. Phys. **20**, 1591 (1952). Calculation of vibrational relaxation times in gases

6. R. N. Schwartz and K. F. Herzfeld, J. Chem. Phys. **22**, 767 (1954). Vibrational relaxation times in gases (three-dimensional treatment)

7. C. Kittel, *Quantum Theory of Solids* (John Wiley, New York, 1963).

8. A. A. Maradudin, E. W. Montroll, and G. H. Weiss, *Theory of Lattice Dynamics in the Harmonic Approximation* (Academic Press, New York, 1971)

9. D. R. Tilley and J. Tilley, *Superfluidity and Superconductivity* (Adam Hilger, Bristol, 1990).

10. M. Levy, *Ultrasonics of High-Tc and Other Unconventional Superconductors,* Physical Acoustics, vol. XX (Academic Press, Boston, 1992).

RESONANT ULTRASOUND SPECTROSCOPY
AND MATERIALS PHYSICS

Albert Migliori
Los Alamos National Laboratory

Resonant Ultrasound Spectroscopy

The use of mechanical resonances to extract material properties has become a mature technique. This lecture will describe the history, physics, and mathematics of Resonant Ultrasound Spectroscopy. In addition, detailed examples of its application to physics, metallurgy and non-destructive testing will be given. Instrumentation, electronics design and data analysis will be described, and a hands-on demonstration will be available.

Reference:

Resonant Ultrasound Spectroscopy, John Wiley and Sons, New York, Migliori and Sarrao.

This will serve as the text for the course, and all transparencies will roughly come from it.

SCANNING ACOUSTIC MICROSCOPY:
LENSES, TIPS AND SONOELECTRONICS

C. F. QUATE
Stanford University

HIGH-RESOLUTION, HIGH-CONTRAST ACOUSTIC IMAGING (*)

J. A. CUNNINGHAM and C. F. QUATE
Stanford University, Stanford, California, U.S.A.

Résumé. — Nous présentons, dans cet article, deux dispositifs qui donnent une imagerie acoustique de haute résolution et de fort contraste. Dans le premier, l'objet est irardié par deux faisceaux acoustiques croisés, les champs propagés formant un système d'ondes stationnaires modulées par les propriétés de transmissions de l'objet. Dans le second, un faisceau acoustique unique tombe normalement sur l'objet qui, là encore, modifie les champs transmis. Dans les deux dispositifs, une couche mince d'une émulsion de petites sphères de latex dans un liquide détecte non linéairement ces ondes transmises dans le champ, tout près de l'objet, pour former une image. Le mécanisme primaire de la détection est la pression de radiation exercée sur les sphères.

Nous présentons des résolutions supérieures à 10 microns pour des densités de puissance modeste de 10^{-3} à 10^{-2} watt/cm². Nous pouvons obtenir une image de spécimens biologiques avec un assez bon contraste, ce qui donne une application pratique de ce dispositif d'imagerie.

Abstract. — In this paper we present two systems capable of high-resolution, high-contrast acoustic imaging. In the first system, the object is illuminated by two intersecting acoustic beams, the transmitted fields forming a standing-wave pattern modulated by the object's transmission properties. In the second, a single acoustic beam is normally incident upon the object which again modifies the transmitted fields. In both systems a thin film emulsion of small (1 micron) latex spheres in liquid nonlinearly detects these transmitted waves in the very near field of the object to form an image. The primary detection mechanism is radiation pressure exerted on the spheres.

We demonstrate system resolutions better than 10 microns at moderate power levels of 10^{-3} watt/cm² to 10^{-2} watt/cm². We are also able to image biological specimens with rather good contrast, one practical application of such an imaging system.

1. Introduction. — We wish to discuss our work on high frequency acoustic imaging where the resolution is comparable with that obtained with an optical system. There are two primary advantages of an acoustic imaging system when it is compared to its optical counterpart. First, optically opaque objects can be probed rather easily with acoustic waves. Second, and perhaps more important, higher contrasts can be achieved for those objects which are nearly transparent to optical waves. This comes about since the change in elastic properties over the cross section of the object is much greater than the corresponding change in the index of refraction. The research that has been carried out on acoustic imaging and the realization of the above goals have been reviewed recently by Mueller [1].

The work has traditionally been divided between acoustic holography and direct acoustic imaging. In either system the method used to convert the acoustic patterns to a visible display is the central problem.

We have previously demonstrated [2] a system for high-resolution acoustic holography which consists of a reference beam and a beam as scattered by the object. These two acoustic beams intersect at a solid surface which contains the imaging film. The two intersecting beams are quite analogous to the optical system for holography [3], but the imaging film is quite different from photographic film. Our imaging film is an emulsion of small latex particles suspended in liquid. The forces exerted on these spheres, or particles, by the two acoustic beams condense the particles into a pattern that reproduces the acoustic hologram. The holograms as recorded in this way, are then reconstructed by conventional optical methods. We believe that this application of « acoustic radiation pressure » is most appropriate for a colloquium in honor of Paul Langevin, who himself made original contributions to this particular branch of physical acoustics [4].

In this report we want to present a continuation of our earlier work [2]. We use the same technique for imaging (as described below) but we replace the large crystals used in the holography experiment with a thin mylar film. The mylar film supports the emulsion of suspended particles and, since the imaging region is very near the object, it is reminiscent of the techniques used for « contact printing » in photography. In this new system we are not able to introduce a reference beam that is separate from the beam as scattered by the object and, as a result, we do not form holograms. However, the images are interesting

in themselves and the resolution is higher than that reported previously. The determining factor for resolution is the wavelength of the sound waves in the liquid object cell and for a fixed frequency the wavelength in liquid is much smaller than that in a solid.

In our research we have developed two rather simple systems with the required acoustic imaging capabilities of high resolution and high contrast. In the first system, the object is illuminated by two intersecting acoustic beams, the transmitted fields forming a standing-wave pattern modulated by the object's transmission properties. In the second system a single acoustic beam is normally incident upon the object, the transmitted fields again being modified by the object's transmission properties. Both imaging systems rely on the nonlinear detection, in the near field, of these transmitted fields. The detector again is a thin liquid film in which small (1 micron) polystyrene spheres are suspended. We believe the main detection mechanism to be the radiation pressure on these spheres with a contributing effect of acoustic streaming created in the liquid. Utilizing these systems we demonstrate resolutions better than 10 microns at sensitivities on the order of 10^{-3} watt/cm^2 to 10^{-2} watt/cm^2.

2. Theoretical considerations. — The general theoretical basis for both our imaging systems is illustrated in figure 1. In our system the particles respond to the
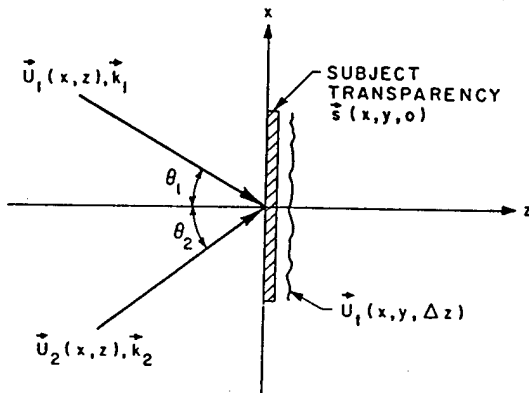
FIG. 1. — General acoustic imaging geometry.

acoustic intensity within the film. We will proceed with the calculation of this parameter for the situation of figure 1. In the two-beam system the object, characterized by its complex transmission function $s(x, y, 0)$, is illuminated by a plane acoustic beam $U_1(x, z)$ with wavevector k_1 at an angle θ_1 and another beam $U_2(x, z)$ with wavevector k_2 at an angle θ_2. Considering the time factor $e^{-i\omega t}$ to be understood for all field quantities we have

$$U_1(x, z) = U_1 \exp[jk_1(-\sin\theta_1 x + \cos\theta_1 z)], \quad (1)$$
$$U_2(x, z) = U_2 \exp[jk_2(\sin\theta_2 x + \cos\theta_2 x)], \quad (2)$$

and

$$s(x, y, 0) = s(x, y, 0) \exp[-j\varphi(x, y, 0)]. \quad (3)$$

Since the two beams are coherent, we have for the total incident field,

$$U_i(x, z) = U_1(x, z) + U_2(x, z). \quad (4)$$

In our cases we have chosen $k_1 = k_2 = k$ and $\theta_1 = \theta_2 = \theta$. If we now make the reasonable assumption that $U_1 = U_2 = U$, then we have for the incident field

$$U_i(x, z) = U \{ \exp[jk(-\sin\theta x + \cos\theta z)] + \\ + \exp[jk(\sin\theta x + \cos\theta z)] \}. \quad (5)$$

Since we are going to detect in the very near field of the object, the transmitted acoustic field may be written to a good approximation as

$$U_t(x, y, \Delta z) \cong U_i(x, \Delta z).s(x, y, 0) \quad (6)$$

or

$$U_t(x, y, \Delta z) \cong U \{ \exp[jk \cos \Delta z] \} \times \\ \times \{ \exp[jk \sin\theta x] + \exp[-jk \sin\theta x] \} s(x, y, 0) \quad (7)$$

where $s(x, y, 0)$ is defined in eq. (3). Hence,

$$U_t(x, y, \Delta z) \cong 2 U \{ \exp[jk \cos \Delta z] \} \times \\ \times \cos (k \sin\theta x) s(x, y, 0) \exp[-j\varphi(x, y, 0)]. \quad (8)$$

The acoustic intensity detected is then given by

$$I(x, y, \Delta z) \equiv | U_t(x, y, \Delta z) |^2 \cong \\ \cong 2 U^2 \cos^2 (k \sin\theta x) s^2(x, y, 0) \quad (9)$$

where the factor of $\frac{1}{2}$ accounts for averaging over time. Using the identity $\cos^2\alpha = \frac{1}{2}(1 + \cos 2\alpha)$ we finally obtain

$$I(x, y, \Delta z) \cong U^2[1 + \cos (2 k \sin\theta x)] s^2(x, y, 0). \quad (10)$$

The main feature of the intensity distribution of eq. (10) is the standing wave along the x-axis with a fringe spacing of

$$\Delta x = \pi/2 k \sin\theta = \lambda/2 \sin\theta. \quad (11)$$

This, of course, is superimposed upon the image intensity distribution $s^2(x, y, 0)$.

The forces on the immersed particles are related to this intensity. In a classical paper [5], King has calculated the force on a rigid sphere in a standing acoustic wave. For our case $k_x a = 0.583$, where a is the radius of the spheres, $k_x = k \sin\theta$, and k is the wavenumber in the solid. The appropriate radiation pressure equation is then

$$\frac{\bar{P}}{\pi a^2} = 2 k_x a \frac{I}{c} \sin 2 k_x xF (\rho_1/\rho_0), \quad (12)$$

where

$$F(\rho_1/\rho_0) = \frac{\rho_1/\rho_0 + [2(\rho_1/\rho_0 - 1)/3]}{1 + 2 \rho_1/\rho_0} \quad (13)$$

and c is the speed of sound in the fluid, ρ_1 and ρ_0 being the respective densities of the spheres and liquid. King has shown that for $F(\rho_1/\rho_0) > 0$ the spheres will move toward a velocity loop of the standing wave and for $F(\rho_1/\rho_0) < 0$ toward a velocity node. This theory has been extended to include the compressibility of the sphere, $\sigma = c_1/c_0 = k_0/k_1$, by Yosioka and Kawasima [6] to yield

$$\frac{\overline{P}}{\pi a^2} = 2 k_x a \frac{I}{c} \sin 2 k_x x F\left(\frac{\rho_1}{\rho_0}, \sigma\right), \quad (14)$$

where now

$$F\left(\frac{\rho_1}{\rho_0}, \sigma\right) \doteq$$

$$= \left[\frac{\rho_1/\rho_0 + \{2(\rho_1/\rho_0 - 1)/3\}}{1 + 2\rho_1/\rho_0} - \frac{1}{3\sigma^2 \rho_1/\rho_0}\right]. \quad (15)$$

The same conditions apply to $F(\rho_1/\rho_0, \sigma)$ as did to $F(\rho_1/\rho_0)$. For the polystyrene spheres in solution used in our experiments, both theories yield a particle distribution corresponding to that expected from eq. (10).

The situation for single-beam imaging is similar to that of figure 1, but now we have

$$U_i(z) = U \exp[jkz]. \quad (16)$$

For the transmitted field this yields

$$U_t(x, y, \Delta z) \cong$$
$$\cong U_t \exp[jk \Delta z] s(x, y, 0) \exp\left[- j\varphi(x, y, 0)\right], \quad (17)$$

and hence an intensity distribution given by

$$I(x, y, \Delta z) \cong \tfrac{1}{2} U^2 s^2(x, y, 0). \quad (18)$$

In this case we must consider the radiation pressure on a sphere due to a plane progressive wave, provided that we neglect the diffraction that takes place in the distance Δz. At the frequency of 1.1 GHz used for single-beam imaging $ka = 2.58$. Fox [7] has extended King's theory for radiation pressure due to a plane progressive wave to include this range yielding

$$\frac{\overline{P}}{\pi a^2} \cong 0.820 \frac{I}{c} \quad (19)$$

for $\rho_1/\rho_0 \simeq 1$. The above analysis will be considered again in a later section when we discuss system sensitivity.

3. The water cell and image detector.

— The water cell and acoustic detection film are shown in the inserts of figures 2 and 5. In both imaging systems the object to be viewed is immersed in the thin water cell. The acoustic beam or beams are passed into this cell by means of acoustic anti-reflection coatings. On top of this cell is stretched a 5 micron mylar film which is acoustically matched to the liquid surroundings. It has an optical anti-reflection coating to avoid direct viewing of the object. This film serves
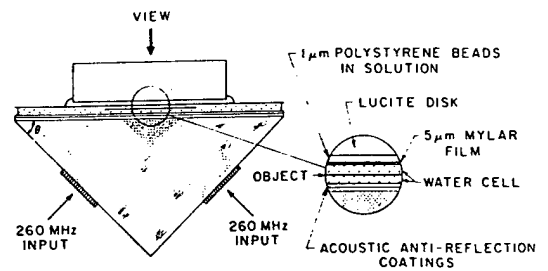
FIG. 2. — Experimental arrangement for two-beam imaging.

the purpose of separating the object space from the detector space. On top of this film a lucite disk uniformly compresses an emulsion of 5 % by volume 1 micron polystyrene spheres in a 90 %-10 % water-glycerine solution. The acoustic image is recorded as a redistribution of these spheres on a 1 : 1 basis. Magnification is then obtained optically by viewing this distribution with a microscope.

4. Experimental results.

— The experimental arrangement used in our two-beam imaging system is shown in figure 2. Two plane acoustic beams at 260 MHz ($\lambda = 23$ microns) are generated in the fused quartz prism by matched $Y/35°$ $LiNbO_3$ plate transducers and intersect at the surface at an angle $\theta = 45°$. This determines a linear fringe spacing in the intensity pattern of 16.2 microns as predicted by eq. (11). The complex transmission function, $s(x, y, 0)$, of the object determines the amplitude of the fringes.

It should also be noted here that since the acoustic fields are passed into the detecting film by means of a mylar membrane, a standing wave pattern is also set up in this membrane. Acoustic streaming effects of the type noted by Jackson and Nyborg [8] could therefore be a contributing factor in this imaging system. A resultant limiting flow velocity similar to that of their eq. (13) would assist in moving particles toward velocity loops of the standing wave pattern.

One of the objects imaged in this system is a nickel mesh with rather ragged circular holes as shown optically in figure 3a. The acoustic image is shown in figure 3b with the superimposed 16.2 μm fringe pattern. Here the object is totally transmissive. The raggedness of the holes is evident in the acoustic image. We also used a square 75 mesh ($\cong$ 3 line/mm) in this system, the optical image of which is shown in figure 4a and its acoustical counterpart in figure 4b. The image quality is good and the fringe pattern stands out clearly. This image demonstrates a resolution of at least 60 microns. An attempt to image a 500 mesh ($\simeq$ 20 lines/mm) showed that ambiguities result in the image when the object has the same periodicity as the standing wave pattern. No attempt has been made to improve image quality by removing the fringes, but this can be easily accomplished by optical spatial filtering [9].
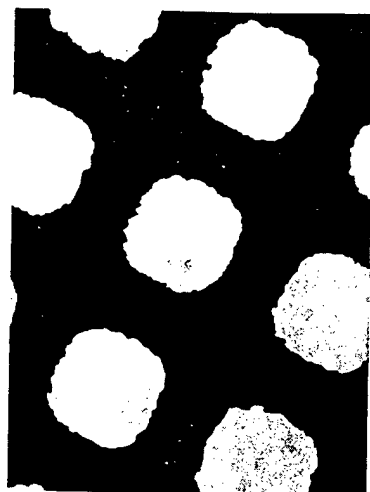
(a)
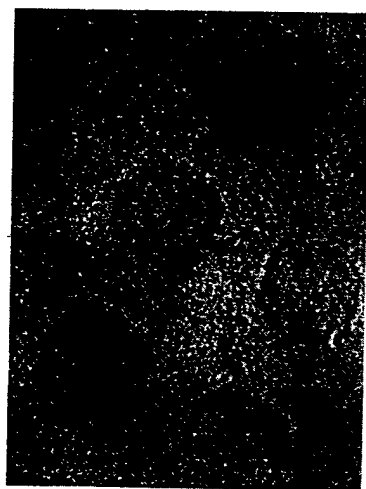
(b)

FIG. 3. — Circular nickel mesh (300 μm holes on 420 μm centers).

a) Optical image : b) Acoustical image.
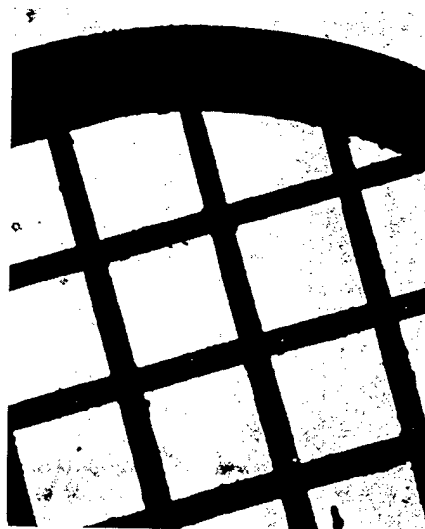
(a)
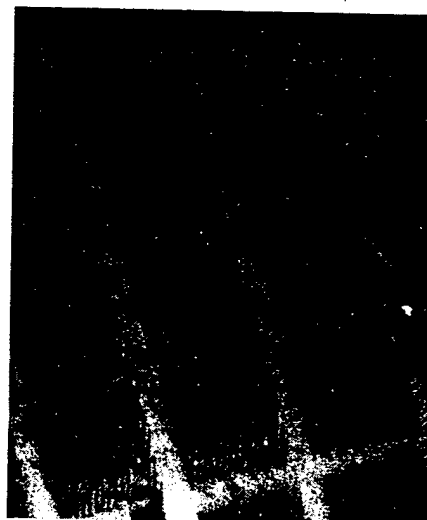
(b)

FIG. 4. — Square 75 mesh (280 μm holes and 60 μm bars).

a) Optical image : b) Acoustical image.

Our single-beam experimental set-up is shown in figure 5. Here a thin film ZnO transducer excites a 1.1 GHz plane acoustic beam into the buffer rod. This buffer rod can be almost any solid material, but we have chosen to use YAG because of its low acoustic attenuation and high isotropy factor. It should be noted that this system is capable of an extremely large field of view, being limited only by the size of the transducer. Again, as in the case of the two-beam system, the mylar film becomes a source of acoustic streaming as well as radiation pressure [10]. Due to velocity gradients created by the differential transmission properties of the object, streaming again has an effect on the acoustic imaging [11].

As an estimate of the resolution capability of this system we have imaged a portion of the 1951 USAF Test Resolution Chart shown optically in figure 6a. As can be seen in the acoustic image of figure 6b, Group + 5 subgroup (5) is clearly distinguished. This demonstrates a resolution of better than 51 lines/mm or slightly less than 9 microns. The wavelength in the emulsion is 1.35 micron. The resolution is lower than we would expect for two reasons. First, due to the minuteness of the elements of the resolution chart they may not be filling with water and hence not passing acoustic energy. Second, the diffractive spreading of the beam within the thickness of the mylar film imposes a lower limit on resolution. If we assume the mylar thickness to be Δz (= 5 microns)

146
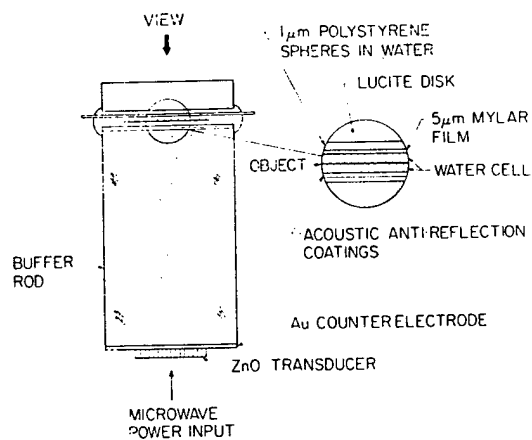
FIG. 5. — Experimental arrangement for single-beam imaging.

and two object points to be separated by a distance d, then in order to be resolved we must satisfy the inequality

$$d - 2 \Delta z \tan \alpha \gtrsim 0, \qquad (20)$$

where

$$\sin \alpha = \lambda_{mylar}/d. \qquad (21)$$

(a)

$\rightarrow| \; |\leftarrow \quad 100 \; \mu m$
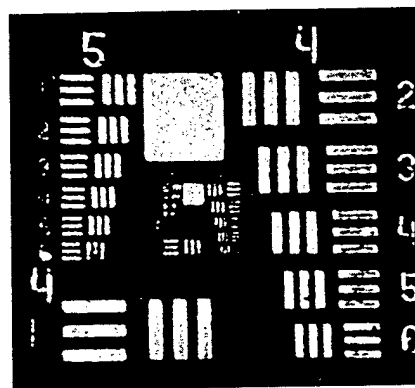
(a)
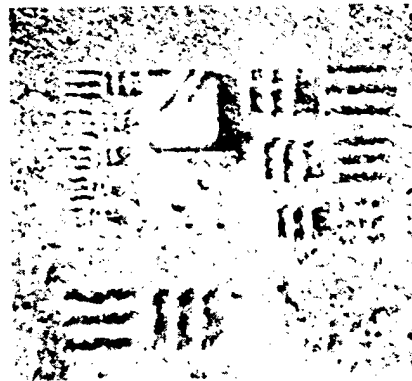
(b)

FIG. 6. — 1951 USAF Resolution Test Chart ; Groups + 4 and + 5.

a) Group + 4		b) Group + 5	
(1) 16.00 Li/mm		(1) 32.00 Li/mm	
(2) 17.96	—	(2) 35.92	—
(3) 20.32	—	(3) 40.64	—
(4) 22.80	—	(4) 45.60	—
(5) 25.56	—	(5) 51.12	—
(6) 28.51	—	(6) 57.02	—

$\rightarrow| \; |\leftarrow \quad 100 \; \mu m$

(b)

FIG. 7. — Microtomed section of human lung tissue.
a) Optical image : b) Acoustical image.

147

For the frequency at which we are operating this would place a lower limit on our resolution capability of approximately 4 microns. Hence, we are only a factor of two above the diffraction-limited resolution.

One primary purpose of our research has been the high-resolution acoustic imaging of biological specimens with good contrast. Shown optically in figure 7a is a microtomed section of human lung tissue which is embedded in epoxy. As can be seen from the acoustic image in figure 7b, the resolution achieved is as good as that obtained optically and the contrast is very high. There also seems to be structure in the acoustical image which does not appear in the optical image. We see this ability to image biological specimens as one practical application of such an imaging system.

5. **Sensitivity.** — The downward pressure on the spheres due to gravity counteracted by buoyancy is given by

$$\frac{\overline{P_D}}{\pi a^2} = (\rho_1 - \rho_0) \frac{4}{3} ag \qquad (22)$$

and is on the order of 4×10^{-3} dyne/cm^2 for our experimental conditions. In the two-beam imaging scheme the theories of King or Yosioka and Kawasima predict radiation pressures on the order of 4 (dyne/cm^2)/(watt/cm^2). Since we must only push the spheres sideways and not lift them against gravity, we may safely assume a force of 0.1 $P_D/\pi a^2$ needed to form our fringe pattern. This would predict a sensitivity of approximately 10^{-4} watt/cm^2. Our experimentally determined sensitivity for this configu-

ration, however, lies in the range of 10^{-3} watt/cm^2. For the single-beam imaging system, which is operating at a much higher frequency and thus higher losses, the theory of Fox predicts a sensitivity of approximately 10^{-3} watt/cm^2. Again, our experimentally determined sensitivity is an order of magnitude higher at approximately 5×10^{-2} watt/cm^2. This order-of-magnitude discrepancy in both systems can be accounted for in terms of the increased attenuation in the detecting emulsion over that of pure water. This attenuation has been found to increase linearly with particle concentration for all conditions. At high frequencies or for large radius spheres, scattering becomes the dominant loss factor [12], [13].

6. **Summary and conclusions.** — In this paper we have presented two systems capable of high-resolution acoustic imaging by the nonlinear detection of acoustic fields in a thin emulsion film. We have demonstrated a resolution capability of 9 microns at a sensitivity on the order of 5×10^{-2} watt/cm^2 with our latter system. We have also been able to image biological specimens with rather good contrast, as expected. By using thinner mylar films and operating at higher frequencies we should be able to increase our resolution capabilities without degrading contrast sensitivity.

References

[1] Mueller R. K., « Acoustic Holography », Proc. IEEE 59 (1971) 1319.

[2] Cunningham J. A. and Quate C. F., « Acoustic, Interference in Solids and Holographic Imaging » in Acoustical Holography, vol. 4, G. Wade, Ed. (Plenum Press, New York, 1972), 667.

[3] Collier R. J., Burckhardt C. B. and Lin L. H., Optical Holography (Academic Press, New York, 1971).

[4] Langevin P., Lectures given at the Ecole Supérieure de Physique et de Chimie, Paris, France (1923).

[5] King L. V., « On the Acoustic Radiation Pressure on Spheres », Proc. R. Soc. London A 147 (1934) 212.

[6] Yosioka K. and Kawasima Y., « Acoustic Radiation Pressure on a Compressible Sphere », Acustica 5 (1955) 167.

[7] Fox F. E., « Sound Pressure on Spheres », J. Acoust. Soc. Am. 12 (1940) 147.

[8] Jackson F. J. and Nyborg W. L., « Small Scale Acoustic Streaming near a Locally Excited Membrane », J. Acoust. Soc. Am. 30 (1958) 614.

[9] Goodman J. W., Introduction to Fourier Optics (McGraw-Hill, New York, 1968), chap. 7.

[10] Cady W. G. and Gittings C. E., « On the Measurement of Power Radiated from an Acoustic Source », J. Acoust. Soc. Am. 25 (1953) 892.

[11] Kolb J. and Nyborg W. L., « Small-Scale Acoustic Streaming in Liquids », J. Acoust. Soc. Am. 28 (1956) 1237.

[12] Epstein P. J. and Carhart R. R., « The Absorption of Sound in Suspensions and Emulsions I. Water Fog in Air, J. Acoust. Soc. Am. 25 (1953) 553.

[13] Allegra J. R. and Hawley S. A., « Attenuation of Sound in Suspensions and Emulsions : Theory and Experiments », J. Acoust. Soc. Am. 51 (1972) 1545.

DISCUSSION

P. Das. — 1) Could you tell us something about the transducers, like bawdwidth, insertion loss, etc...?

2) How much power did you use for the pictures you showed in the slides ?

C. E. Quate. — 1) The transducers are standard

as make of ZnO films (sputtered). The insertion loss is 5 dB and the bandwith is 100 MHz.

2) For the holograms which were shown at first we use approximately 10^{-3} watt/cm^2. For the others we use about 10^{-2} watt/cm^2.

Acoustic microscope—scanning version

R. A. Lemons and C. F. Quate

Microwave Laboratory, Stanford University, Stanford, California 94305
(Received 22 October 1973)

This letter reports the development of a mechanically scanned acoustic microscope showing 10-μm resolution. Using single-surface lenses an acoustic beam is focused with negligible spherical aberration in a water cell. The image is formed by mechanically·scanning an object through this focused beam in a raster pattern. Transmitted power is detected with a piezoelectric transducer, and this signal modulates the synchronized raster of a CRT display. By employing piezoelectric detection, sensitivities of 10^{-8} W/cm^2 are obtained, yielding images of excellent clarity and contrast.

A microscope can be made with acoustic waves rather than optical waves. We have set out to explore the potential of such an instrument and compare its performance with that of the optical version. The acoustic microscope would provide us with a new method for delineating the detail of microscopic objects if it could be made with a resolution comparable to that of the optical instrument.[1] In principle this can be done. With a frequency of 1 GHz the wavelength for sound in water is near 1.5 μm. The utility of such an instrument arises because the scattering of acoustic waves is dependent on the change in elastic properties, whereas it is the change in the index of refraction that determines the scattering of optical waves. This fundamental difference suggests that acoustic radiation, responding to structures within the object, may allow one to resolve details which are different from those recorded with optical waves. Often, objects which are transparent to optical radiation show considerable acoustic contrast. This increase in contrast will allow us to view structural details within an object without chemical staining. Moreover, the optically opaque objects frequently allow acoustic transmission, enabling internal structure to be observed. This is the basis for our interest in an acoustic microscope.

Such an instrument has not yet been fully exploited and we would like to report here on our progress toward a scanning version. This system may have some advantage over the two other systems which have been previously reported.[2-4] We have determined that it is possible with a most simple lens to focus an acoustic beam with negligible spherical aberration. Thus the potential resolution is limited only by the acoustic wavelength. This is in contrast to the previous two systems wherein the recording process itself can influence and degrade the ultimate resolution. In addition, the detector in the scanning version is a piezoelectric, which is the most sensitive detector available for acoustic radiation. We can therefore work with average sound intensity levels of 10^{-8} W/cm^2, monitoring either variations of acoustic

(a)

200 μm

(b)

200 μm

FIG. 2. Comparison of the acoustic (a) and optical (b) images of a 200-mesh copper electron microscope grid.
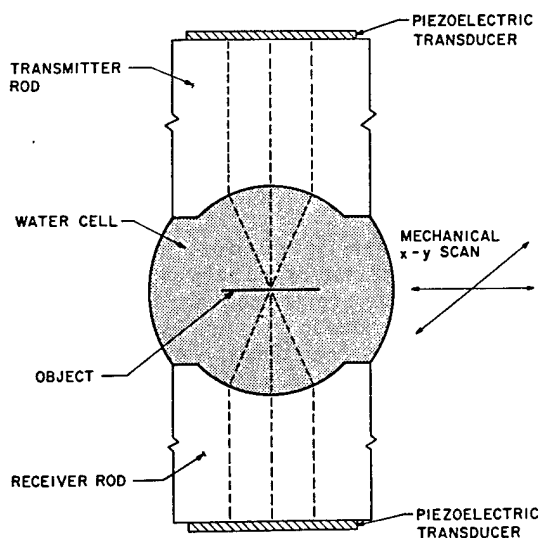
FIG. 1. Schematic showing the lens configuration.

(a)   300 μm
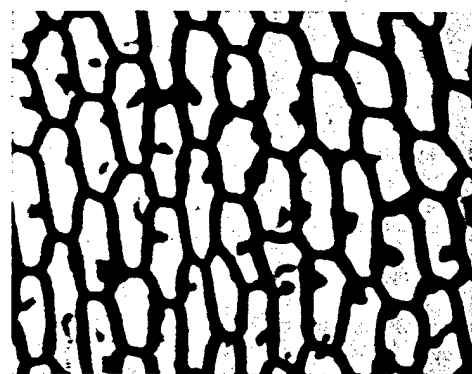
(b)   300 μm

FIG. 3. Comparison of the acoustic (a) and optical (b) images of onion cells.

intensity or variations of acoustic phase. This level is well below the damage threshold for biological specimens. Finally, the scanning rate is sufficiently high that image formation time is reduced to less than 1 sec.

The heart of our system is a pair of single-surface acoustic lenses. Each is a polished concave spherical surface ground into the end of a crystal rod. These two lenses are positioned as mirror images with a water cell filling the space between them (see Fig. 1). The water serves the dual function of providing a slower refracting medium and of making acoustic contact between the lens surfaces and the object to be imaged. On the end of each rod, opposite the lens surface, a piezoelectric transducer is applied. One of these transducers acts as the transmitter, generating a plane acoustic wave. This acoustic power is then focused by the lens into the water cell. The specimen itself is placed at the minimum waist of this beam. Since the two lenses are spaced so that their foci are coincident, the power transmitted by the specimen is recollimated by the second lens. Thus, all the transmitted power from a specimen point is detected with the second piezoelectric

transducer. Accordingly, this transmission geometry has the dual advantage of eliminating spurious signals while fully utilizing the sensitivity of piezoelectric detection.

To produce an image with this system the specimen is mechanically translated through the beam in a raster pattern. This motion is synchronized with a raster on a CRT display, and the output of the detector modulates the display beam current. In this way, modulation of the acoustic beam by the object is converted directly into a brightness pattern or image.

The specimen is attached to a small piston which rides in a cylinder that constrains both lateral and vertical motions to less than 1 μm. The horizontal component of the mechanical motion is then achieved by affixing the free end of the piston to the cone of a dynamic speaker. In practice this speaker is driven sinusoidally and can provide up to 3-mm displacement with scan rates of 300 lines/sec. Since the voltage which drives the speaker also drives the X axis of the display, synchronization is simplified. The vertical motion is provided by mounting the speaker assembly on a motor-driven stage. This motor also drives a potentiometer, providing the Y-axis deflection for the display.

Clearly, the resolution of this system is determined by the diameter of the focused beam in the object plane. If an optical system of the same design were constructed, one would find that the spherical aberration inherent to a single spherical surface would severely limit the resolution. However, the precision with which the beam is focused depends crucially upon the ratio of the propagation velocities at the crystal-water interface. By maximizing this ratio, the spherical aberration can be greatly reduced. This can be understood in a qualitative way by a simple application of Snell's law ($\sin\theta_2 = C_2/C_1 \sin\theta_1$). As the ratio of propagation velocities (C_2/C_1) is reduced, the angle θ_2 between the refracted ray and the intersecting radius is likewise reduced. In the limit $C_2/C_1 \to 0$ all rays would converge on the center of curvature, and the spherical aberration would be exactly zero. A more quantitative calculation based on first-order aberration theory shows that the spherical aberration is roughly proportional to $(C_2/C_1)^2$. Fortunately, in acoustics available materials allow much smaller values of C_2/C_1 than one is accustomed to in optics. For a sapphire-water system $C_2/C_1 = 0.134$. This means an $f/0.8$ lens can be diffraction limited down to a resolution of 1 μm.

In our present system the lenses are ground with a 1.59-mm radius of curvature and have an f number of 0.92. The piezoelectric transducers are 35 Y-cut LiNbO$_3$ platelets operated in their fifth harmonic at a frequency of 160 MHz. Accordingly, the acoustic wavelength in the water is 9.4 μm, and the diffraction-limited resolution of the lens is ~10 μm.

The image-forming capabilities of this system are illustrated in Figs. 2 and 3. In each case the acoustic image can be compared with the corresponding optical image. In Fig. 2 the acoustic image of a 200-mesh copper electron microscope grid is shown. The excellent contrast and definition seen in the acoustic image

Appl. Phys. Lett., Vol. 24, No. 4, 15 February 1974

150

is one of the chief advantages of this microscope. In Fig. 3 the acoustic image of a layer of onion cells is shown. The most prominent details here are the individual cell walls. Both the marked contrast of their structure and the appearance of internal cell detail suggests the potential usefulness of acoustic microscopy. In addition, analysis of these and other images indicates that the resolution of our system is very near the 10-μm diffraction limit.

These results are sufficient to allow us to comment on the improved resolution that can be attained by moving to higher frequencies. We will be limited by the acoustic attenuation in water which increases as the square of the frequency. We can partially compensate for this increase by shortening the liquid path length and this in turn will require a lens with a smaller radius of curvature. We have successfully fabricated and tested a pair of lenses with a focal length of 0.46 mm, and we are confident that these can be operated at 400 MHz with a resulting resolution of 3 μm. At 1000 MHz the acous-

tic attenuation in water is approximately 145 dB/mm (at 35 °C)[5] and lenses with 0.1-mm radius of curvature would be needed. The resolution of an acoustic microscope at this frequency would be equal to 1.2 μm.

[1]This was pointed out to us by Professor R. Kompfner who was the first to realize the potential of acoustic microscopy.
[2]L.W. Kessler, P.R. Palermo, and A. Korpel, *Acoustic Holography*, Vol. 4 (Plenum, New York, 1972), p. 51.
[3]J.A. Cunningham and C.F. Quate, J. Phys. Colloq. C-6 Suppl 33, 42 (1972).
[4]J.A. Cunningham and C.F. Quate, *Acoustical Holography*, Vol. 5 (Plenum, New York, to be published).
[5]J.M.M. Pinkerton, Nature 160, 128 (1947).

Acoustic Microscopy: Biomedical Applications

Variations in elastic properties within a single cell can now be visualized with the acoustic microscope.

Ross A. Lemons and Calvin F. Quate

The microscope in its various forms is one of the most powerful and widely used scientific tools. In the fields of biological and medical sciences advances of great significance have resulted from observations made with these instruments. The microscope enables us to explore the structural details beneath the limits of our direct visual perception. The variety of phenomena in this miniature world is so large that it is imperative that we examine all forms of radiation that can be used for such exploration. Optical waves in the optical microscope and electron waves in the electron microscope are the most common forms. It is the wavelength of this radiation that determines the minimum size of the object that can be distinguished. With recent technological advances it is now a straightforward matter to generate acoustic energy with a wavelength comparable to that of light. It follows that this could provide us with a new instrument for microscopy (*1*).

A scanning acoustic microscope based on acoustic radiation with this short wavelength has now been developed to the point where it can be used to resolve detail in biological material at the subcellular level (*2*). This new form of radiation in the field of microscopy should permit us to increase our understanding of that material and to gain further insight into the processes that go on within cells and microorganisms. This expectation is based upon the fundamental distinction between light and sound and upon the interaction of these different forms of radiation with the material of the specimen. In contrast to the dielectric properties of the object which determine the content of an optical micrograph, it is the elastic properties that determine the content of an acoustic micrograph. Variations in the elastic constants can be revealed by changes in the amplitude and phase of an acoustic wave which passes through the specimen. In this article we will deal primarily with the changes in

amplitude. This varying attenuation can be produced either directly by absorption or by the scattering arising from the acoustic impedance mismatch between a region of the specimen and the surrounding material.

With the work to date we have learned that biological material—unstained—can exhibit high contrast when subjected to acoustic radiation. For example, hemoglobin (*3*) and connective tissue (*4*) show strong attenuation and these components can be easily identified in the acoustic micrographs. Furthermore, the instrument is sensitive to changes in viscosity and this should be of some advantage in rheological studies.

In this article we demonstrate the potential of the acoustic microscope and show that it represents an alternative technique for recording microscopic images that can complement and enhance the information that is provided by the light microscope. We hope that it will attract the interest of those who are professionally engaged in research with biological systems and stimulate them to consider the most appropriate areas for exploiting this new technology.

Instrument Design

The acoustic microscope that we will discuss is based upon a scanning principle wherein the specimen is moved through a focused acoustic beam. This instrument has been operated with acoustic frequencies as high as 1000 megahertz. In liquid water the wavelength corresponding to this frequency is 1.5 micrometers and the resolution approaches 1 μm (*5*). We are confident that further work will allow us to improve the resolution, but the present version is sufficient to reveal significant features in biological systems.

The critical components of the acoustic system are shown schematically in Fig. 1.

Incoming electromagnetic energy is converted to acoustic energy by a piezoelectric film at the surface of the sapphire crystal. The resulting acoustic wave propagates as a collimated beam through the sapphire. At the opposite end of this crystal a concave spherical lens transmits the sound waves into a water cell. Since the velocity of sound in water is lower than its velocity in sapphire, the concave surface acts as a positive lens, focusing the beam in the water.

At first glance a single surface lens might seem wholly inadequate to focus a collimated beam to a waist of micron diameter. In an optical system of this design the spherical aberration inherent to such single surface lenses would limit the waist of the beam to a diameter of several wavelengths, and the resolving power would be seriously degraded. Optical intuition, however, is not directly applicable to an acoustic lens. This is because of the large velocity difference between sound waves in sapphire and in water. The ratio of velocities for sound in these two materials is 7.45 whereas a typical ratio for optical waves would be 1.5. A simple geometric calculation shows that the spherical aberration is inversely proportional to the square of this ratio and, therefore, this limitation on resolution can be neglected in the acoustic system (*6*).

The remaining limit on resolution is imposed by the effects of diffraction and this can only be extended by using the highest frequency in order to reach the shortest possible wavelength. Unfortunately, the attenuation of sound increases drastically as the frequency is increased. To give some idea of the magnitude of this effect, the attenuation in water at an acoustic frequency of 1 gigahertz is approximately 200 decibels per millimeter. This means that the power in the wave is reduced by 20 orders of magnitude each millimeter it propagates. The decibel loss per unit length will increase as the square of the frequency and it is necessary, therefore, to fabricate lenses of very small radii of curvature. In the present system the lenses have a focal length of 0.15 mm. We are not limited by spherical aberrations and these small lenses can be made with an F number of 0.65. This value is necessary if we are to maximize the resolution. It is the simplicity of single surface lens design that permits us to realize the combined requirements of small size and large aperture.

The sapphire crystal with its generating transducer on one face and a lens surface on the opposite face can be thought of as a
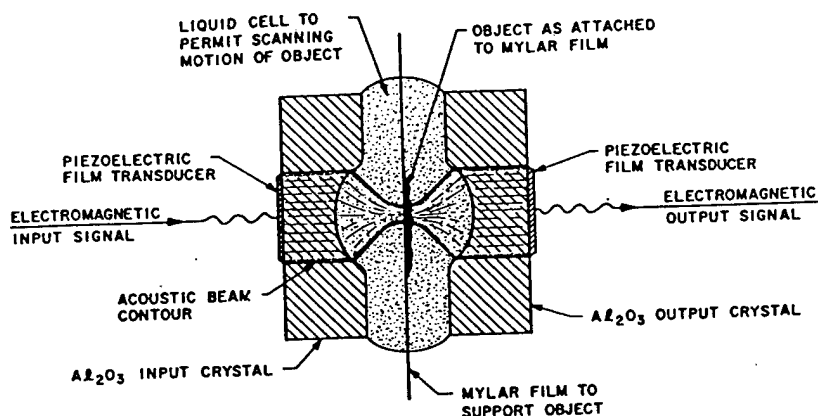
Fig. 1. Schematic diagram of the acoustic system showing the lens configuration.

transmitter which generates an acoustic probe of micrometer size. The remainder of the acoustic system consists of a receiving crystal of exactly the same design. This second crystal has a lens surface which faces the transmitting lens. The two lenses are confocal and in this configuration the energy which diverges from the acoustic focus will be collected and collimated in the receiver crystal. At the output end of this crystal a piezoelectric film reconverts the oscillatory strain of the incident acoustic wave into an electromagnetic signal.

The symmetry of this confocal geometry gives us an additional advantage. When the receiver has its focus coincident with that of the transmitter, the effective resolution of the complete system is enhanced. The basis for this enhancement lies with the directional sensitivity of the detecting transducer. The piezoelectric film is sensitive both to the amplitude and to the phase distribution of the wave incident on its surface. This produces a pattern of sensitivity at the focus which is of the same form as the energy distribution generated by the transmitter. The product of these two functions gives an effectively narrowed focal beam diameter. A detailed analysis shows that with a large signal-to-noise ratio a confocal system can resolve spatial frequencies twice as high as can a single lens alone (7). Another advantage of the confocal geometry is that the signal-to-noise ratio of the system is improved. There is no background signal: the information that is detected is limited to that coming from the focus.

The acoustic properties of a specimen are measured by first affixing the sample to a 2-μm-thick Mylar membrane. In turn this membrane is mechanically translated through the focal plane of the instrument. Within the limited working space between the lenses the Mylar membrane provides a thin, strong support for the sample. Mylar

has an additional advantage in that it has an acoustic velocity only slightly greater than the water which surrounds it. This helps minimize the diffraction spreading and reduces the acoustic reflections at the interface.

To obtain the two-dimensional image of the object, the object itself is scanned through the beam waist in a raster pattern. The use of mechanical scanning has been nearly abandoned in the field of macroscopic acoustic imaging, but on a microscopic scale it proves to be a highly practical technique. By definition, a microscopic object will be not only of small size but of small mass as well. Because of this, scanning can be done quite rapidly. A typical image covers a square 0.25 mm on a side and contains 5×10^4 elements of information. This area can be scanned in approximately 1 second.

The mechanism that provides the mechanical scanning of the object is shown schematically in Fig. 2. For clarity, the two lenses have been drawn apart as they are when a new specimen is mounted. The Mylar membrane that supports the object is stretched across a thin metal ring. This ring is then connected to the movement of a loudspeaker cone by a rod running through a closely machined sleeve. The loudspeaker is driven sinusoidally in time and thus provides the fast, or x-directed, scan. The y motion is brought about by displacing the whole assembly of loudspeaker and object holder in a vertical direction in a slow and uniform manner. This assembly is displaced by a hydraulic piston driven with a small pump.

The acoustic image is formed by displaying the output signal from the receiver element on a cathode-ray display. There the electron beam of the display is moved in synchronism with the mechanical motion of the object, and the signal arising from the transmitted acoustic beam is used

to modulate its intensity. In this way the acoustic transmission of each point on the object will be mapped onto the display screen. An advantage of this display technique is that any desired magnification can be realized by simply adjusting the voltages that deflect the electron beam on the display. Focusing of the acoustic image is accomplished by displacing the entire object assembly with a differential micrometer in the axial or z direction. When the object plane coincides with the narrowest cross section of the acoustic beam the sharpest focus is achieved.

When the lenses are brought together in the viewing position the liquid that surrounds the object is held in place purely by surface tension. This liquid serves to maintain a continuous acoustic path as the object is scanned relative to the focused beam. The choice of water was made not only because it is compatible with biological systems. Water also has one of the lowest attenuations known to occur in those liquids with small acoustic velocities. Small acoustic velocity is, of course, desirable because this improves the resolution that can be achieved for a given input frequency. The large absorption of liquids, although limiting the highest obtainable resolution, does have one desirable feature. At the acoustic frequencies normally used in our instrument the attenuation in the water cell is high enough so that stray acoustic radiation caused by multiple reflections does not cause degradation in the image, provided care is taken to shield the electrical input and output connections from each other.

The remaining mechanical components of the acoustic microscope serve to align one lens with respect to the other. A coincidence of the lens foci is necessary and, for a beam of micrometer diameter, this requires highly precise alignment. To achieve this alignment our system employs an orthogonal pair of elastic levers to position in the x-y plane. The separation in the z direction is adjusted with a differential micrometer mounted on a precision translation stage. This translation stage also allows the lenses to be widely separated when the sample is changed. Adjustment of the relative lens position is made to give a maximum signal through the acoustic system.

It is the basic simplicity of the construction for this kind of scanning microscope that makes it possible to approach the theoretical limit in resolution. Other more elegant but more complicated schemes have been tried but their performance is inferior to that permitted by diffraction. Mechanical scanning seems to be a small price to pay for such a large benefit.

Work with Mammalian Cells and Tissue Sections

The acoustic microscope in the form as described above has been used to examine a variety of specimens. We have accumulated a number of acoustic images and are beginning to catalog the response of various biological materials to acoustic waves. In this initial survey we have seen clear distinctions between the corresponding acoustic and optical micrographs.

The samples that we have worked with can be conveniently divided into three groups. First we have results with simple cell systems containing isolated cells of only a few types. We next deal with living cells grown in culture. The last and largest group are tissue sections which contain a complex collection of cell types. We include here a selection of acoustic micrographs from each of these categories. The majority of these images were obtained with an acoustic frequency of 600 Mhz (λ = 2.5 μm). The more recent images, however, were made at a frequency of 900 Mhz (λ = 1.7 μm) with a resulting resolution which approaches 1 μm. In presenting these images a convention was chosen such that points on the sample with greater acoustic transmission are lighter in the corresponding image. Thus, dark areas correspond to regions of large acoustic attenuation.

Simple cell systems. As an illustration of this category we show in Fig. 3 a comparison between the acoustic and optical images of a human bone marrow smear (*8*). This sample was prepared by smearing the fresh material onto the 2-μm-thick Mylar support membrane. The smear was then fixed in formalin. The most numerous components in this smear are the erythrocytes and they show the largest acoustic attenuation. In addition, a number of granulocytes can be seen. These cells are the developmental elements of the leukocyte family. In the acoustic micrograph they appear larger than the erythrocytes and they have a markedly lower acoustic attenuation. In some instances, details within the granulocytes can be seen. The 900-Mhz acoustic image corresponds closely with its optical counterpart. That the shapes and outlines of individual cells are so clear is impressive to us because the red corpuscles average 7.5 μm in diameter and this is less than 4.5 acoustic wavelengths. The most striking feature, however, is the superb contrast seen in the acoustic image. For the optical image, Wright's-Giemsa's stain was applied to enhance the contrast. Thus the nuclei stained blue, allowing them to be easily distinguished from the cytoplasm; without the stain the cellular detail in the
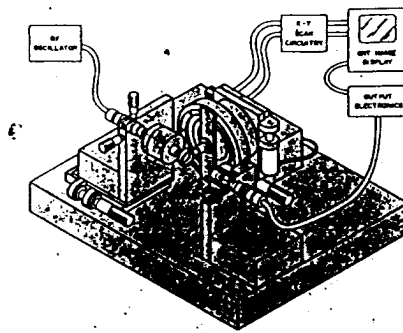
Fig. 2. Generalized diagram of the scanning acoustic microscope.

light micrograph would be poorly defined.

In the acoustic image, the contrast that highlights the red blood corpuscles is due primarily to the increased absorption of the hemoglobin within the cell. This absorption has been the subject of previous studies with acoustic waves at low frequencies (*3, 4*). These studies show the absorption to be related both to the viscosity and to the molecular structure of the material. The acoustic microscope is therefore sensitive to these parameters, and can map them on a microscopic scale. Accordingly,

the acoustic microscope may be useful in studying cells which exhibit abnormalities in these two properties.

Living normal human diploid cells. Cell cultures represent an important component in experimental work on living cells, normal and abnormal. We have only begun to explore the application of the acoustic microscope to this field of work, but the micrograph shown in Fig. 4 demonstrates the compatibility of the acoustic microscope with a culture of living human cells. To date we have demonstrated that cells can be cultured on the Mylar support membrane used in the acoustic microscope and that they stick to the membrane with sufficient tenacity to withstand the forces that are encountered in a mechanical scanning system of this type. In addition, a growth medium can be substituted for the water between the acoustic lenses in order to sustain the life of these cells.

Even though the lung fibroblast cells shown in Fig. 4 are nearly confluent, the extented spindle shape of individual cells can be seen. The nuclear region of an individual cell is acoustically more attenuating than are the processes which extend from it. In some instances, small areas of re-
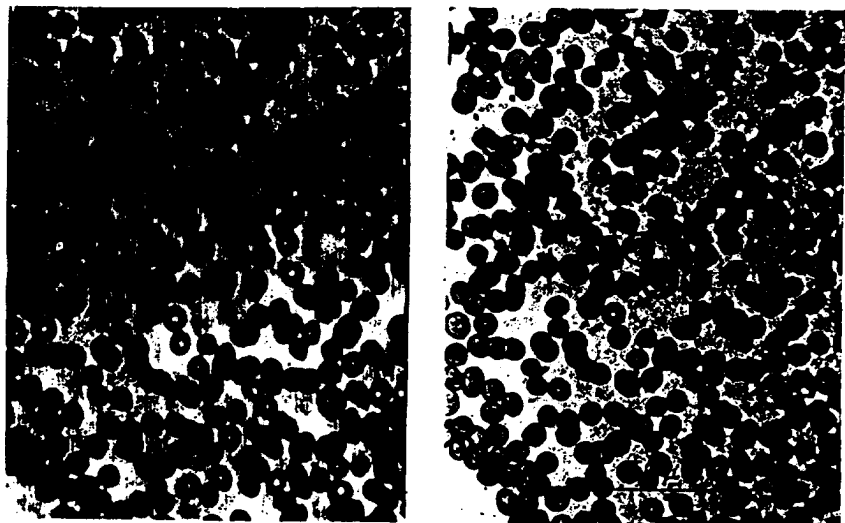
Fig. 3. Comparison of the 900-Mhz acoustic image (left) with the optical image (right) of a human bone marrow smear.

Fig. 4. Acoustic image (900 Mhz) of a living culture of normal human diploid lung fibroblasts.

154

duced transmission can be seen within the nuclear region; these probably correspond to the nucleoli, which often occur in pairs. The acoustic contrast of these cells is not great. In an optical microscope, however, the cells would be so nearly transparent that phase contrast techniques would have to be used to visualize them. By using acoustic phase contrast techniques it might also be possible to enhance the contrast of cells examined with the acoustic microscope.

The prime advantage of the acoustic microscope lies in the different sources of acoustic and optical contrast. For example, in liquids the acoustic absorption increases with the viscosity. Therefore, changes in viscosity should appear as changes in the absorption of the specimen. With the acoustic microscope it should be a direct and simple matter to observe those viscosity changes. Such a study might well give us new insight into the differences between normal and abnormal cells.

Normal tissue sections. In the next series of micrographs we present the images of a number of tissue sections that illustrate the power of the acoustic microscope in viewing complicated cell systems. Samples of tissue were originally fixed in formalin and were subsequently embedded in paraffin. Each of the specimens was then cut from the paraffin block with standard microtome techniques to a nominal 5-μm thickness. These sections were then mounted onto the Mylar supporting membrane and the paraffin was removed. All of the tissues were unstained and, therefore, the acoustic response is typical of what can be expected from the tissue as altered only by the procedures of fixation and embedding.

In Fig. 5 a number of the characteristic features of human lung tissue are evident. In the lower left corner is an alveolar duct with a number of adjacent alveoli. The individual cells which comprise the walls of the alveolar sacs can be seen along with several capillaries. At the lower right is the folded epithelium of a bronchiole. A small vessel is also evident just above the center of the image.

Figure 6A shows a section of the human fallopian tube with its characteristic deep branching folds of the mucous membrane. The most striking feature of the image is the sharp contrast between the internal matrix of connective tissue and the epithelial layer. As will be seen in the following figures, the large acoustic attenuation of connective tissues holds for a variety of specimens. In this instance the connective tissue is loose with numerous fibroblasts. The columnar structure of the epithelium is evident in some regions of the image, and some of the subtle detail within cells can also be seen. Of particular interest are the highly attenuating points at the outer boundary of certain epithelial cells. The epithelial layer is composed of both ciliated columnar cells and nonciliated secretory cells. The points of high attenuation seen in the acoustic image probably indicate the ciliated components.

The most prominent feature of the section of human spleen shown in Fig. 6B is the oblique view of a small artery. Inside the vessel a great number of red blood cells can be seen. Within the wall of the vessel the internal elastic membrane is the most distinctive structure. This membrane strongly attenuates the acoustic beam and thus appears as a folded black line around the inside of the artery. In contrast, optical micrographs show the elastic membrane as a translucent structure unless it has been specifically stained. The band of muscle tissue surrounding the artery is also clearly distinguished in the acoustic image. Outside of the vessel the highly absorbing erythrocytes are contrasted with the much lower average attenuation of the spleen tissue.

As a final illustration of the acoustic appearance of normal human tissue, Fig. 6C shows a section of mammary gland. At the top of the figure a small group of secretory tubules can be seen surrounded by a highly attenuating mass of connective tissue. The attenuation of this intralobular connective tissue is on the average considerably larger than that of the interlobular connective tissue as seen at the lower right. Using this normal tissue as a point of reference we will now proceed to discuss the acoustic appearance of breast tissue afflicted with cancer.

Fig. 5. Acoustic image (600 Mhz) of an unstained section of human lung tissue.

155

Tissue Sections Showing Pathology

Whenever a cell is modified by disease or other agents it is likely that some of its elastic properties are changed as well. This makes an investigation of the acoustic response of such material interesting and important. Our preliminary work in this area indicates that there may be significant diagnostic applications of the acoustic microscope.

This potential can be illustrated with some comparative acoustic and optical images of a sample from a malignant tumor of the human breast. A panoramic acoustic view of a section of this material is presented in Fig. 7A. After the acoustic image of this specimen had been recorded, the section was stained with hematoxylin-eosin (HE) so that a comparative optical micrograph could be made. Thus, Fig. 7B shows an optical image of the same area. This micrograph was made by means of a Zeiss Photo-Microscope II with a × 10, numerical aperture 0.32, planapo objective. A close inspection of Fig. 7 will show that the basic structural features seen in the acoustic micrograph correlate on a one-to-one basis with the optical image. This carcinoma is of the infiltrating ductile variety, as evidenced by the abnormal

duct-like structure seen in the upper right corner of each image. In order to facilitate a comparison of the detail in the acoustic image with that seen optically, one area of interest has been selected for enlargement. The acoustic and optical images of this enlarged area are shown in Fig. 8, A and B, respectively. Each individual cell can be compared in these two images.

The emphasis of particular details is, however, very different in the two images. In the optical image the HE stain provides a clear distinction between the cell nucleus and the cytoplasm, while distinctions between cell types are less marked. In contrast, differences in acoustic attenuation tend to discriminate between cell types. In some instances a difference between two areas can be quite subtle in the optical image while being obvious in the acoustic micrograph. An example of this can be seen in Fig. 8A. In the central area of this image several small localized regions show a very large acoustic attenuation compared to the surrounding material. The same regions in the optical image, Fig. 8B, appear little different from the bulk of the tissue. These distinctive areas are most probably compact bundles of connective tissue with an increased collagen content. For some soft tumors the recognition of increases in collagen content and connective tissue can be an important diagnostic criterion.

We are aware that connective tissue and collagen deposits can be accentuated in a light micrograph by applying a specialized stain such as Masson's trichrome; however, many of these stains require as much as 12 hours to apply. There are important situations, such as biopsies on patients in the operating room, where this time interval is prohibitive. For those cases it has been necessary to work with frozen sections that are stained with HE. A rapid evaluation of a frozen section with th acoustic microscope should assist in making a more accurate diagnosis.

As a second example of the acoustic appearance of diseased tissue, Fig. 9 shows section of lymph node afflicted with Hodgkin's disease. Several reasonably normal-looking lymphatic nodules can be seen i this image as circular areas of high acous tic transmission. The individual cells of these nodules are not well defined. In contrast, a highly attenuating matrix of mate rial can be seen between the nodules. Ev dently these are strands of collagenous material which have formed in response to the disease. Since these strands have a vastl greater attenuation than the normal tissue the acoustic microscope can easily show minute amounts of this material. In some instances this might prove useful as a diag nostic technique.

In the final illustration, Fig. 10, we present a comparison between the acoustic images of cancer cells within a lymph nod and the lymph node discussed in connection with Fig. 9. The neoplastic cells originated with a breast tumor which metastasized and spread to the lymph nod This comparison is made to point out a interesting difference in the acoustic response of the two cell types. As mentioned before, the cells of the lymph nodule show in Fig. 10B are not well distinguished. Tt
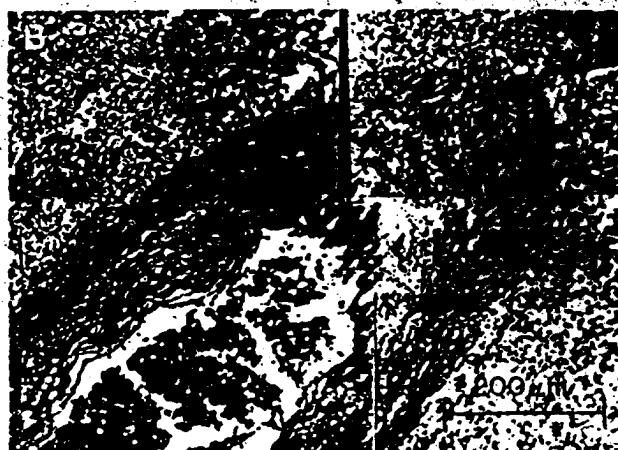
Fig. 6. Acoustic images (600 Mhz) of (A) a section of human oviduct; (B) human spleen, showing a small artery; and (C) normal human breast tissue.

200μm

156
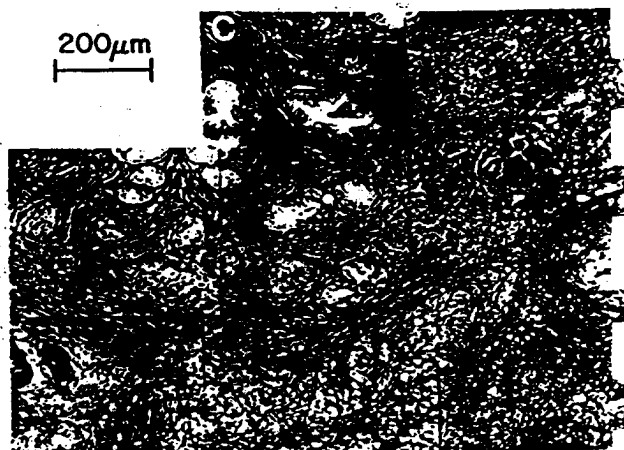
cancerous cells seen in Fig. 10A can, however, be readily distinguished by the clearly outlined nuclei. These nuclei are particularly conspicuous because points of large attenuation can be seen.

Future Developments and Applications

The transmission mode as illustrated in this article is but one of several modes of operation. In an alternative mode we can image those areas which exhibit uniform transmission by recording the acoustic phase shift through the specimen. We do this by comparing the phase shift of the signal through the instrument with a reference waveform of fixed phase as derived directly from the signal generator. Further than this we have studied the reflection mode as a method for viewing the minute detail of integrated circuits (9). This can be readily adapted to biological specimens mounted on a glass microscope slide. The large discontinuity of acoustic impedance at the glass interface will provide the reflected signal for this method of examination. There are still other possibilities based on the interference patterns of multiple beams—each displaced in frequency—passing simultaneously through the instrument.

The resolution can be further improved by going to higher frequencies in liquid cells of smaller dimensions and by using fluids such as liquid nitrogen and argon. These cooled liquids have an attenuation per unit length that is lower than that of water and a sound velocity that is nearly half that of water. Therefore, we will in principle realize a wavelength that is one-third that of what we are now using and the resolution should be improved by a corresponding factor.

In another study (10) we have found that certain biological specimens have a large nonlinear response to acoustic waves. This response is rather easily measured by tuning the output circuit to the second harmonic of the input signal and recording the scanned image at this double frequency. This information is distinct from that which we obtain in the linear imaging process and it provides a method for further increasing the informational content of the images of a given specimen.
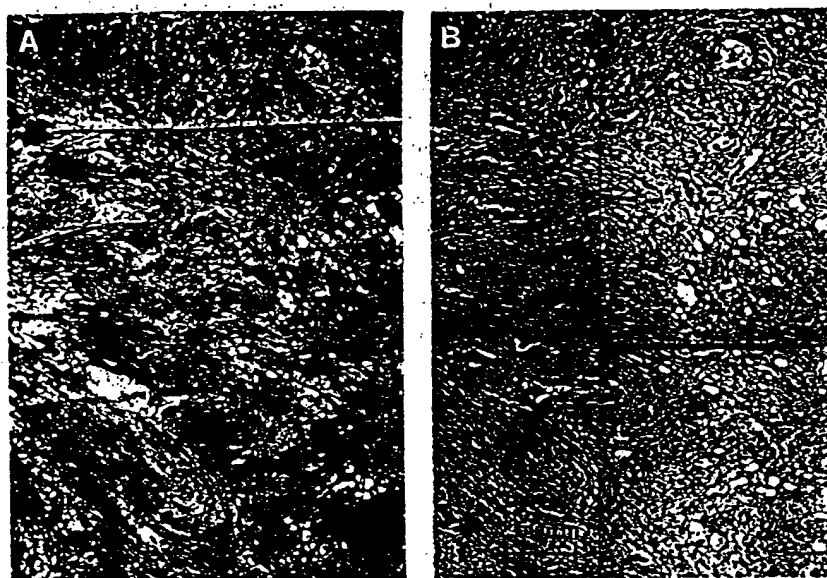
Fig. 7. (A) Acoustic image (600 Mhz) of a malignant tumor of the human breast, shown under low magnification. (B) Comparative optical micrograph (stain: hematoxylin-eosin).
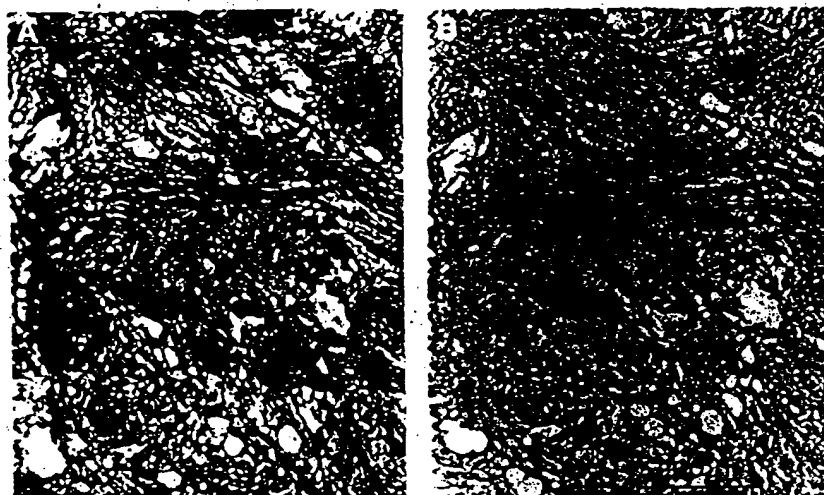
Fig. 8. (A) Acoustic image of a small region of the tumor shown in Fig. 7. (B) Optical image of the same area.
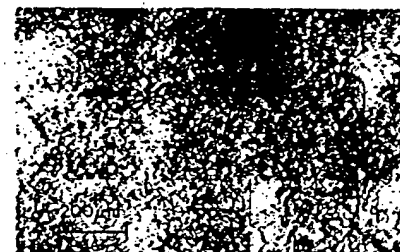
Fig. 9. Acoustic micrograph of a section of human lymph node afflicted with Hodgkin's disease.

Fig. 10. Comparison of the acoustic images (950 Mhz) of (A) a cancerous lymph node and (B) lymph nodule of the sample shown in Fig. 7. Both are shown under high magnification.

157

Conclusions

The technology of acoustic microscopy has now been advanced to the point where it can be used to record micrographs of cells and of cell complexes. The acoustic microscope responds to the elastic properties of the specimen and it therefore provides information which is distinct from that of an optical microscope. We believe that this instrument can be used in the fields of biological and medical research to obtain new information or to obtain some information more rapidly than is possible with present techniques.

The micrographs we have presented show that unstained biological material can be acoustically imaged with good contrast. Certain materials such as collagen and connective tissue have a particularly marked attenuation. As a result, the acoustic microscope can reveal information that is now only available through time-consuming staining techniques. Recently it has been shown that the acoustic microscope is compatible with frozen sectioning techniques and, with proper development, this may provide improved diagnostic techniques.

That the absorption of acoustic energy is proportional to the viscosity of the medium can be of particular importance in in-vestigating living systems (*11*). It might be possible to use the acoustic microscope for monitoring the viscosity within the cytoplasm of cells undergoing mitosis. The acoustic microscope might also be useful for monitoring the various rheological states of the contractile systems that are responsible for movement in several cellular systems. An increased understanding of the rheology of protoplasm or of cell division is a goal worthy of a large and extended effort.

With the work that has been done to date it is not possible to pinpoint the possible areas of application with great accuracy, but we do claim to have demonstrated that this is a new method for viewing the microscopic world. With this in mind it is important to examine carefully the features and limitations of this new tool and to search for areas where it can be used to extend our knowledge.

References and Notes

1. A. Korpel, in *Ultrasonic Imaging and Holography*, G. W. Stroke, W. E. Kock, Y. Kikuchi, J. Tsojivchi, Eds. (Plenum, New York, 1974), pp. 345-362; L. W. Kessler, *J. Acoust. Soc. Am.* 55, 909 (1974); R. K. Mueller (*Inst. Electr. Electron. Eng.*) *Proc.* 59, 1319 (1971). These reviews cover the state of the art of acoustic microscopy up to the spring of 1973. The series entitled *Acoustical Holography* (Plenum, New York, 1970-1974), vols. 1 to 5, contains several articles on the various ideas that have been put forth for an acoustic instrument.
2. R. A. Lemons and C. F. Quate, *Appl. Phys. Lett.* 24, 165 (1974).
3. E. L. Carstensen and H. P. Schwan, *J. Acoust. Soc. Am.* 51, 305 (1959).
4. A clear demonstration of this effect for low frequency sound has been given by Anderson where he compares the attenuation in liquid blood with the absorption in an organized blood clot [R. E. Anderson, in *Acoustical Holography*, P. S. Green, Ed. (Plenum, New York, 1974), vol. 5, p. 508, fig. 3].
5. This has been pointed out by F. Dunn and W. J. Fry in *The Encyclopedia of Microscopy*, G. L. Clark, Ed. (Reinhold, New York, 1961), p. 544; and even earlier S. Sokolov recognized this potential for an acoustic imaging system [see G. Devey, *Radio-Electron. Eng.* (a translation) (February 1953), p. 8].
6. R. A. Lemons and C. F. Quate, *1973 Ultrasonic Symposium Proceedings*, IEEE Cat. 73CHO807-8 SU (January 1974), paper E-6, pp. 18-21.
7. E. M. Slayter, *Optical Methods in Biology* (Wiley Interscience, New York, 1970), p. 241.
8. W. Bloom and D. W. Fawcett, *A Textbook of Histology* (Saunders, Philadelphia, ed. 9, 1968).
9. R. A. Lemons and C. F. Quate, *Appl. Phys. Lett.* 25, 251 (1974).
10. R. Kompfner and R. A. Lemons, in preparation.
11. A. D. Keith and W. Snipes, *Science* 183, 666 (1974); R. D. Allen, *Acta Protozool.* 11, 75 (1972). These authors include a discussion of the various techniques that have been used in the past to measure the rheological properties of cells.
12. The research was funded by a grant from the John A. Hartford Foundation, Inc., and we are grateful for that support. Drs. N. K. Wessells, L. Hayflick, R. Dorfman, R. Kempson, and M. Billingham gave us great encouragement and pointed to areas of possible utility. Drs. E. Farber and F. Ebaugh guided us in the work with cell complexes and tissue sections. The cell cultures of Fig. 6A were prepared in Dr. Hayflick's laboratory (Department of Medical Microbiology) and the tissue sections were prepared in Dr. R. Lawson's laboratory (Department of Pathology). Dr. W. Bond taught us how to fabricate and accurately assemble the components of the prototype. Dr. R. Kompfner originally suggested this project to us, and he and Dr. M. Chodorow were of great help in discussions concerning the physics of imaging.

MICROWAVE LABORATORY

M L

2 3 7 9

STANFORD UNIVERSITY

Acoustic microscopy

Electroacoustical transducers and acoustic lenses work at megahertz frequencies as "miniature sonar systems," forming high-resolution images that show properties not seen in optical micrographs.

Calvin F. Quate

The acoustic microscope is a new entry in the field of microscopic imaging. It comes after a delay of many years, but now appears to be well established. At last year's conference of the Royal Microscopical Society in London, for example, the booth traditionally reserved for new instruments was used to introduce commercial versions of acoustic instruments.[1] Visitors to the exhibit learned that one can focus acoustic waves in water to a diffraction-limited waist and use them as a probe for microscopic examination. They found in examining their own samples that they could observe new detail with a resolution comparable to that achieved by optical microscopes. (See figures 1 and 2.)

It is not easy to explain the delay in the introduction of acoustics into microscopy. Much of the technology has been around for some time. The idea surfaced in the Soviet Union as early as 1949, when Sergei Y. Sokolov pointed out[2] that the wavelength of 3-GHz sound in water is equal to the wavelength of green light. He suggested that an imaging system at that frequency might compete with optical instruments. There were some attempts[3] in this direction in 1960, but the serious work on acoustic microscopy did not start until the late 1960s. At the Zenith Laboratory in Chicago,

Acoustic micrographs. Right: Image of a bipolar transistor. The connecting lines are made of aluminum and are 2 microns wide. Opposite page: Acoustic image of a myxobacterium, which is a member of a social spore-forming class of bacteria. The diameter of the bacterium is 0.5 microns. (Bipolar transistor courtesy of TRW Inc.)   Figure 1
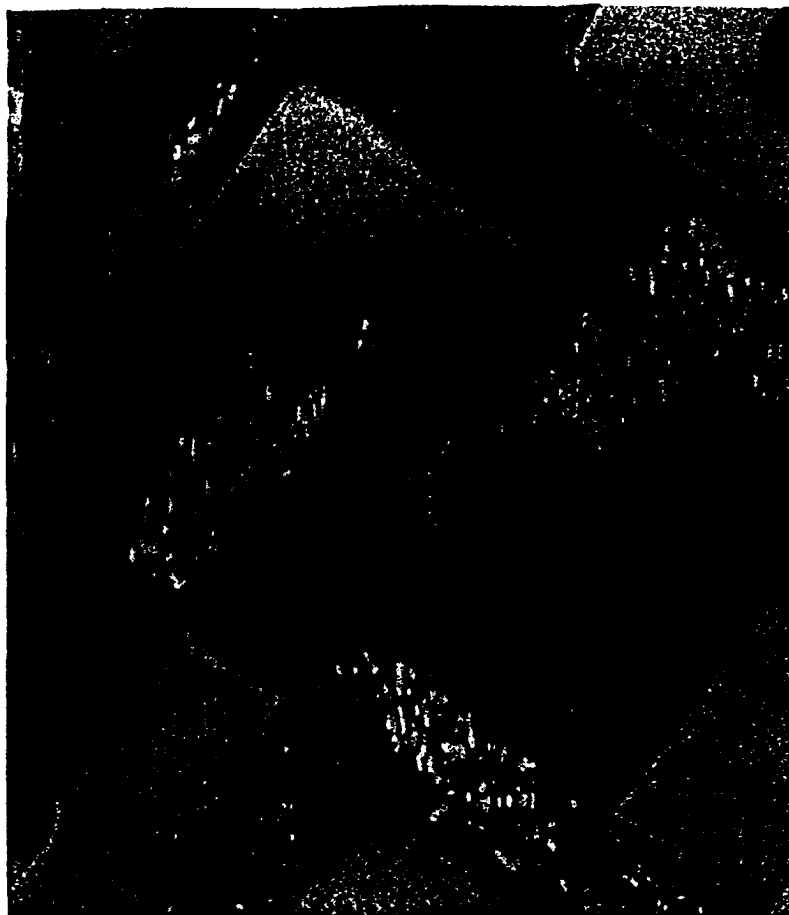
Larry Kessler and Adrian Korpel developed an acoustic system in which a scanned laser beam reads the image from a plastic membrane deformed by the acoustic signal. At Stanford, Bertram Auld and his students demonstrated and analyzed several systems of microscopic acoustic imaging. These workers and others in that early period established many of the basic principles of acoustic microscopy, and their work influenced all that followed.

In this article I discuss the physics of acoustic microscopy and consider the technique's applications and potential for further development. While I draw many of my examples from our research program at Stanford, excellent work on acoustic microscopy is under way all over the world. At Tokohu University in Sendai, Japan, a group under the leadership of Noriyoshi Chubachi has shown how to get quantitative information from the character of acoustic pulses reflected from samples. At the Université des Sciences et Techniques du Languedoc, in Montpellier, France, Jacques Attal has demonstrated how one can get information from the interiors of solids. At University College London, Eric Ash's acoustic imaging group has made effective use of the phase information in acoustic reflections to obtain contrast and to do internal imaging. And at the General

Electric Company in Schenectady, New York, Robert Gilmore has pioneered the use of low frequencies for scanning in depth over large surfaces. Scientists elsewhere have built instruments to demonstrate other characteristics of acoustic imaging; a special issue[4] of the *IEEE Transactions on Sonics and Ultrasonics* gives a more complete description of their work.

The advent of acoustic microscopy is only part of the renaissance that is taking place in the field of imaging. Other new instruments are using ion beams,[5] tunneling electrons[6] and x rays to study the microscopic world (see the article in this issue on x-ray microscopy, page 22, and the article on x-ray optics, April 1984, page 44). The table on page 36 compares the resolving powers of the various instruments. The new probes challenge the dominant position of optical waves and electron beams, but it is unlikely that the conventional methods will be displaced, as each viewing instrument extracts different information from the specimen. The ion microscope is powerful because it can analyze the masses of atomic species that come from the specimen during sputtering. The tunneling microscope allows us to study surfaces on an atomic scale, and is bound to have a large impact in surface science. The scanning x-ray

microscope has the potential for far greater resolution, and this will be realized as aberrations in x-ray lenses are corrected.

The acoustic system

The acoustic microscope differs from optical instruments in several significant ways. The change from optical to acoustic energy allows us to use frequencies in the microwave region. We must use a different kind of lens for focusing. We can not use photographic film, but must rely on piezoelectric transducers and electronic detectors. In combination these differences give us a new instrument, an instrument that extends our capability for observation beyond the limit set by optical systems.

We can operate the acoustic instrument at 3 MHz in water, where the wavelength is 500 microns, and at 8 GHz in helium, where the wavelength is 300 Å. This wavelength range, exceeding four orders of magnitude, is equivalent to operating an optical system from the microwave region, through the infrared, through the visible and beyond the vacuum ultraviolet.

Calvin Quate is professor in the departments of applied physics and electrical engineering at Stanford University. He is a senior research fellow at the Xerox Palo Alto Research Center.
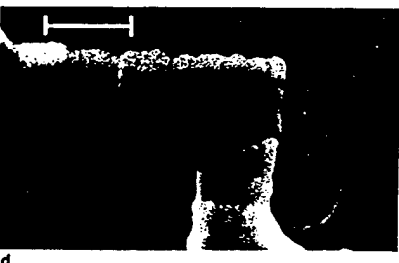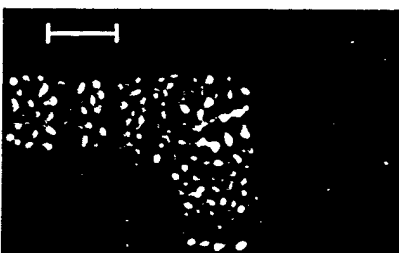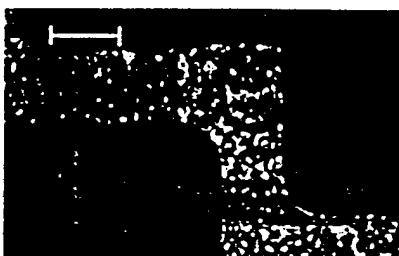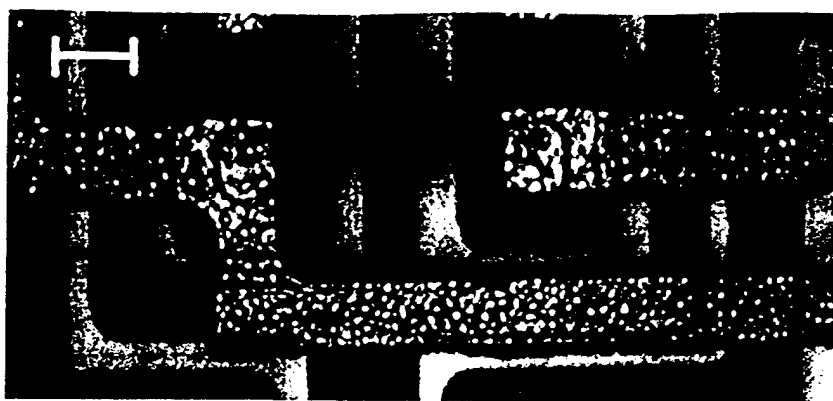
Furthermore, the piezoelectric detectors permit us to exploit coherent radiation, where information is available from both the amplitude and phase of the reflected signals. The piezoelectric films, typically made from zinc oxide, are efficient. As generators these films can convert 50% of the electromagnetic energy into sound. As detectors they can convert 50% of the sound into electromagnetic energy.

The lens is the heart of the acoustic microscope. It is deceptively simple, but it makes all the difference in the imaging work. The lens can have various forms, such as the one shown in figure 3a, which is a sketch of a lens that we have used at high frequencies at Stanford. The lens is a sapphire crystal with a spherical cavity; it is immersed in a fluid that makes contact with the specimen. When a plane wave of sound is launched into this crystal by a piezoelectric film, it is converted to a spherical wave at the solid–liquid interface. Focusing, limited only by diffraction, is easily obtained with this single-surface lens.

There are no aberrations because the velocity of sound in the fluid is small compared to the velocity of sound in the solid, and therefore there is a large angle of refraction at the interface. The direction of a ray crossing the interface into the liquid lies close to the normal. In this situation the focal length is nearly equal to the radius of curvature of the surface and the spherical aberrations are reduced to negligible values. There is, of course, a penalty. The change in velocity is accompanied by a strong discontinuity in the acoustic impedance, a problem that must be overcome to transfer energy efficiently into the liquid. We have found that a layer of carbon makes an excellent antireflection coating.

An acoustic-microscope system resembles a miniature radar or sonar system. The acoustic beam, after reflection from the sample, passes back through the lens and into the piezoelectric transducer that generated it. The transducer, now working as a receiver, detects the signal, whose amplitude and phase carry information about the sample. The acoustic lens moves in a raster pattern over the specimen, examining each point sequentially. The system requires several seconds to cover a single frame containing 512×512 pixels, each of which is typically 0.5 microns in size. The returning pulses modulate the intensity of the beam in a cathode-ray-tube monitor. The display system is similar to that in a scanning electron microscope.

The interface wave

What is it that we "see" in acoustic images that we do not "see" in optical images? Here are a few representative examples: We can see fibers inside various composite materials; we can see features of biological cells that are difficult to image in the optical microscope; and we can see contrast that tells us about the structure and the adhesive properties of metal films and other coatings. Abrupt changes in elastic properties show up in acoustic micrographs where there are no corresponding changes in the optical index of refraction. For example, with polycrystalline samples of anisotropic metals, the optical image of a polished surface is smooth and featureless. In the acoustic images, individual grains stand out, showing their random orientations.

There is contrast between individual grains and their neighbors because there is a wave that travels along the interface between the sample and the overlying liquid, and this "interface wave" is sensitive to small variations in the elasticity of the exposed surface. The interface wave is responsible for highlighting the elastic properties, or rather the changes in these properties, in acoustic images. Surface waves of this type, common in acoustic systems, have no counterpart in electromagnetic systems. When we first encountered them in our work we were unprepared for the remarkable changes that they produce in images when the sample is moved away from the focal plane, toward the lens.

The effect of this defocusing is most easily demonstrated by measuring the output voltage V from the transducers while varying the lens-to-sample spacing z. In this measurement, one suppresses the lateral translation of the sample so that the output is a single curve of voltage versus spacing. For a classical focusing system with coherent

Scanning instruments

Technique	Resolution
Optical microscopy	2000 Å
X-ray microscopy	500–700 Å
Ion microscopy	400 Å
Acoustic microscopy	200 Å
Tunneling microscopy	2 Å

a

Lens configuration (a) and transducer output (b). The lens is a spherical cavity in sapphire. The plots show the measured transducer output power as a function of the distance between the lens and an aluminum specimen in water at about 40° C. The acoustic frequency is 1.8 GHz. The angle θ is the opening angle of the acoustic lens used to collect the data. Figure 3

b
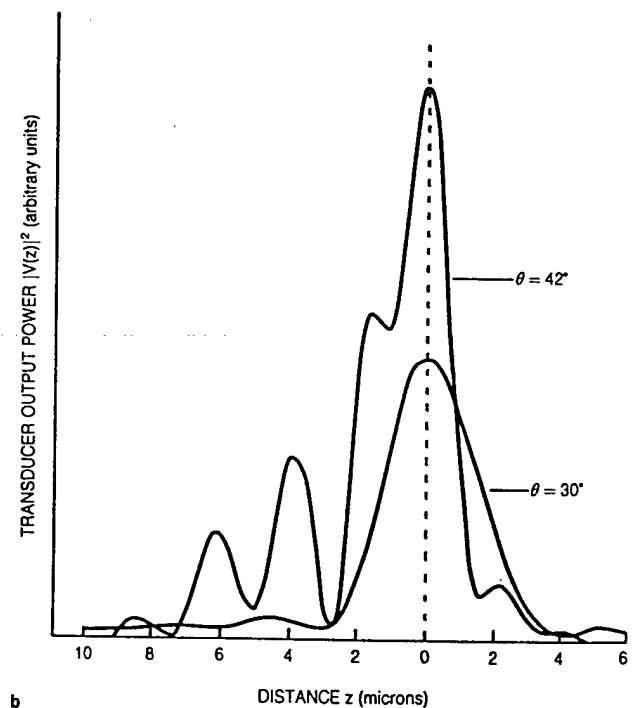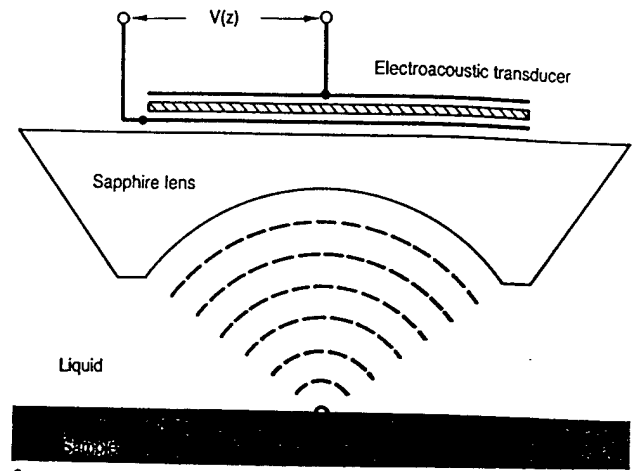
radiation and a phase-sensitive detector, the voltage curve has the form $\sin z/z$. In an acoustic system with a small-aperture lens, these surface waves are not excited and we do indeed measure such a curve. The experimental curves in figure 3b, recorded at Stanford by Larry Lam, demonstrate this point. The sample was aluminum with a polished surface, and the working fluid was water. For these materials, the critical angle for coupling to the interface wave is 33°. The opening angle for the lens used to collect the data represented by one of the curves is 30°, which is less than the critical angle. The interface wave is not excited and the classical $\sin z/z$ curve is the result. The other curve shows the response for a lens with an opening angle of 42°, which is large enough to encompass the critical angle. The form of this response curve is fundamentally different because the surface wave is now generated. The deep minima come from the interference between the ray at normal incidence and the ray at the critical angle.

We can clarify this interference still further by reference to figure 4. The transducer depicted at the top of the figure is a piezoelectric film that generates short pulses of acoustic energy. Each pulse is a plane wave of sound that propagates downward through the crystal into the liquid. As the wave passes through the surface of the lens, it is transformed into a spherical wave that converges toward a focal point. After reflecting from the sample, the acoustic pulse returns through the lens to the transducer, which converts it back to an electrical pulse that is the output signal.

This system has a very small field of view. This means that the transducer responds only to those rays that appear to be reflected from a small region very near the focal point of the lens. When the surface of the reflecting sample is placed near the focal plane, the entire beam is reflected and the transducer output is maximized. When the surface of the reflecting sample is displaced from the focal plane, it still

reflects much of the beam, but in such a way that it appears to come from a region far from the focal point. The transducer does not respond to those rays because they do not meet it at normal incidence. The transducer can only respond to a narrow central cone, because that cone always meets it at normal incidence. That cone, marked A in figure 4a, appears to originate in the focal plane regardless of the position of the reflecting sample. There is, however, a second mechanism by which reflected rays can appear to come from the focal plane. This involves the surface wave produced by the incoming radiation that meets the interface at the critical angle. The surface wave travels along the interface and continually reradiates energy back into the liquid at the same critical angle. The

ray marked B in figure 4a illustrates this. The direction of this ray makes it indistinguishable from a ray reflected from the focal plane.

The two components represented by the ray paths marked A and B in figure 4b contribute to the transducer output with an amplitude that is nearly independent of the position of the reflecting sample. There is a phase shift between the two components, and this sets up an interference pattern that changes when the lens-to-sample spacing is varied. The distance between minima in this interference pattern measures the velocity of the interface wave along the surface of the sample. This, in turn, tells us something about the elastic properties of the sample. The pattern of interference illustrates the variation in elastic constants over the

162

surface of the sample and provides the contrast that we record in the display.

History of the V(z) effect

The interference between the central cone and the surface wave produces remarkable images. Robert Wilson at Hughes Research Laboratories, working with Rolf Weglein, once sat on a high stool for hour upon hour and recorded acoustic images as he moved the sample toward the lens in discrete intervals.[7] That series was pivotal; the striking regularities were difficult to explain with any model that anyone was using. In our own lab, Abdullah Atalar, who first introduced the reflection instrument while he was a graduate student, tried in vain to tell us that there might be new information in his

out-of-focus images.[8] We were inclined to attribute the changes to aberrations in the lens. It was only after he recorded images of the grains in an alloy of nickel that we were convinced.

At that point we knew that the changes in contrast depend on the nature of the sample. We realized for the first time that we might use the V(z) effect to image the elastic properties of materials, and we began work in earnest. We found that one can clearly delineate the grain structure of metals and ceramics, and that one can study subsurface structures if they are within reach of the interface wave, which extends about one wavelength into the solid. This distance is 8 microns in silicon at 1 GHz, and 2 microns at 4 GHz.

We set about the task of correcting our model. In the process, we uncovered the work[9] of Mack Breazeale at the University of Tennessee, who had characterized steel samples by exploiting the fact that acoustic waves reflected from a liquid–solid interface are displaced laterally, an effect known as the Schoch displacement. Kumar Wickramasinghe, now at IBM, calculated[10] the transducer output voltage V as a function of the lens-to-sample spacing z—the V(z) curve—in a concise manner, providing additional insight into this important and distinctive source of contrast. Henry Bertoni of the Polytechnic Institute of New York published[11] a ray theory that also increases our understanding of the interactions that take place in the out-of-focus geometry.

The recorded images, which I will now consider, illustrate what can now be done with acoustic microscopy.

The images

The texture, or grain structure, of polycrystalline materials is the first and most dramatic difference that shows up in comparisons of acoustical and optical images. Inconel and titanium samples illustrate these differences. We see in figures 5a and 5b contrast changes from grain to grain. We don't yet fully understand the source of the contrast that highlights the boundaries. That remains as a problem for the commercial instruments to unravel. Figure 6 shows state-of-the-art images from commercial instruments. The detail and contrast in these images, of aluminum lines on silicon and of glass fibers embedded in epoxy, demonstrate the power of electronics in presenting images.

Integrated circuits, with their dominant role in today's technology, require better and better performance from imaging instruments. Acoustic imaging should play a role in selected areas such as the problem of achieving adhesion in the deposition of thin films. Photomasks for the fabrication of integrated circuits are commonly made of

163

Contrast enhancement. a: Three images of a polished inconel surface, showing the grain structure with increasing contrast as the sample is moved from the focal plane (z = 0) toward the lens. The operating frequency is 1.8 GHz. The rectangular area imaged is about 80 microns wide. b: Acoustic images of a polished surface of titanium. The area imaged is about 70 microns wide.

Figure 5

chrome on glass. The lack of adhesion of the chrome films is a serious defect that is difficult to identify with optical methods.

Heat dissipation in integrated-circuit chips is another area where acoustics will be important. Chips must be securely bonded to their heat sinks, and manufacturers require some method for monitoring this bond. The solder used for bonding often contains voids and fissures that reduce the heat conductivity to unacceptable levels. Acoustic instruments that operate at moderate frequencies with reduced resolution can now monitor problems of this kind.

The remarkable penetrating power of acoustic radiation is demonstrated in the work of Butrus Khuri-Yakub at Stanford. In one experiment, he used a lens in the form of a spherical shell at an instrument operating at a frequency of 3 MHz—a wavelength of 500 microns—to image a honeycomb structure used as the skin of aircraft. The structure consists of two sheets of aluminum held apart by a honeycomb grid. He can either focus on the surface, which gives a bland, uninteresting image, or on the internal structure.

The internal image with the hexagonal grid illustrates the unique feature of acoustic imaging because the inside of the sample is inaccessible to optical radiation.

Polymers such as those used to make photoresist patterns represent another area of opportunity for acoustic imaging. Photoresist material can be troublesome if for some reason it contains defects such as bubbles. Lack of contrast in optical images hides these defects, but in acoustic images they appear as points of strong reflection.

In polymer materials, both the velocity of sound and the acoustic impedance have values quite close to those of water, which is the working fluid between the lens and the sample. We do not observe the $V(z)$ effect in these materials because the critical angle is too large. However, the reflection at the interface is small, and this allows the acoustic beam to propagate into the interior. Inclusions and other defects are prominent features in the images. The inclusions that appear deep inside the film have impedances that are much less than that of the surrounding polymer. It is easy to identify small pockets of gel, which are a common

defect in these films.

Distributing high-strength fibers throughout polymers or other materials strengthens those materials and modifies some of their other physical properties. The distribution of fibers is difficult to monitor with optical radiation because the host material is often opaque. In an acoustic field, the fibers produce strong reflections because their acoustic impedance is higher than that of the host material. Figure 7 shows an example of an image in which glass fibers are the only apparent feature. It is in some sense a "binary" image, with two values, black and white.

Finally, we return to the high-magnification acoustic image shown in figure 2. This image, taken with an acoustic microscope operating at 4.5 GHz, rivals optical images in quality and resolution. It represents the high point for acoustic imaging in water.

Biology. Biological cells represent a class of objects that are entirely different from the solid objects discussed above. Optical absorption is not large in these cells, so optical imaging requires special techniques such as differential interference contrast. Acoustic

164
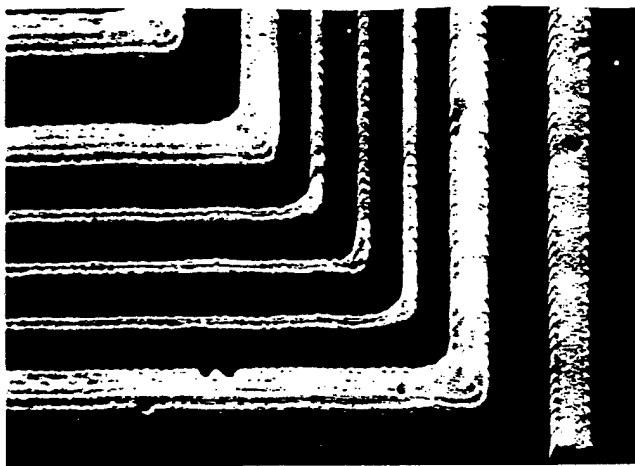
Highlighting in images produced by a commercial acoustic microscope. Top: Aluminum lines on silicon. The aluminum line at the right is about 6 microns wide. Bottom: Glass fibers embedded in epoxy. (Courtesy of E. Leitz Inc.) **Figure 6**

micrographs serve to increase the contrast in biological samples and to bring out new details. Sound waves travel through the entire cell with ease because cellular material is largely water. Nevertheless, there are differences in the acoustic properties of the cellular material, and these represent sources of contrast in the images.

With cells grown on glass substrates, we encounter two reflections, one at the liquid–cell interface at the top of the cell, and the other at the cell–substrate interface beneath the cell. These reflections interfere constructively or destructively as the cell thickness varies. Contours of constant cell thickness appear as dark rings. Changes in the ring structure allow us to study the dynamic properties of the cell. Such features as adhesion to the substrate, movement of particles, and construction and deconstruction of filaments through polymerization and depolymerization are easily observed.[13] New information on cellular movement should be forthcoming from acoustic microscopy.

The changes in elastic properties that take place in cells and tissues should also show up in acoustic micrographs, although the evidence to support this hunch is not yet available.[14]

Imaging such changes would be useful for studying pathological specimens.

Resolving power

Each instrument used for microscopic imaging has features that make it unique and interesting. However, it is an instrument's resolving power, as listed in the table on page 36, that determines where it will be used. The useful operating range for optics is limited by the large absorption in the ultraviolet region. As a rule of thumb, the resolution of the conventional optical microscope is 0.4 microns in air, 0.3 microns in water and 0.2 microns in oil. These numbers set the goals for competing instruments.

In a similar way, it is acoustic absorption in liquids that limits the upper operating frequencies of the acoustic microscope. Sound absorption per unit distance increases as the square of the frequency. If we use wavelengths shorter than those of microwave frequencies, we encounter severe restraints. The electronics dictate that to preserve an acceptable signal-to-noise ratio we must maintain the total attenuation along the signal path at a constant level. If we double the frequency, for example, we must reduce the signal path to one-quarter its pre-

vious length. The lenses must be fabricated with smaller and smaller radii.

We have been able to achieve a signal-to-noise ratio of 30 dB at 4.5 GHz with a lens 16 microns in radius. This gives a resolution near 0.2 microns (see figure 2), the limit of the optical microscope.

If we seek to improve the resolution to observe objects below that size, we must use a liquid with an acoustic velocity less than that of water. There are such liquids at room temperature, but they are not useful because their attenuation is too large. Cryoliquids emerge as the most attractive candidates.

Liquid nitrogen is a convenient and readily available cryofluid. The absorption of sound in this liquid is similar to that in water, while the velocity of sound is one-half that in water. At a given frequency, we can improve the resolution by a factor of two by transferring the instrument from water to liquid nitrogen.

While working with liquid nitrogen, Daniel Rugar, now at IBM San Jose, discovered[15] a new phenomenon that is important for imaging. He found that one can improve the resolution by moving to the nonlinear regime, where the amplitude of the sound wave is high enough to generate harmonics. While the evidence is clear—images with improved resolution—the theoretical explanation is less clear. The improvement may come from second harmonics, because those components produce a sharper image than the fundamental frequency. However, there seems to be a dilemma. How can one observe the second harmonic content by monitoring the output with a detector that operates at the fundamental frequency? The theory that has been published shows that this is indeed reasonable and that the behavior of nonlinear waves in spherically focused beams is quite different from what we expected, but we don't yet have a simple description that I can present here.

Liquid helium. The ultimate resolution with acoustics will come from an instrument immersed in liquid helium. In this liquid the sound velocity is one-sixth of that in water and the absorption of sound decreases with the fourth power of the temperature when the

Glass-reinforced polymer. a: Optical-microscope image. b: Acoustic-microscope image produced with 0.9-GHz acoustic radiation. The width of the area imaged is 400 microns. (Courtesy of Mark Hoppe, E. Leitz Inc.) Figure 7
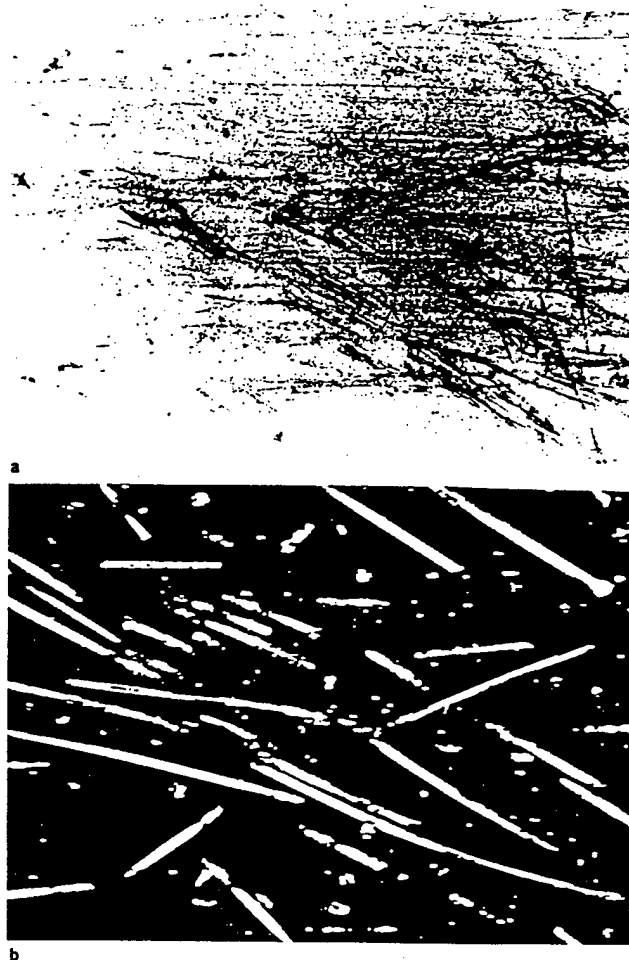
liquid is cooled well below the lambda point. When we operate at 100 mK, for example, the absorption is reduced to the point where it is no longer a limitation on the operating frequency.[16]

The behavior of this liquid baffled us for a time because we could not find the signal reflected from the object. At first we thought it was either a problem of misalignment or a lack of input signal strength. Hence, that is what we worked on—new methods of alignment, efficient transducers, antireflection coatings. We injected higher and higher levels of sound power into the liquid helium in an effort to increase the strength of the reflected signal, but our efforts were in vain, for they never produced a signal.

It took us a very long time to get all of this straight. We now know that the culprit is the nonlinear behavior of the medium. This behavior is emphasized when the intensity is increased by focusing the beam, but it is worse than that. We have learned that increasing the input level of a focused beam above a certain intensity can actually reduce the energy at the fundamental frequency. The energy is effectively lost when it is transferred to the higher harmonic frequency because the piezoelectric detector is unable to respond to the shifted frequencies.

The levels that we commonly use in present-day instruments are 30 dB below the levels that we used in the dark days when there were no signals. We have been forced to use every known strategy to capture and manipulate the weak signals. We have learned from the radioastronomers how to build GaAs FET amplifiers cooled to 4 K to achieve low effective noise temperatures. We have learned from the radar community how to employ chirped signals to lower the peak power levels of input without sacrificing the total energy.

With all of these improvements, our system appears to be working rather well.[17] We produce images at two frequencies—4.2 GHz and 8.0 GHz. At 4.2 GHz the wavelength is 570 Å and the resolution is near 400 Å. The image of the integrated circuit in figure 1 is a typical example. The sample is a bipolar transistor with a line width of 2

microns. The depth of focus for the acoustic beam is approximately 1500 Å, less than the height of the aluminum lines. We are able to record several images—each focused on a different level. We then assign each image a different color and assemble them into a composite with a computer. Thus color in the figure represents height in the object.

When we operate at 8 GHz, where the wavelength is 310 Å, the resolution appears to be better than 200 Å. The most interesting result is the image of a specimen of myxobacteria furnished to us by Stanford biochemist Dale Kaiser. The contrast in the composite image, shown in figure 1, is much greater than that found in a transmission-electron-microscope image.

This result suggests that the helium acoustic microscope, with its narrow depth of focus that makes the surface topology show up with high contrast, will one day play a role in the study of microscopic objects.

. . .

References

1. Proc. Roy. Microscopical Soc. 19, Part 4, MICRO 84 Supplement (July 1984).
2. S. Y. Sokolov, Dokl. Akad. Nauk. SSSR 64, 333 (1949).
3. G. L. Clark, ed., The Encyclopedia of Microscopy, Reinhold, New York (1962), p. 544.
4. IEEE Trans. Sonics Ultrason. SU-32, No. 2 (1985).
5. A. L. Robinson, Science 225, 1137 (1984).
6. G. Binnig, H. Rohrer, Ch. Gerber, E. Weibel, Phys. Rev. Lett. 50, 120 (1983).
7. R. G. Wilson, R. D. Weglein, J. Appl. Phys. 55, 3261 (1984).
8. A. Atalar, J. Appl. Phys. 50, 8237 (1979).
9. M. A. Breazeale, L. Adler, G. W. Scott, J. Appl. Phys. 48, 530 (1977).
10. H. K. Wickramasinghe, Electron. Lett. 14, 305 (1978).
11. W. Parmon, H. L. Bertoni, Electron Lett. 15, 684 (1979).
12. B. Hadimioglu, C. F. Quate, Appl. Phys. Lett. 43, 1006 (1983).
13. J. Hildebrand, D. Rugar, J. Microsc. 134, 245 (1984).
14. R. A. Lemons, C. F. Quate, Science 188, 905 (1975).
15. D. Rugar, J. Appl. Phys. 56, 1338 (1984).
16. J. Heiserman, Physica (Utrecht) 109 & 110B, 1978 (1982).
17. J. Foster, Physica (Utrecht) 126B, 199 (1984).

ACOUSTIC IMAGING
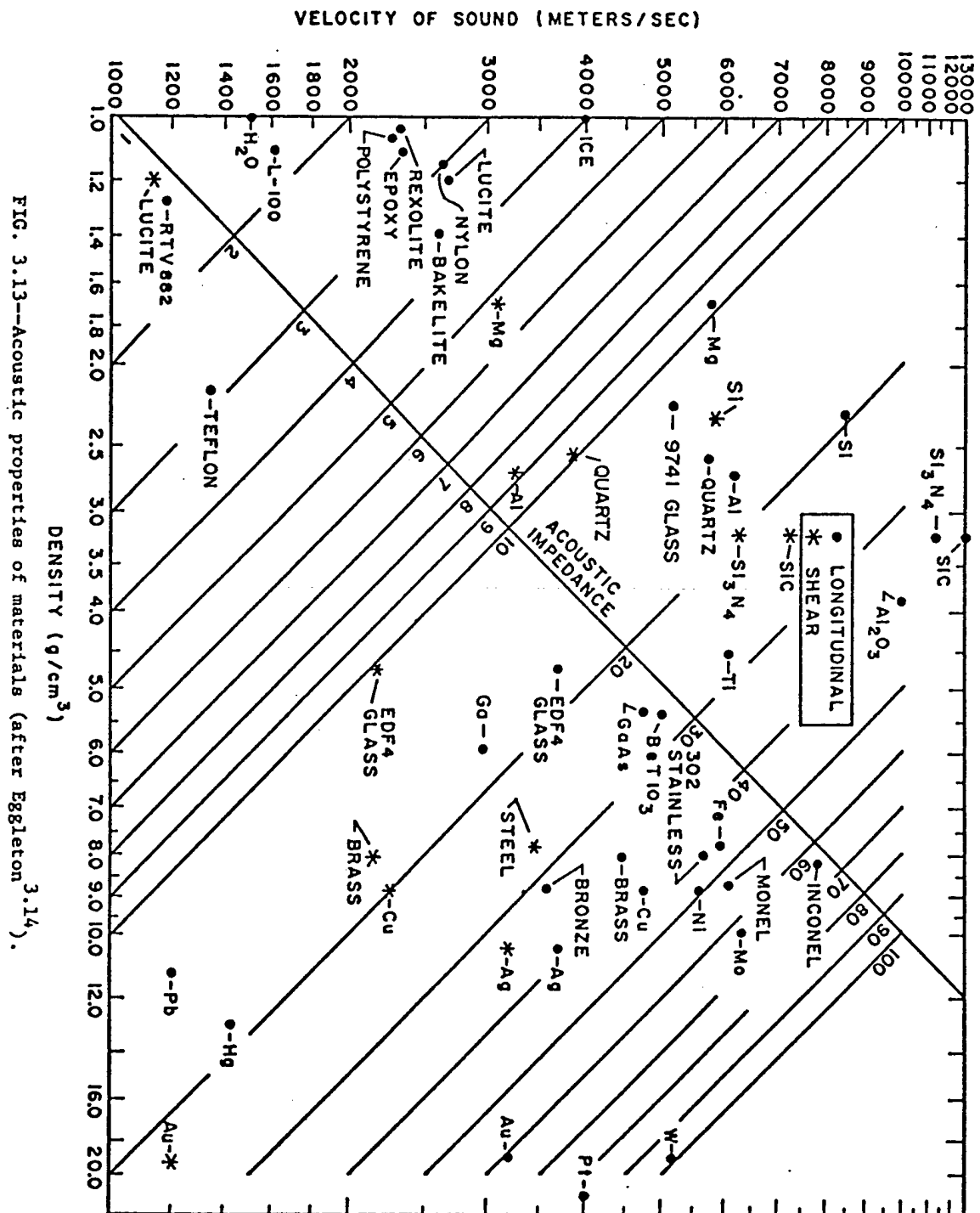
FIG. 3.13—Acoustic properties of materials (after Eggleton[3.14]).

168
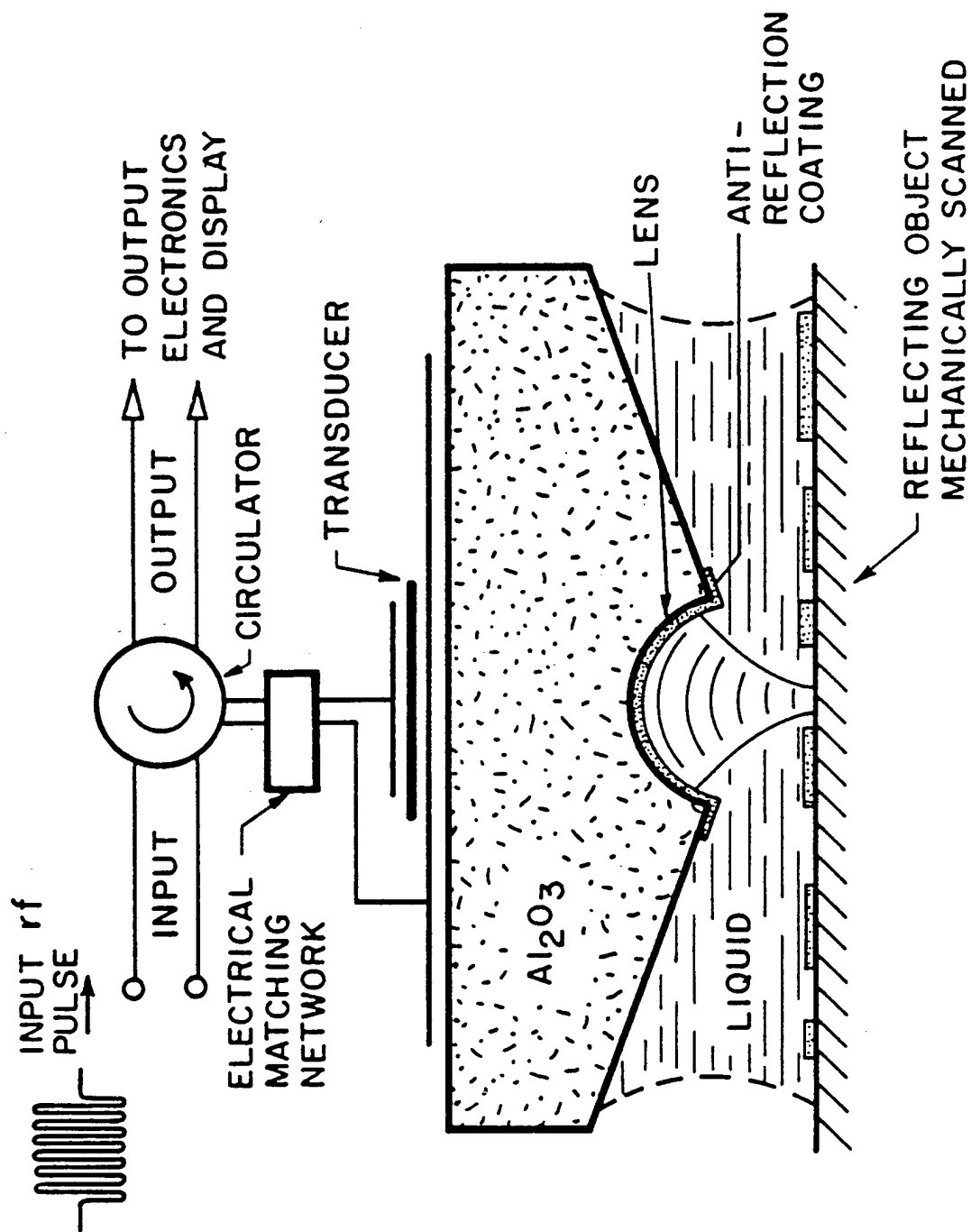
FIGURE 1.1--Configuration of the reflection acoustic microscope.

169

AIR $\big\{$ FLINT GLASS
(rad = .85mm)
$c_2/c_1 = 0.667$

sin θ = 0.4

(10μ DIVISIONS)

sin θ = 0.4

(10μ DIVISIONS)

SAPPHIRE $\big\{$ WATER
(rad = .85mm)
$c_2/c_1 = 0.135$
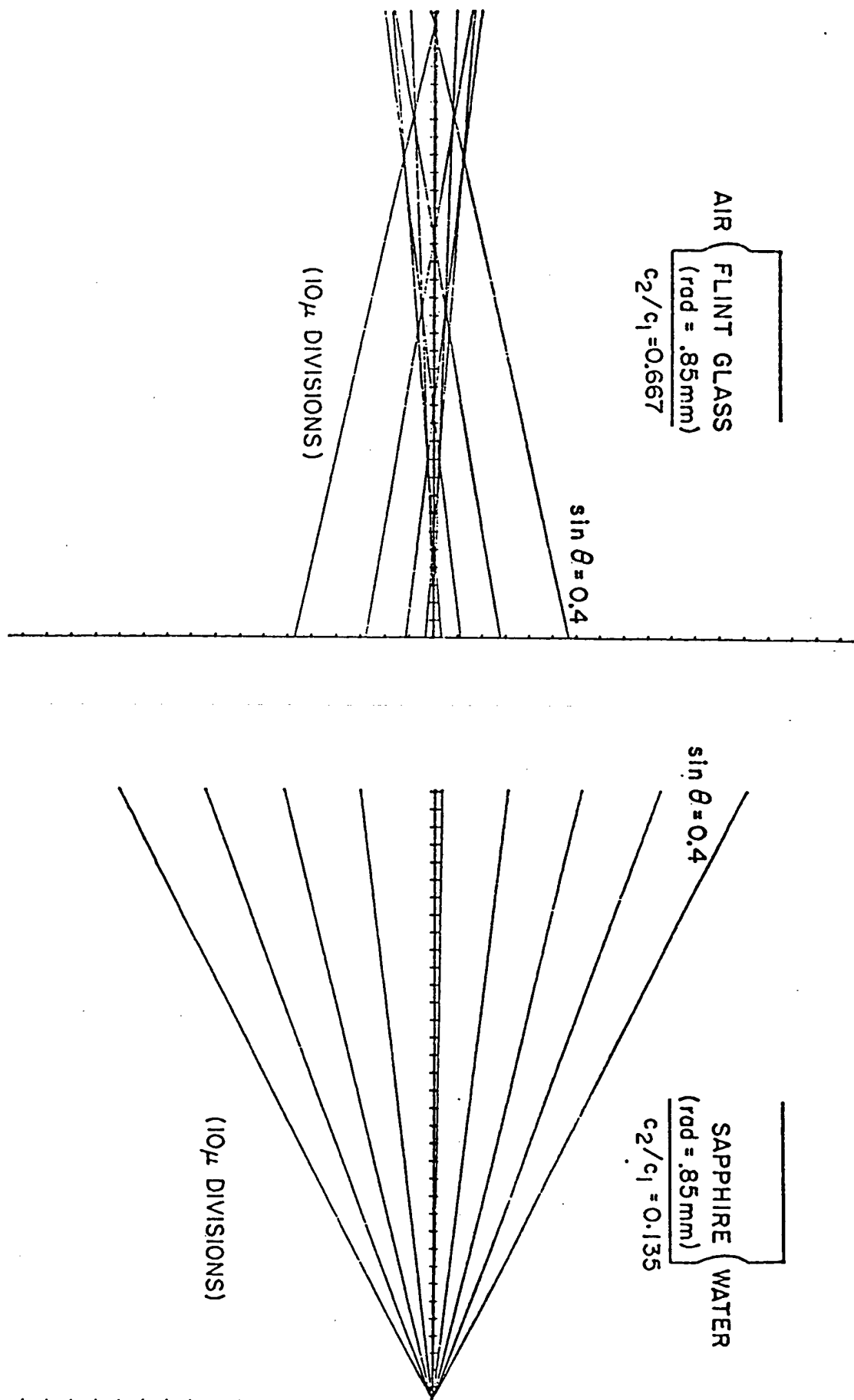
FIG. 3-3--Ray tracing comparison of the performance of a single surface lens in a light optical system (left) and an acoustic system (right). The paraxial focus lies at the ordinate.

170

OPTICAL, UNETCHED   OPTICAL, ETCHED   OPTICAL, ETCHED   ACOUSTIC

Z = -0.5   20 μm
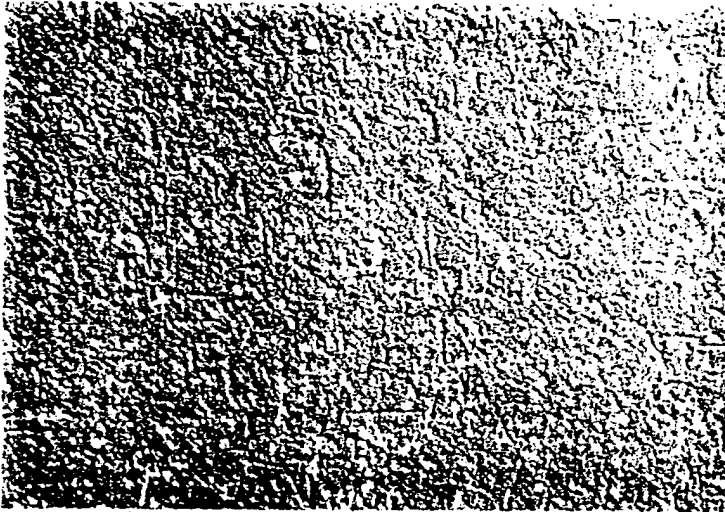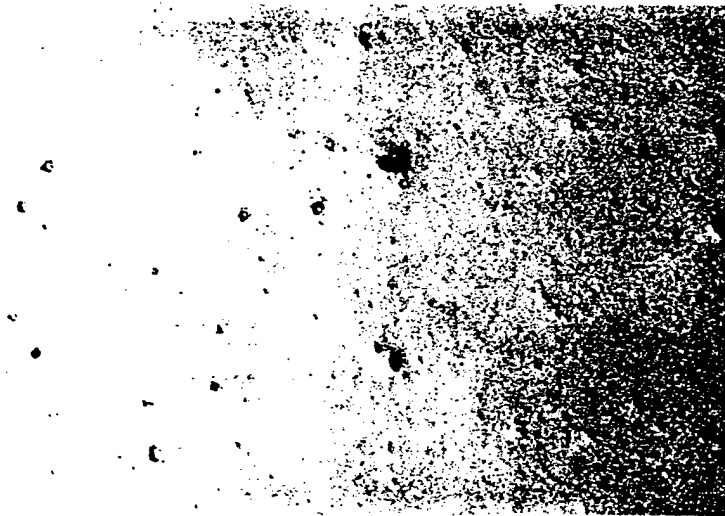
80% NICKEL, 0.6% Y_2O_3, 19% CHROM.

Fig. 1B.2   Optical and acoustic images of a nickel alloy sample.

171

OPTICAL

ACOUSTIC z = 0

ACOUSTIC z = -1

Fig. 1B.3   Optical and acoustic images of a ferrite sample.

Z = 0   20μm

Z = -1

Z = -1.5
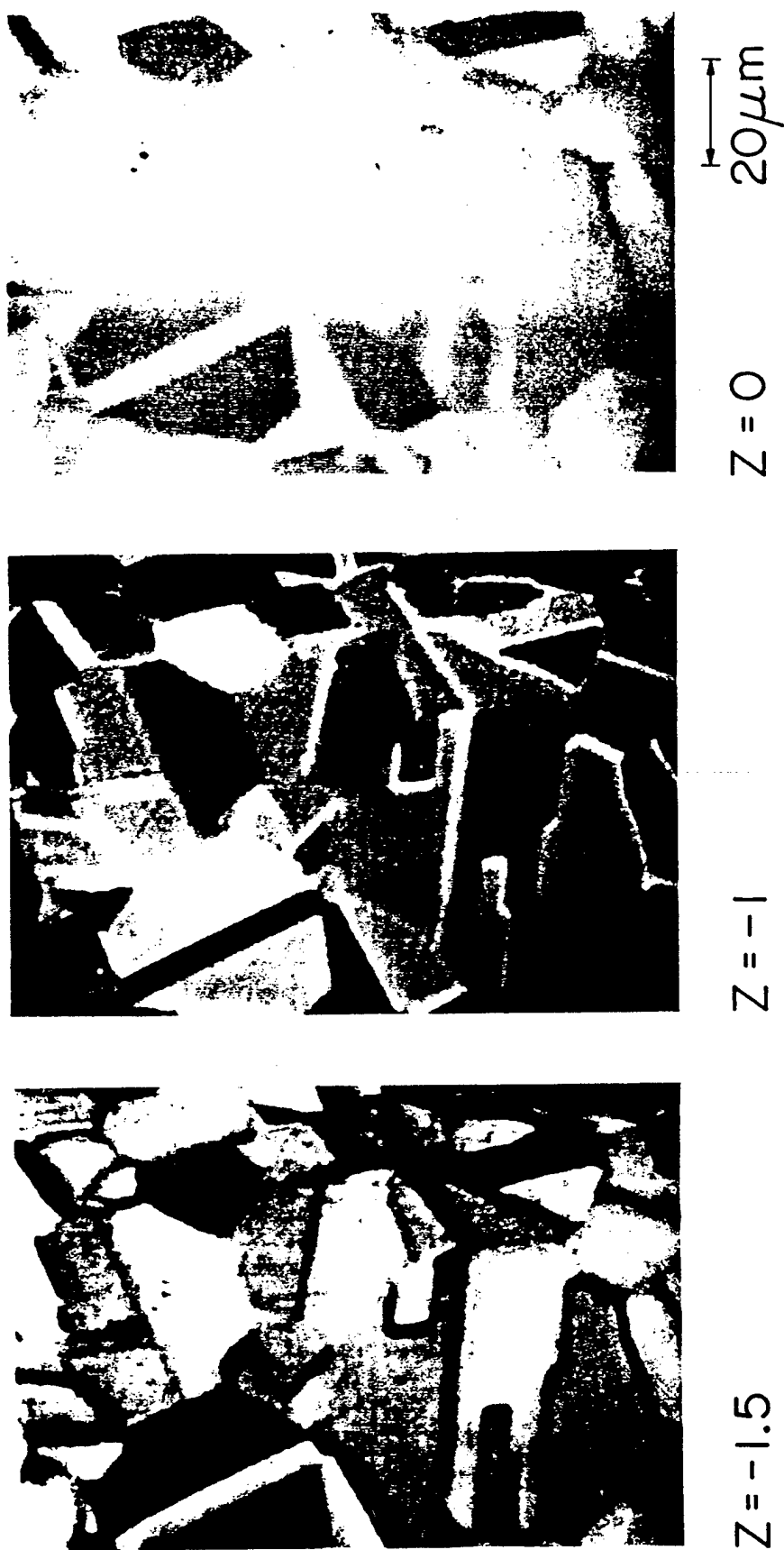
INCONEL

Fig. 3A.4   Three x-y scans of an inconel sample. The reference particle is the pentagon shaped dark particle in the upper right hand corner of the z = 0 image.

173
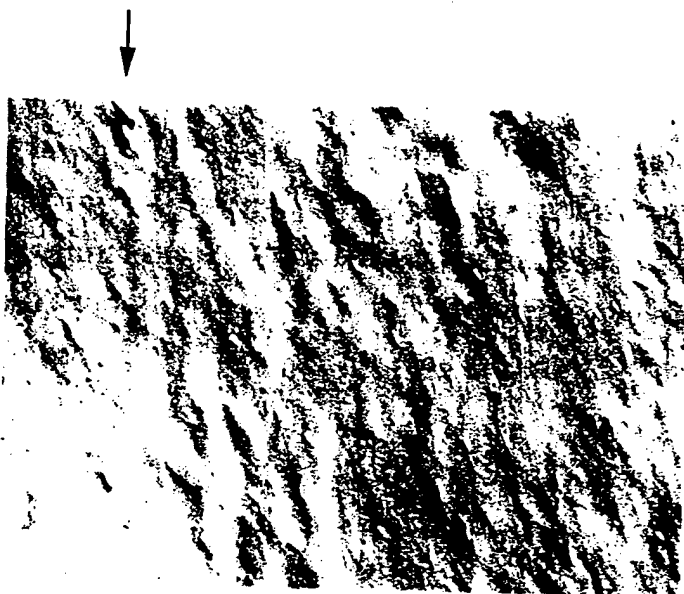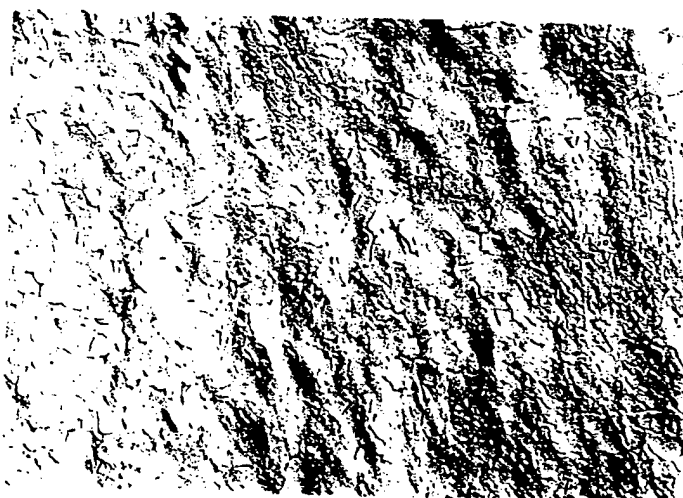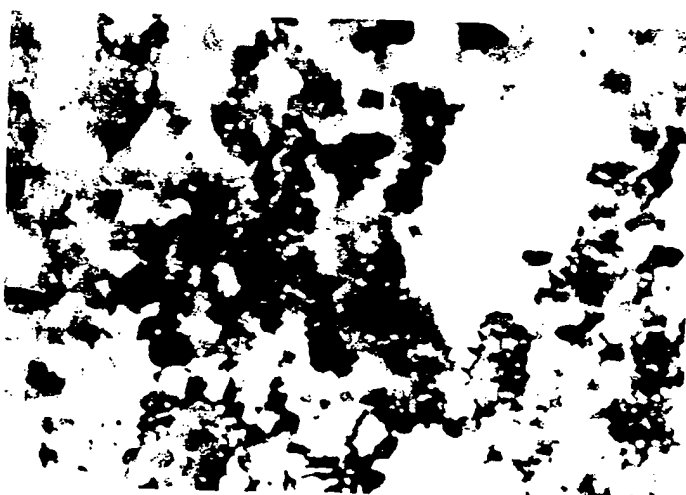
OPTICAL, UNETCHED   OPTICAL, ETCHED   ACOUSTIC, UNETCHED Z = - 0.5

100 μm

TITANIUM DIFFUSION BOND

Fig. 3A.5   Optical and acoustic images of the same region near a diffusion bond on a titanium alloy sample. An arrow at left indicates the position of the diffusion bond between

174

INCONEL ALLOY

(a) OPTICAL, POLISHED
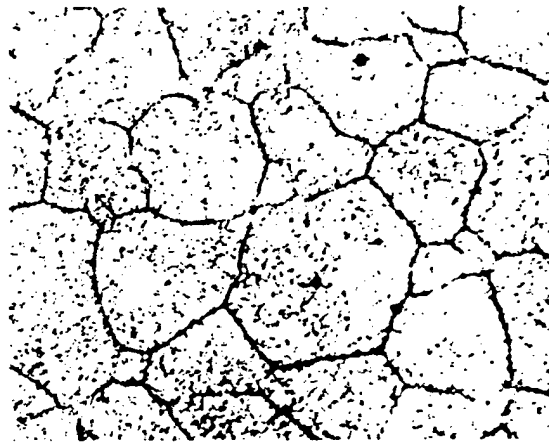
(b) OPTICAL, ETCHED

(c) ACOUSTIC $Z = -0.5\mu$
2.7 GHz

(d) ACOUSTIC $Z = -1.0\mu$
2.7 GHz

FIGURE 3-1

(a) OPTICAL ×620
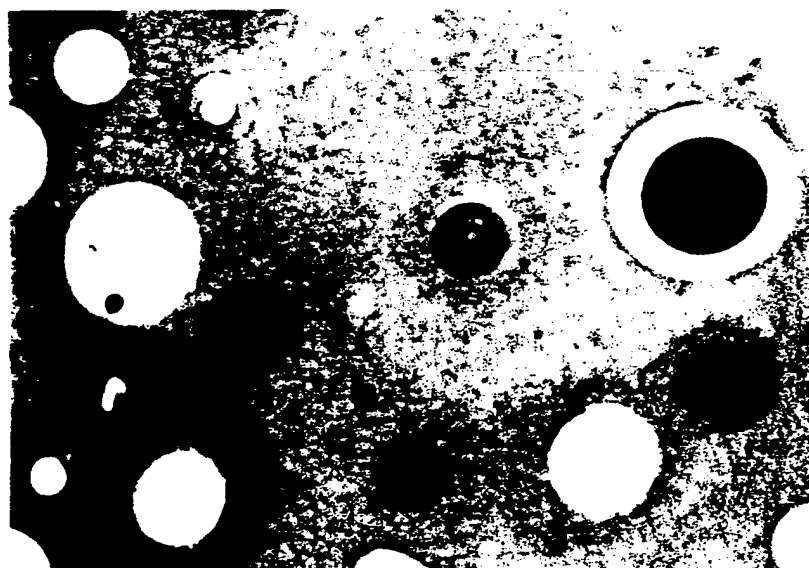
(b) ACOUSTIC (2600 MHz)

FIGURE 3-5

POLYETHYLENE TEREPHTHALATE
WITH IMBEDDED GLASS BEADS
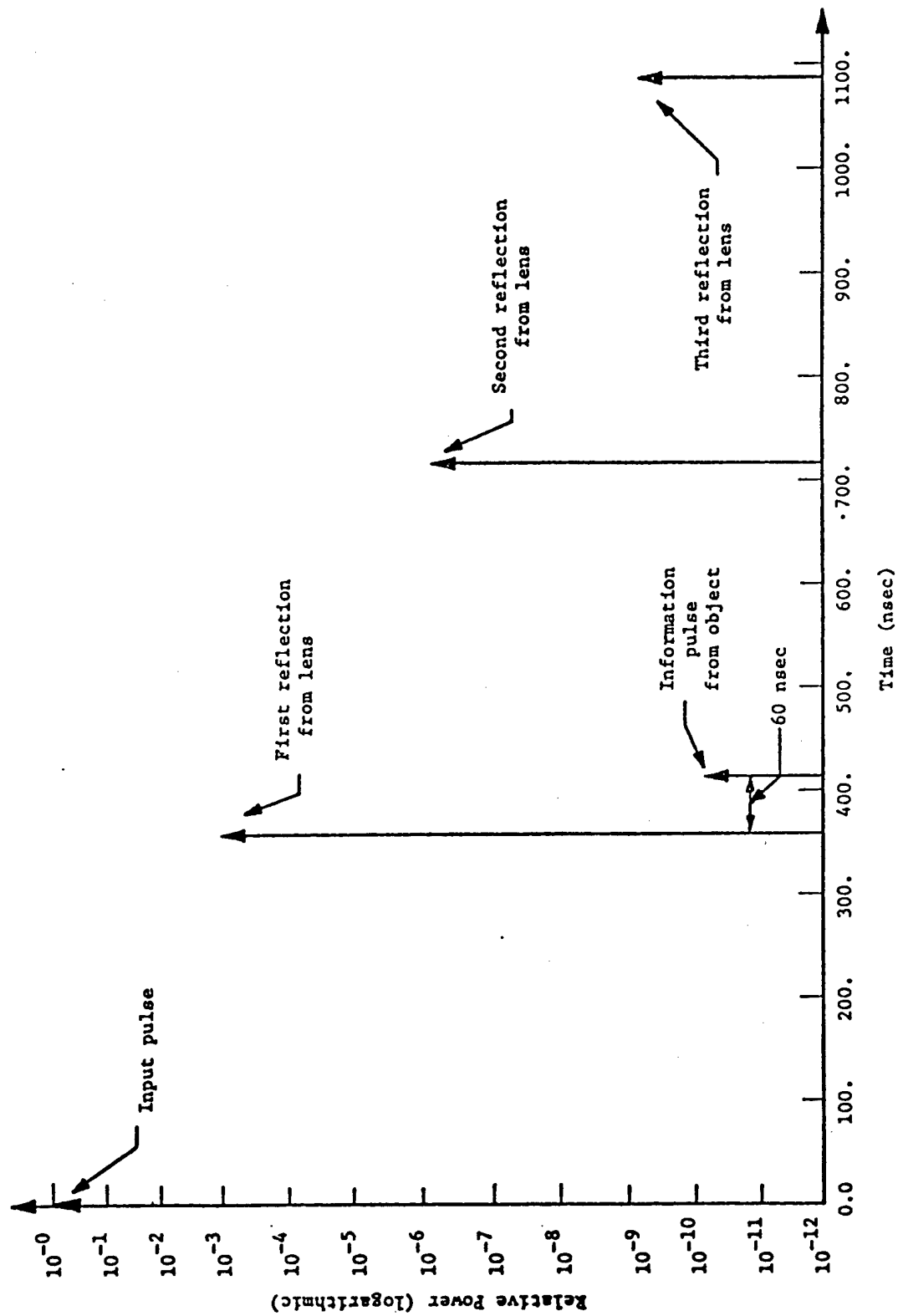
FIG. 2.2--Timing diagram for an acoustic lens: $A\ell_2O_3 - H_2O$ lens; $R = 40$ µm; $d = 2$ mm.

177

A
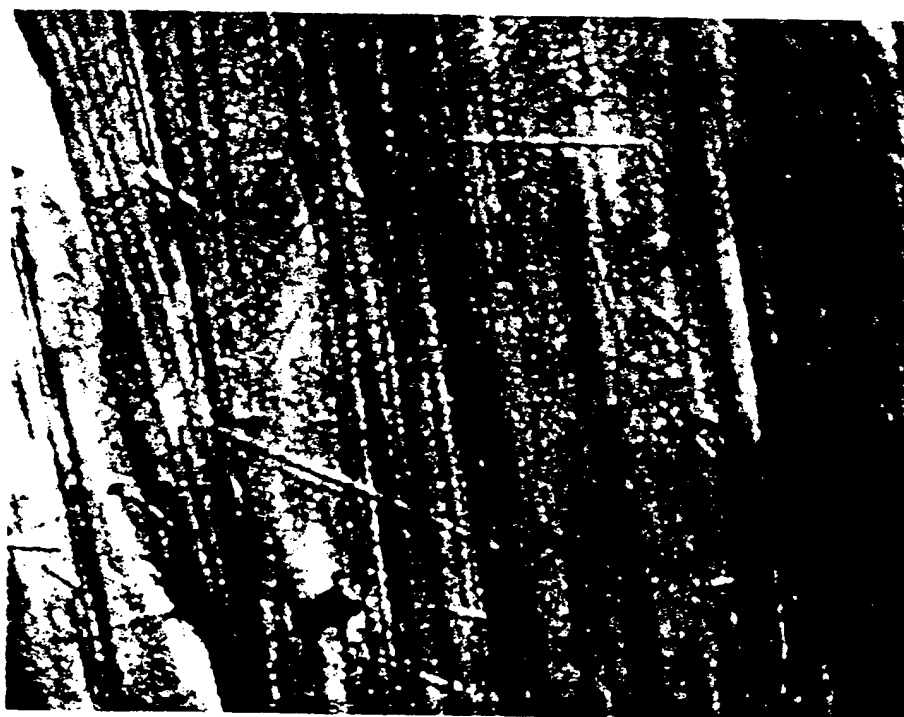
B

FIG. 4.3 Comparison of the optical A and the acoustic B micrograph
of a polished surface of polycrystalline silicon.

(a)

(b)

FIG. 4.6---Optical (a) and acoustic (b) comparison
of polished brass surface. Field of
view is 55 × 90 μm.
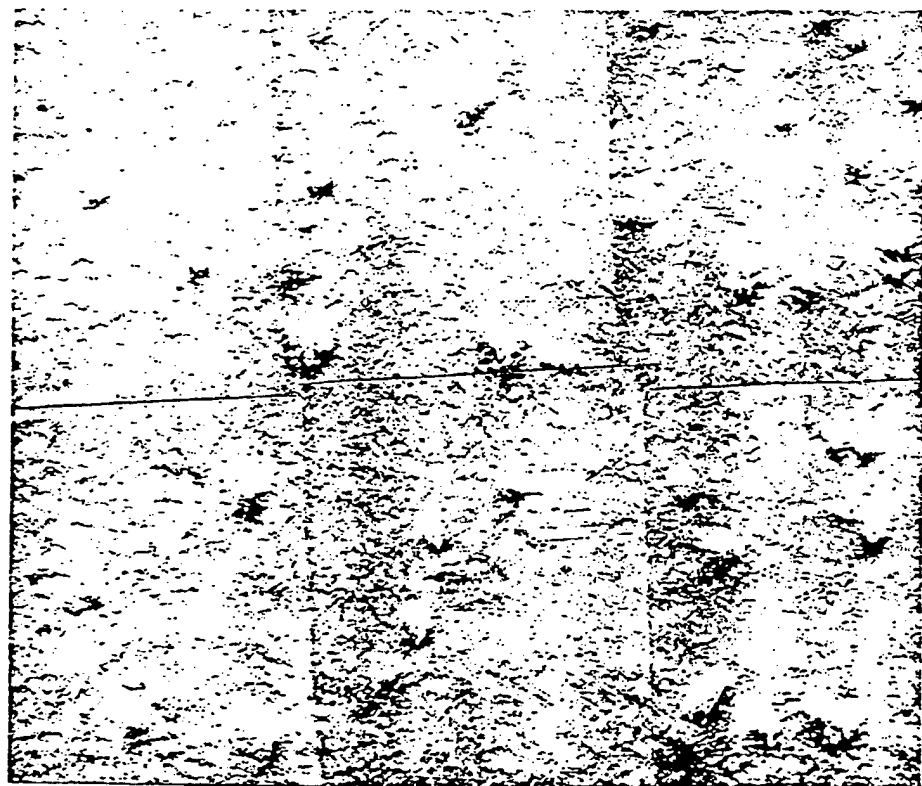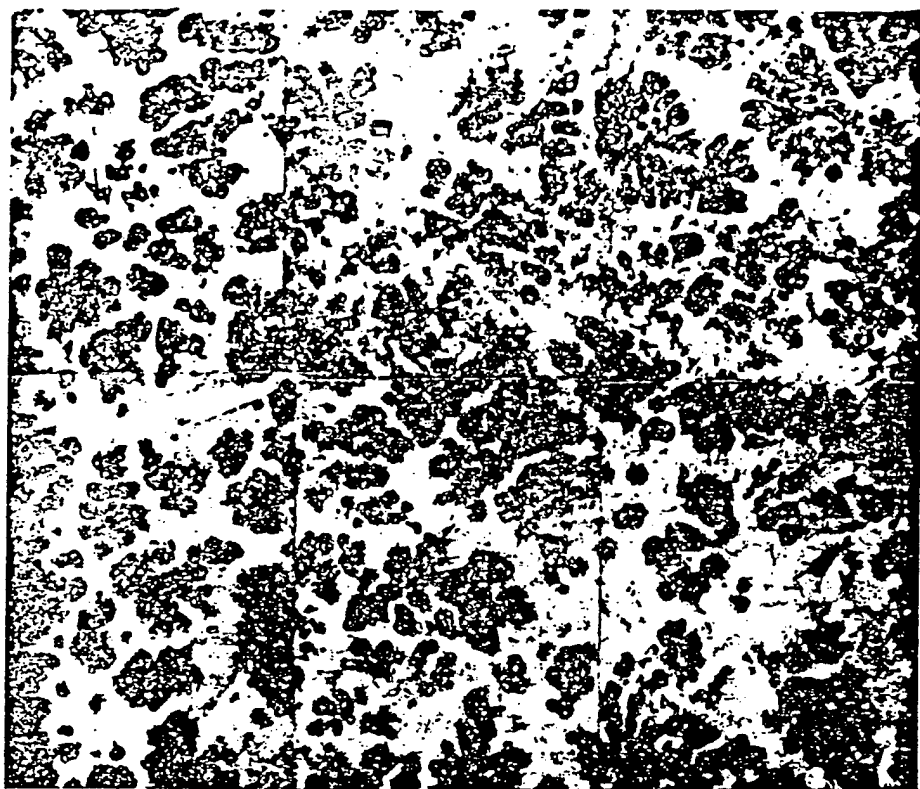
(b) ACOUSTIC (1100 MHz)

(a) OPTICAL DIC

100μm

Co—Ti ALLOY SHOWING 4 PHASES. OVERALL COMPOSITION, 59% Co 41% Ti

FIG. 6.1--Comparison of optical and acoustic images of a Co-Ti alloy.

STUDIES OF ADHESION

1000 A Cr ON GLASS (GOOD ADHESION)

(a) OPTICAL

30 μ
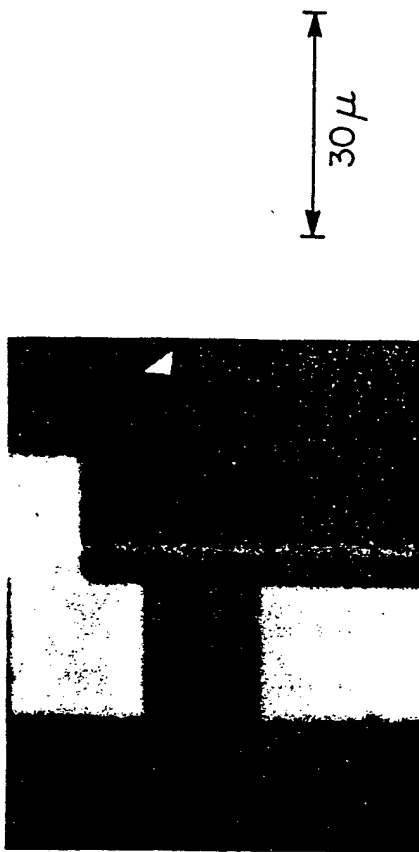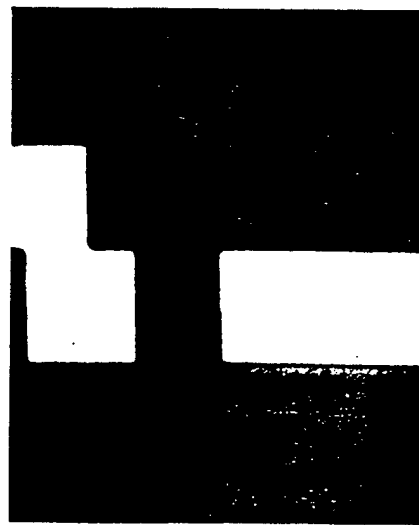
(b) ACOUSTIC, FOCUSED ON Cr
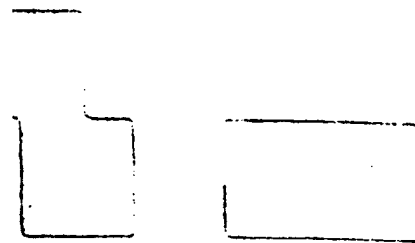
(c) ACOUSTIC, FOCUSED ON GLASS

FIGURE 4-1

182

1000 Å Cr ON GLASS (POOR ADHESION)

100 μ
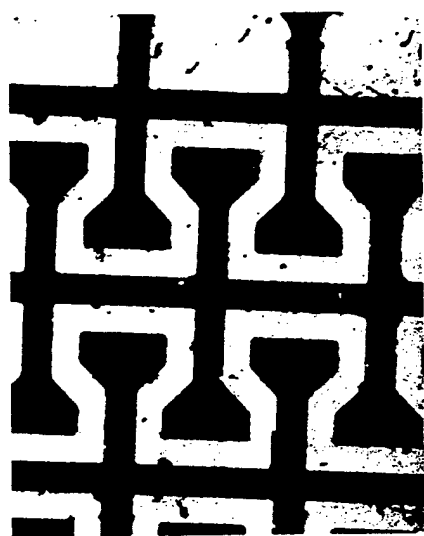
(d) OPTICAL

(e) ACOUSTIC Z = -1μ

(f) ACOUSTIC Z = -0.5μ

FIGURE 4-2

183

1
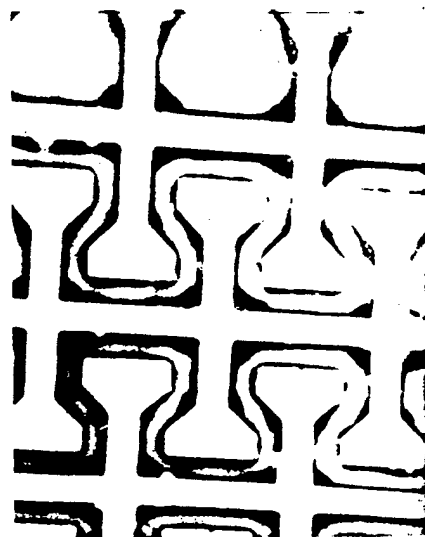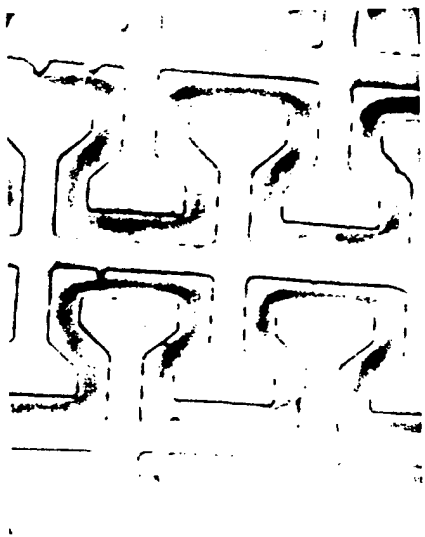
Acoustic Microscopy Analysis of Microelectronic Interconnection and Packaging Technologies

Gabriel M. Crean, Colm M. Flannery, and Séan Cian Ó Mathúna

1.1. Introduction

Several advanced packaging and interconnection technologies are currently under development to meet the requirements of complex, large and/or high-speed microelectronic integrated circuits (IC) and systems. These include the design of multilayer interconnection substrates (see Fig. 1.1) for Multichip Module (MCM) applications[1] and the development of multilayer ceramic pin grid array packages (see Fig. 1.2) for large-area die size, high pin count ICs.[2] A development road map[3] for single-chip packaging technologies is presented in Fig. 1.3.

However, the use of both novel materials and an increasing number of processing steps in such technologies has posed serious yield and reliability issues. The availability of characterization tools for on-line process monitoring and failure analysis of packaging technologies are therefore becoming increas-

GABRIEL M. CREAN AND COLM M. FLANNERY • National Microelectronics Research Centre, University College, Lee Maltings, Cork, Ireland. SÉAN CIAN Ó MATHÚNA • Power Electronics Ireland, National Microelectronics Research Centre, UCC, Ireland
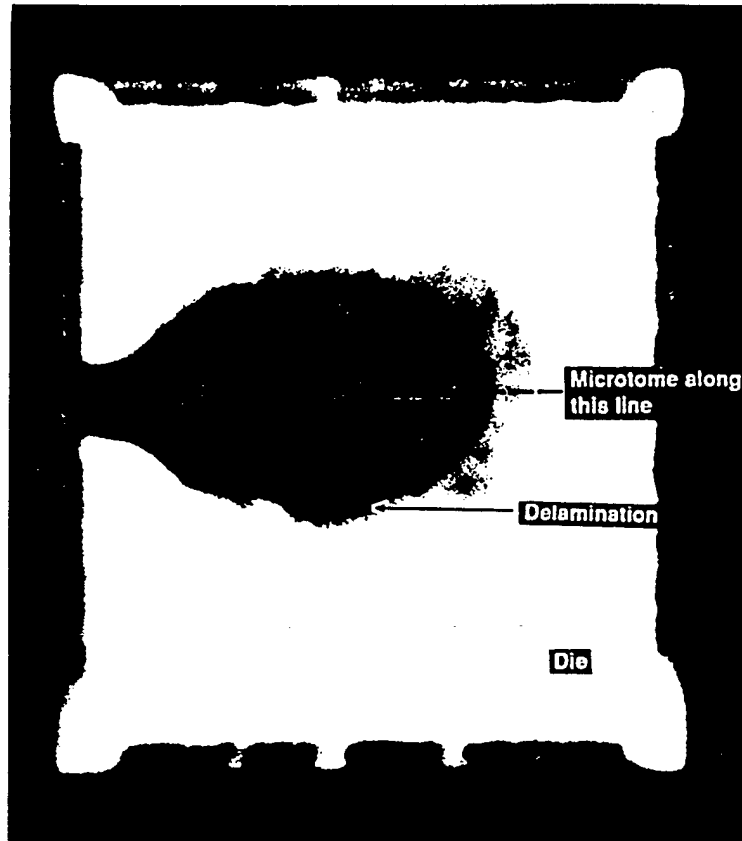
Figure 1.23. Acoustic micrograph of the die–plastic interface of the 160-lead PQFP. The white region corresponds to delaminated areas.

4 μm. The pattern of delamination on the chip surface was correlated with results from finite-element thermomechanical analysis of the package structure.[14] From this modeling, delamination at the material interfaces was found to increase significantly the level of thermomechanical stress. Once delamination commenced, the line of stress concentration followed the adhesion edge on the chip surface, since this was the first point of constraint encountered by the moulding compound. Delaminations may thus propagate and spread throughout the whole package, resulting in the delamination pattern observed in Fig. 1.23.

185
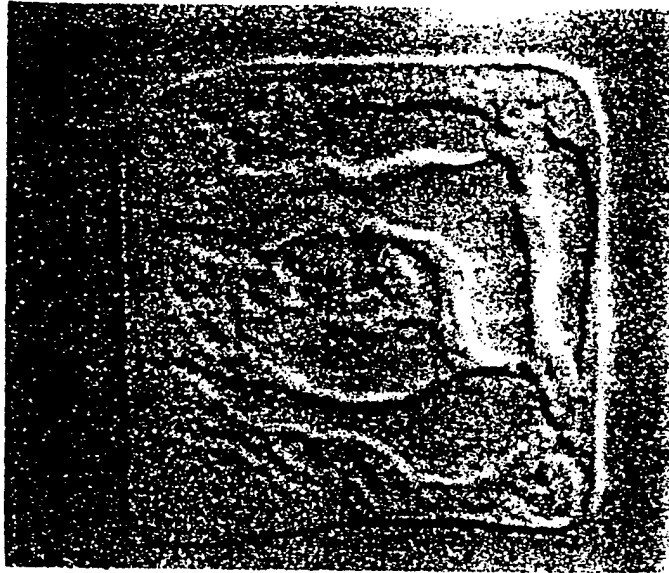
Figure 1.31. A pseudo-three-dimensional X-ray micrograph of the metallic substrate sample detailing voiding in the silver epoxy die attach

both scattering of the acoustic signal at void edges and to voids not extending fully through the thickness of the die attach.

Since the die was removed from the metal substrate with the die attach still adhering, an optical micrograph of the die attach on the silicon die is presented in Fig. 1.33. In regions where no wetting occurred, the dark metal substrate is visible; these correspond to voids extending throughout the thickness of the die attach. Adjacent lighter regions exist where wetting of the metal substrate occurred without adhering to the die. Shapes of the voids as shown in Figs. 1.31 and 1.32 correlate very closely with voids visible in Fig. 1.33. At the edges of the large through-voids, thin layers of the die attach are observable, and the thinner stream like voids do not appear to extend fully through the thickness of the die attach. Enhanced optical microscopy revealed that the die attach at void edges and smaller voids is a thin, textured, 10–15 μm thick layer attached to the back surface of the silicon die. Figure 1.32 shows that all the dark areas in the acoustic micrograph correspond to areas where there are thin, rough die-attach layers, that is, at edges of large voids and at the thinner nonthrough-voids, where the acoustic signal is scattered at the rough, nonplanar void edges.

For several of the metal substrate samples, extensive delaminations were visible in acoustic micrographs at the die-attach–substrate interface, but they were not detected at the die–die-attach interface. This indicated poor adhesion of the metal substrates. Subsequently die shear tests removed the die and die

Figure 1.32. An acoustic micrograph of the metallic substrate sample detailing voiding in the silver epoxy die attach as viewed through the die.

attach from the metal substrates with little effort, indicating a wettability-processing issue. This is consistent with the absence of a gold layer to assist adhesion.

It is evident from the preceding results that interpreting acoustic micrographs of die attach is complex. Both bright and dark areas can indicate the presence of voids in die attach depending on their size, thickness, and shape. In addition the sides of large voids may have sloping edges of die attach that scatter the acoustic signal, resulting in an underestimation of the actual void area. The dominant physical mechanism resulting in dark areas in the SAM images appears to be scattering from nonplanar features.

This study has investigated only one particular type of die attach on a number of substrates. For a thorough inspection of die-attach quality, reflection acoustic imaging from both sides and acoustic transmission imaging are desirable;

Figure 1.33. An optical micrograph of the die attach on the silicon die after removing the substrate.

however they may not be possible for some packages, such as those containing a cavity where only reflection imaging from one side is practicable. Further work on the influence of different die-attach materials, curing processes, and the influence of lead frames should be performed. However for the silver epoxy die-attach material investigated, it is possible to elaborate a methodology to determine the presence of voids and to identify adhesion integrity issues using acoustic microscopy.

1.5. Multilayer Interconnect

The requirement to interconnect complex and high-speed ICs has resulted in the design of several multilayer interconnection technologies and the development of multichip modules, such as those in Fig. 1.1.[1,2] The most innovative
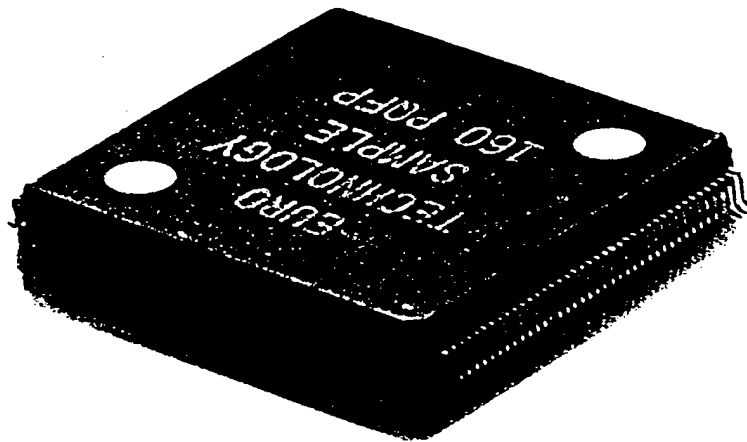
Figure 1.21. A 160-lead PQFP for surface-mount application.
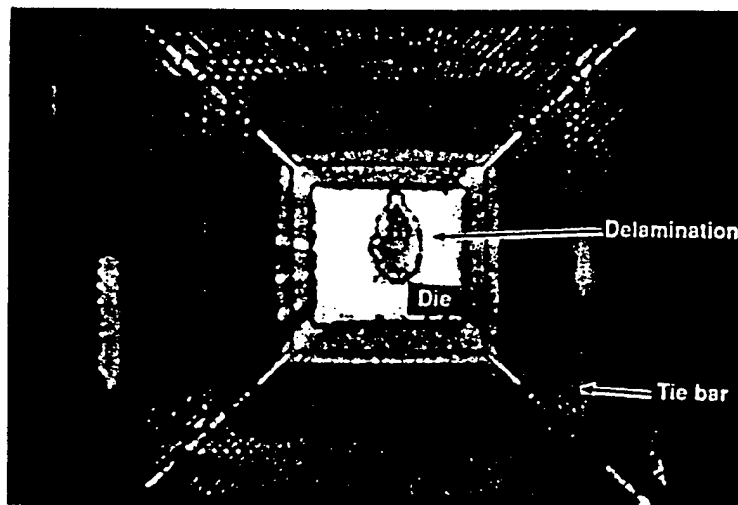
Figure 1.22. Acoustic micrograph of a 160-lead PQFP imaged through the top of the package and detailing the lead frame, tie bar, IC, and die pad.

189

IMAGING AT HIGH FREQUENCIES
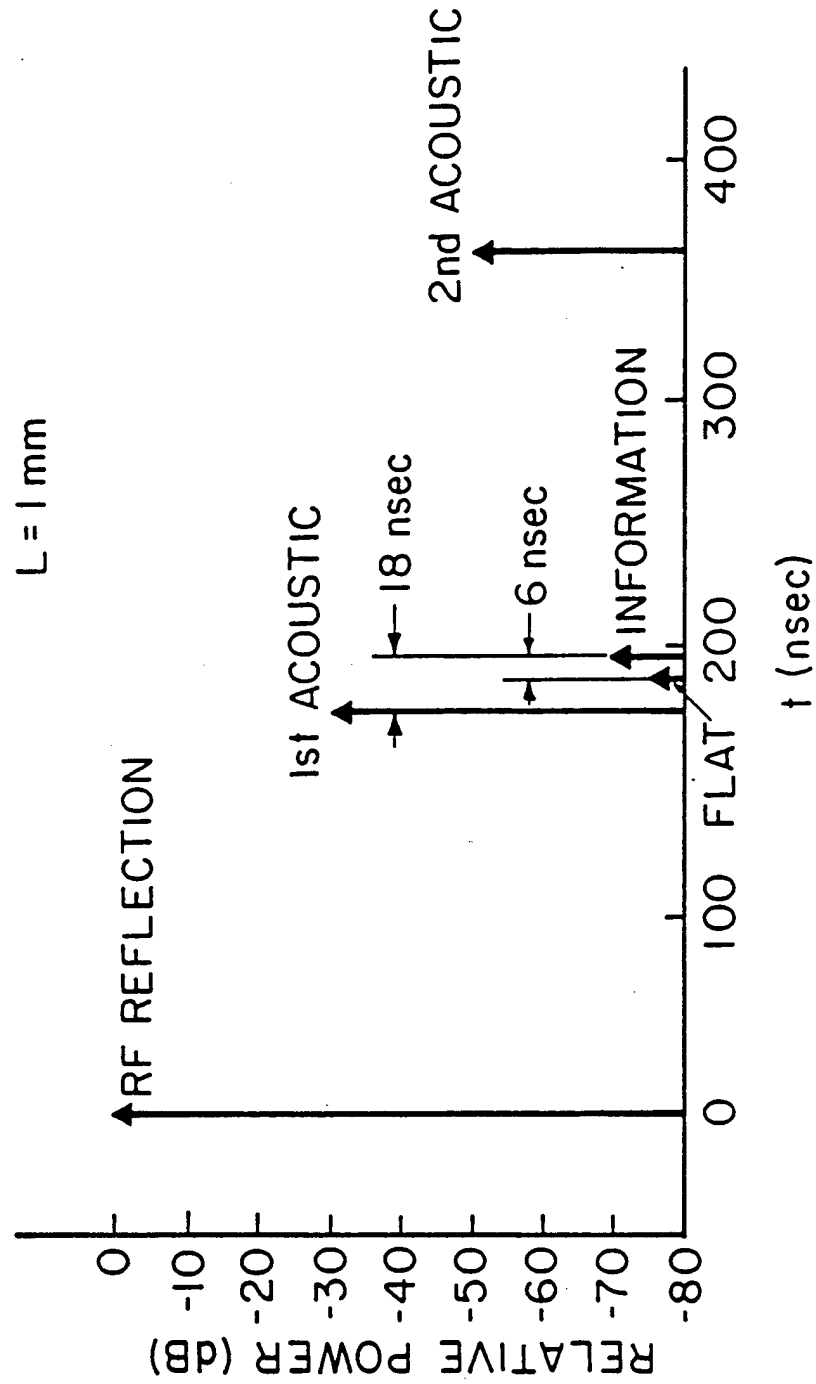
FIGURE 3.1--Typical timing and relative amplitudes of signals received from the transducer
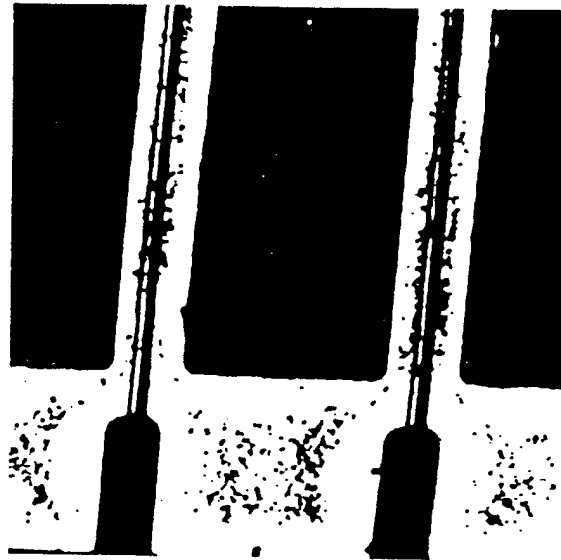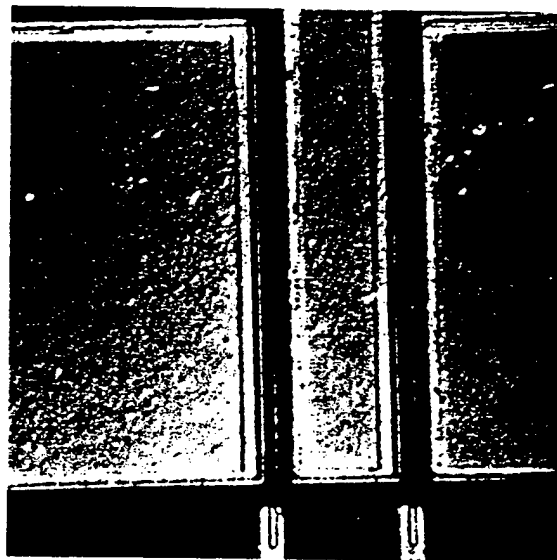
FIGURE 3.6--(a) Acoustic and (b) optical images of a
0.5 µm gate FET. Acoustic microscope
is at 3.7 GHz.
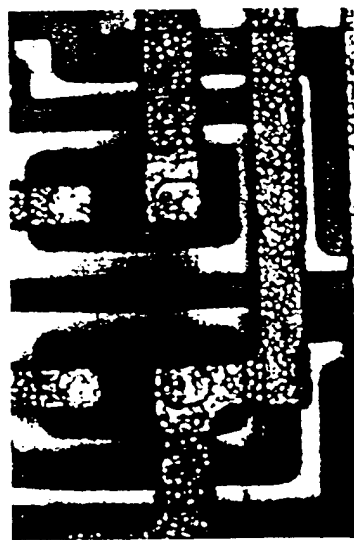
(a)

(b)

(c)

(d)

FIGURE 3.7--(a) and (b) Acoustic (4.4 GHz), (c) optical and (d) SEM images of a bipolar integrated circuit.

193
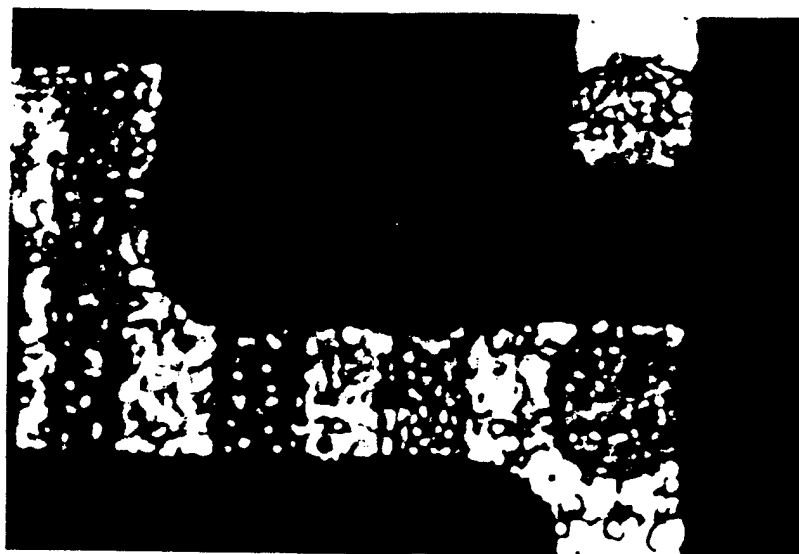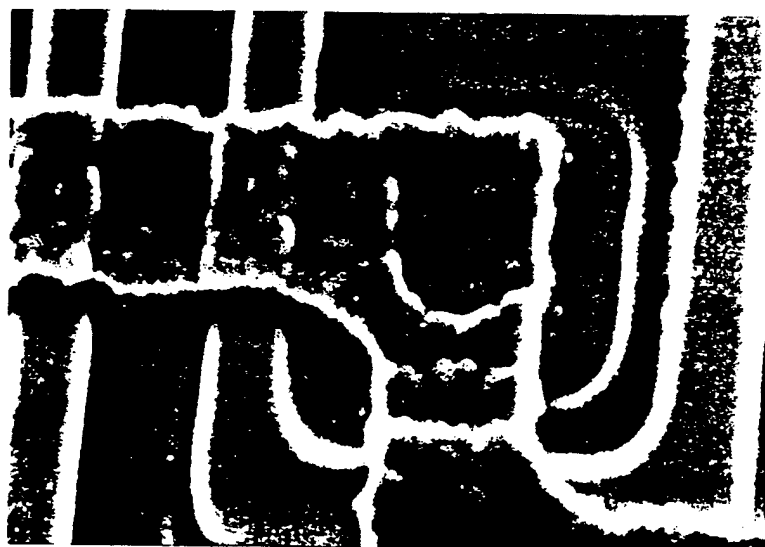
FIGURE 3.8--(a) Acoustic (4.4 GHz) and (b) SEM images of
the sample in Fig. 3.7.

194
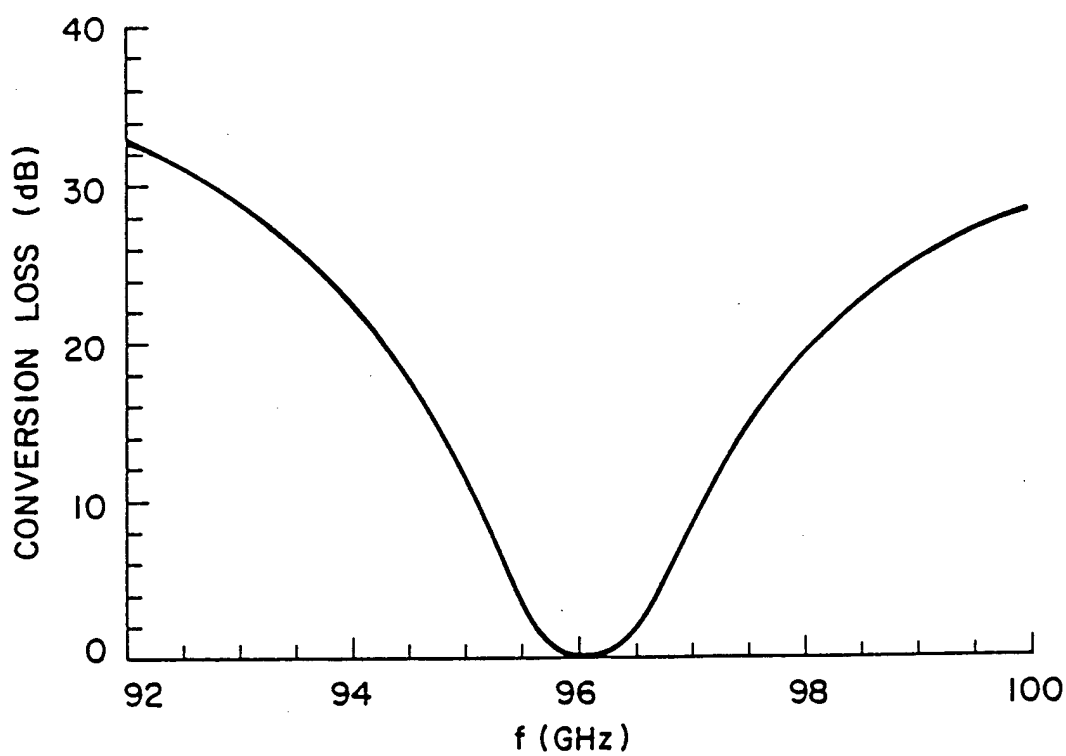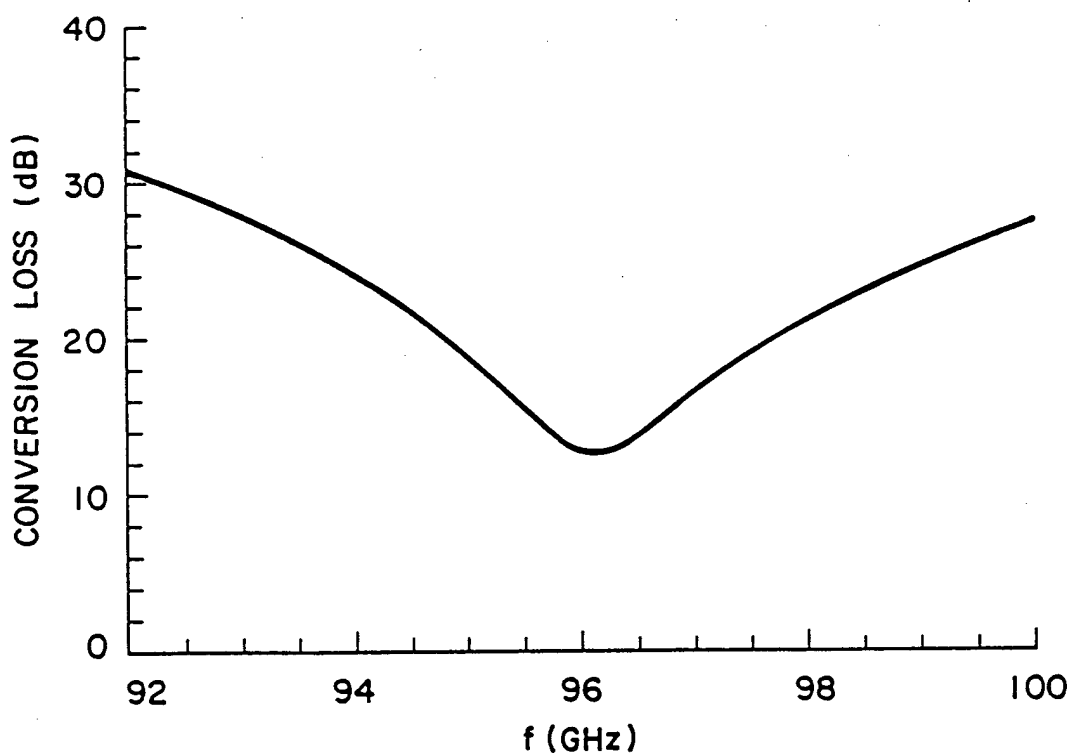
FIGURE 5.2--Conversion loss (theoretical) vs. frequency for 96 GHz
transducers. (a) All losses are neglected. (b) Conductor
losses are included.
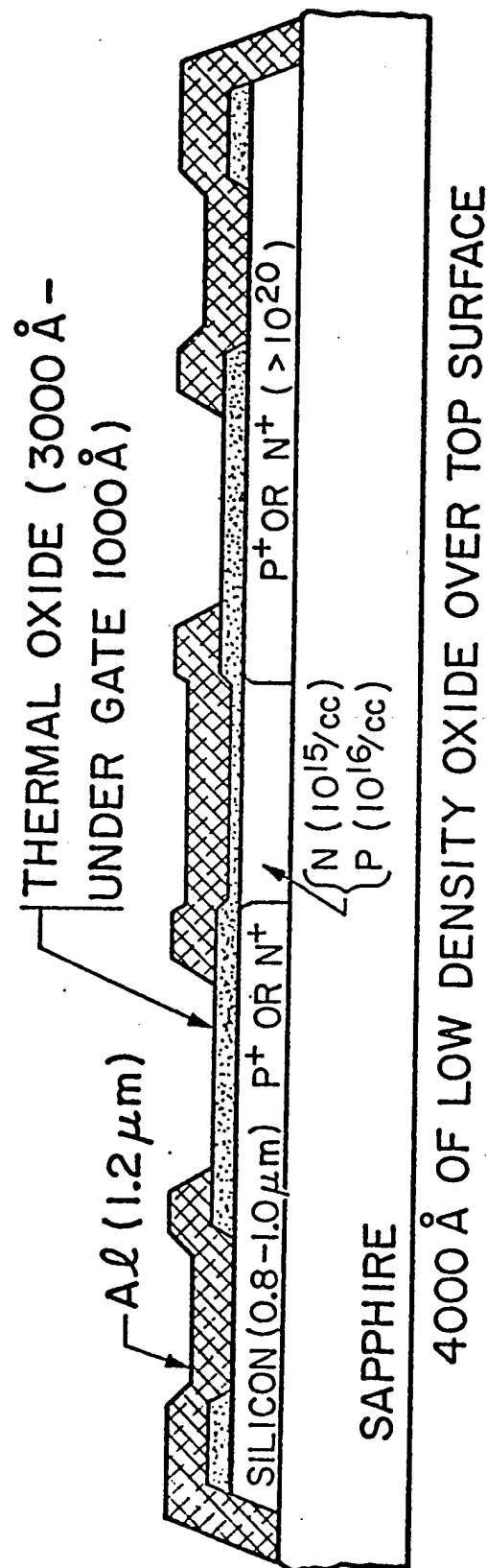
FIG. 6.7--The profile of the H-MOS transistor.

THERMAL OXIDE (3000Å – UNDER GATE 1000Å)

Aℓ (1.2 μm)

SILICON (0.8-1.0 μm) P⁺ OR N⁺

N (10¹⁵/cc)
P (10¹⁶/cc)

P⁺ OR N⁺ (>10²⁰)

SAPPHIRE

4000Å OF LOW DENSITY OXIDE OVER TOP SURFACE

OPTICAL INTERFERENCE CONTRAST by Koch (×1120)
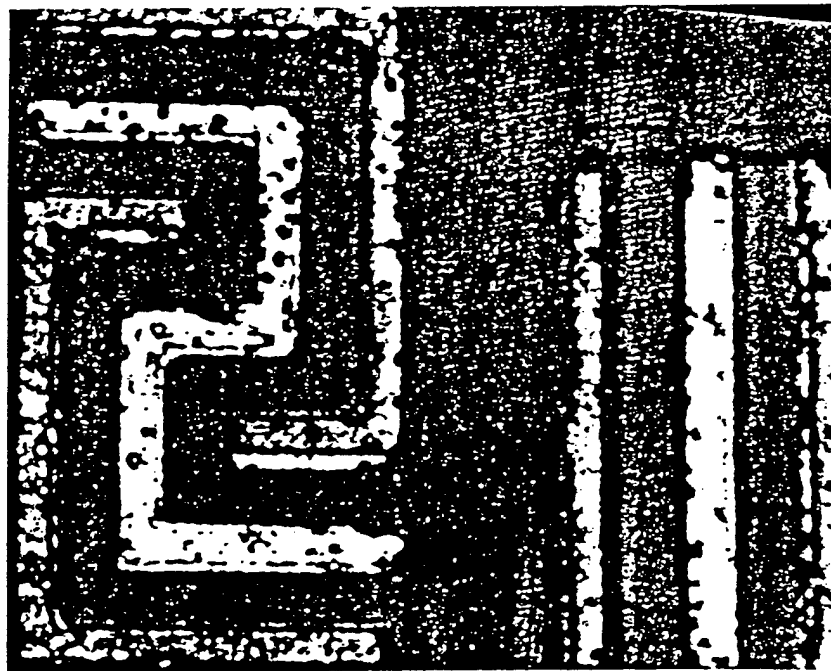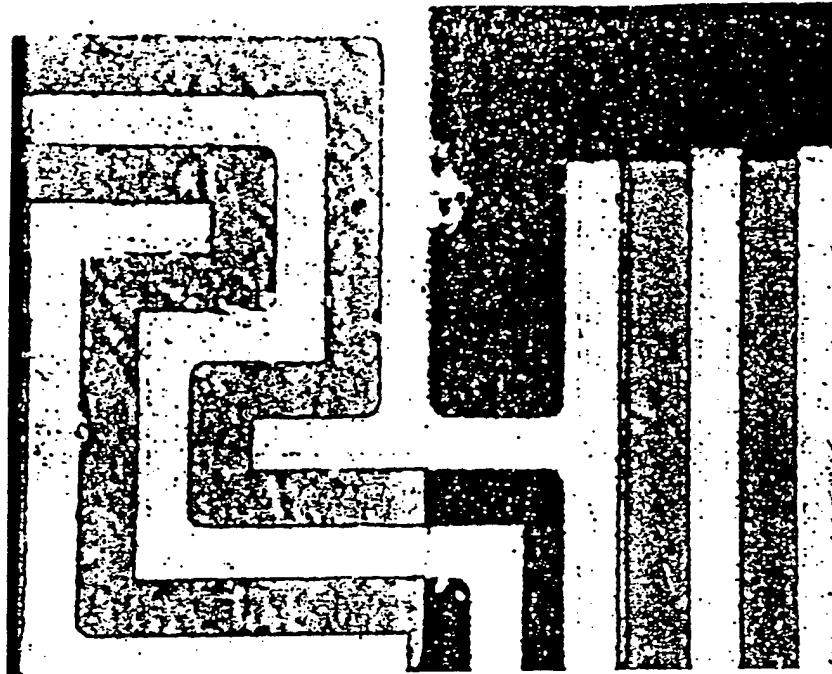
ACOUSTIC f = 1100 MHz

SOS Chip A — STDL-4

Rockwell CMOS

FIG. 6.9--Optical and acoustic images of H-MOS transistors on a SOS chip.
Acoustic image is recorded at Z = -4.5 microns.
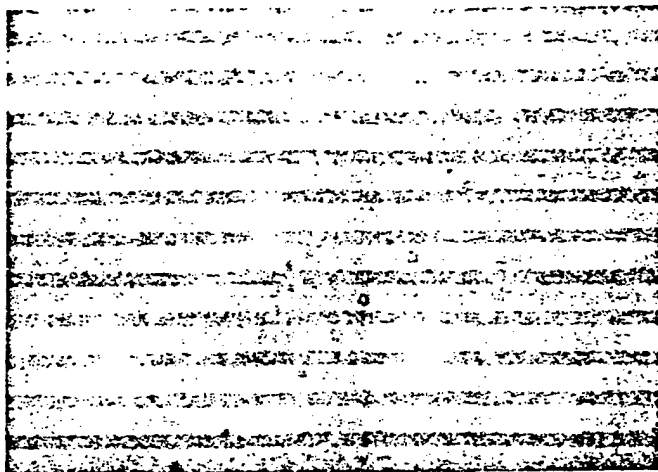
(A) ACOUSTIC
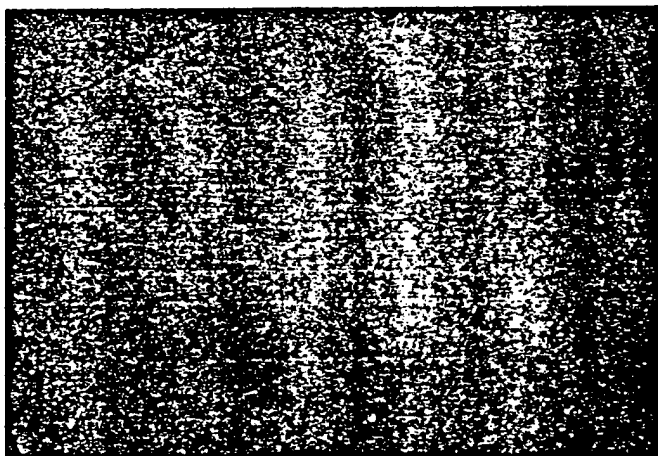
(B) DIC OPTICAL (X300)

FIG. 6.15--Acoustic and optical images of an exposed but undeveloped photoresist layer on a silicon wafer.

NON-LINEAR IMAGING

AT LOW TEMPERATURE
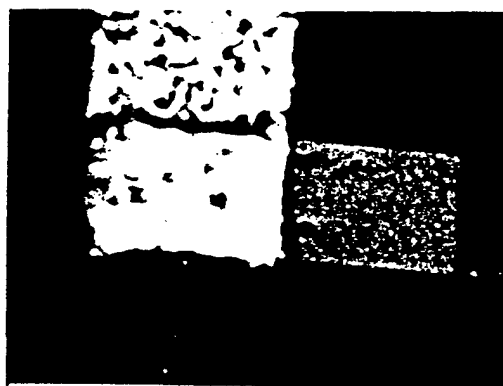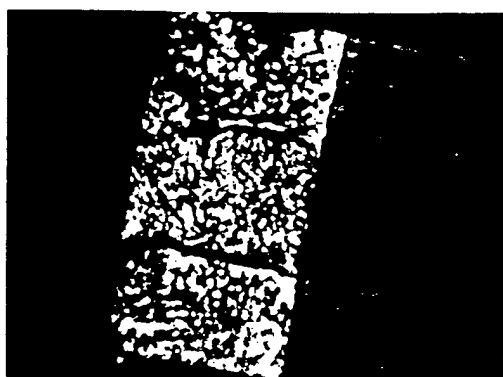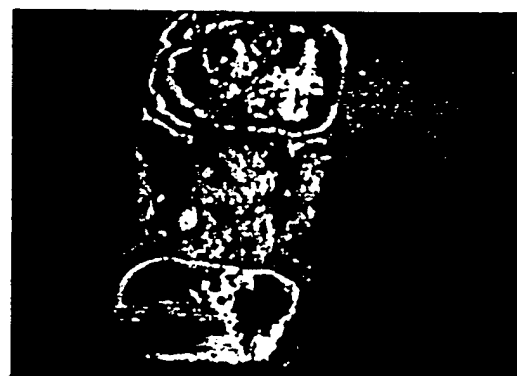
FIG. 3-23. Optical (left) and acoustic (right) images of a silicon-on-sapphire integrated circuit. The acoustic images were taken at 630 MHz in helium at 1.95K.
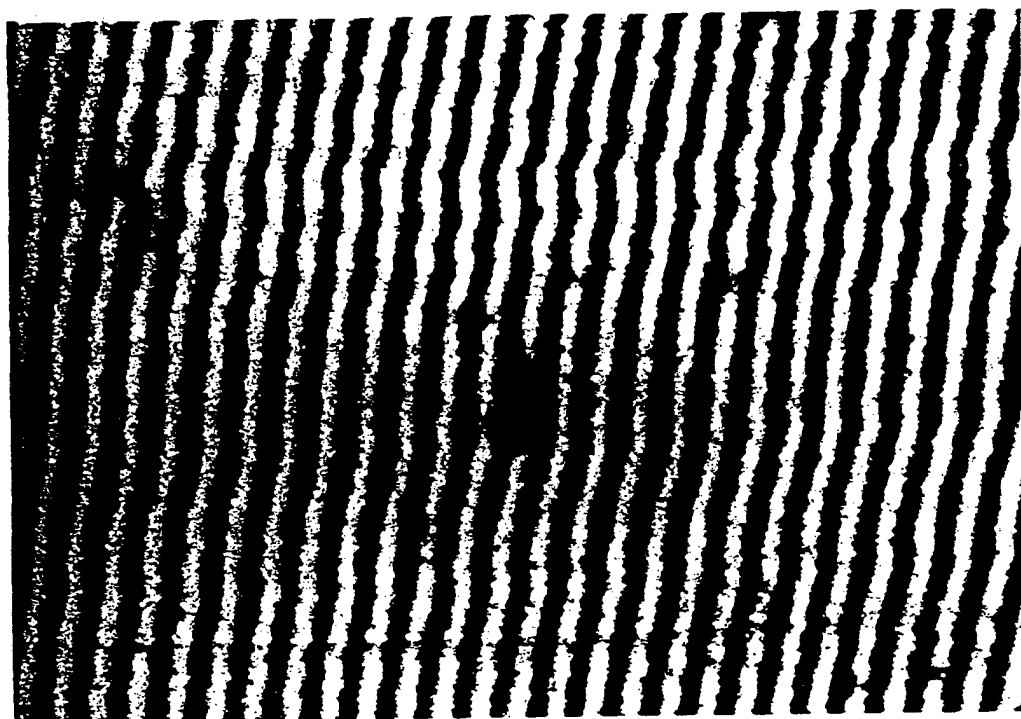
(a) ACOUSTIC IMAGE IN ^{4}He

T = 1.95 K f_0 = 840 MHz

PHOTO RESIST
1000 Å THICK

Aℓ 700 Å THICK

~ 1000 Å

~ 2500 Å
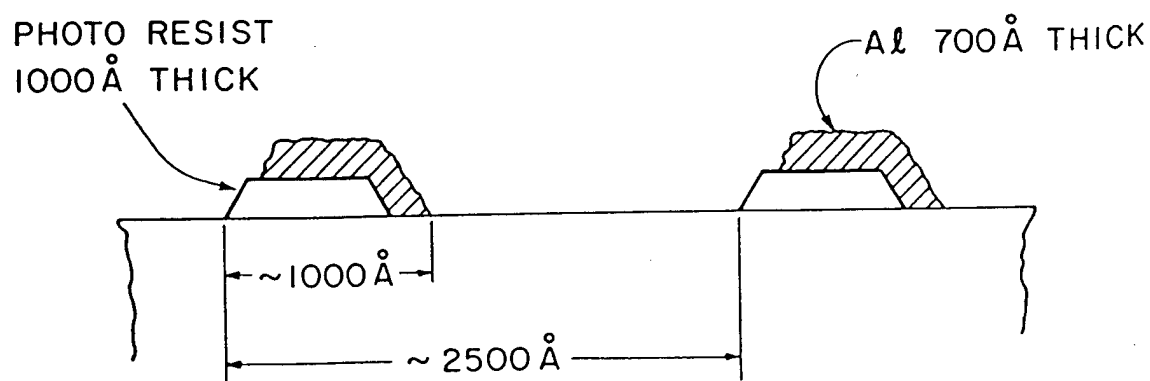
(b) CROSS SECTION

FIG. 3-25.   Acoustic image of grating with 2500 Å period.
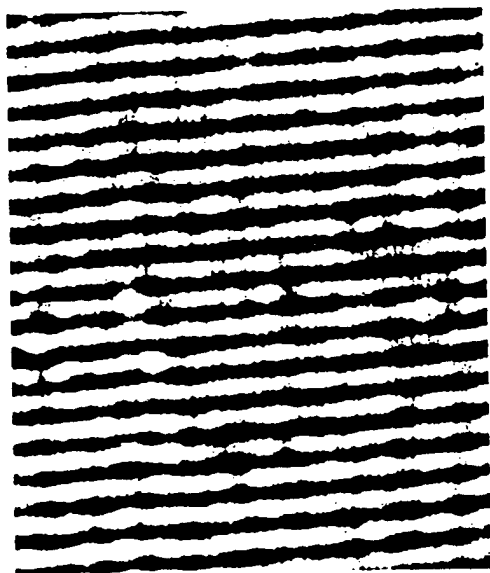
(a)

(b)
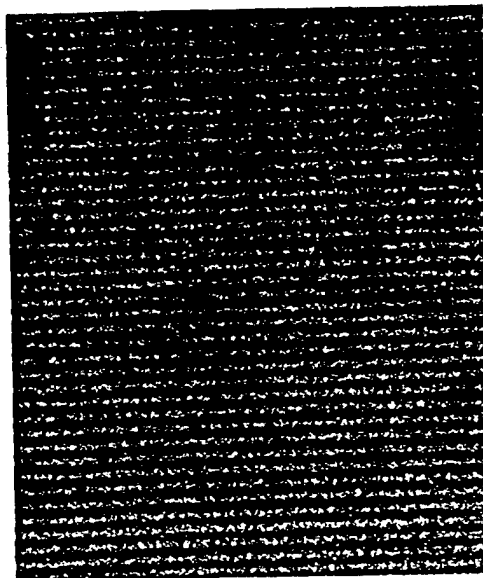
(c)

FIG. 4-3. Images of grating with 4000 Å period. The acoustic image (a) was taken in liquid argon at
[illegible] (b) and (c) are by scanning electron and optical microscopy, respectively.

NONLINEAR RESOLUTION OF 0.29 μm GRATING IN LIQUID NITROGEN

$f_o = 2.0\,GHz, \quad \lambda_o = 0.43\,\mu m$

(a) 10 dBm

(b) 15 dBm

(c) 20 dBm

(d) 25 dBm

FIG. 5-3.   Demonstration of resolution improvement as a function of lens input power.

203
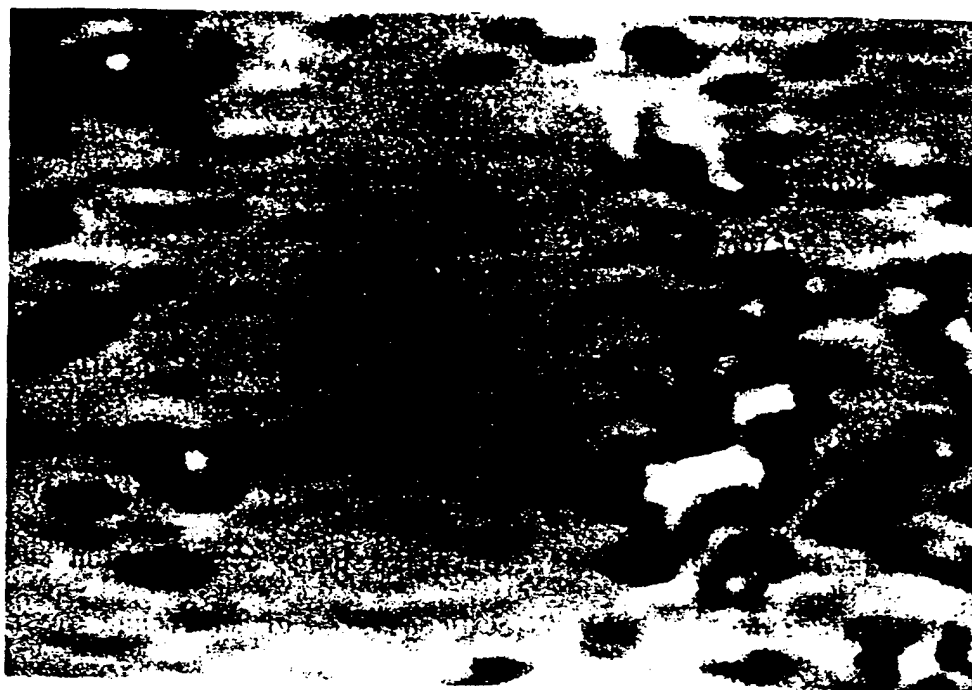
(a)
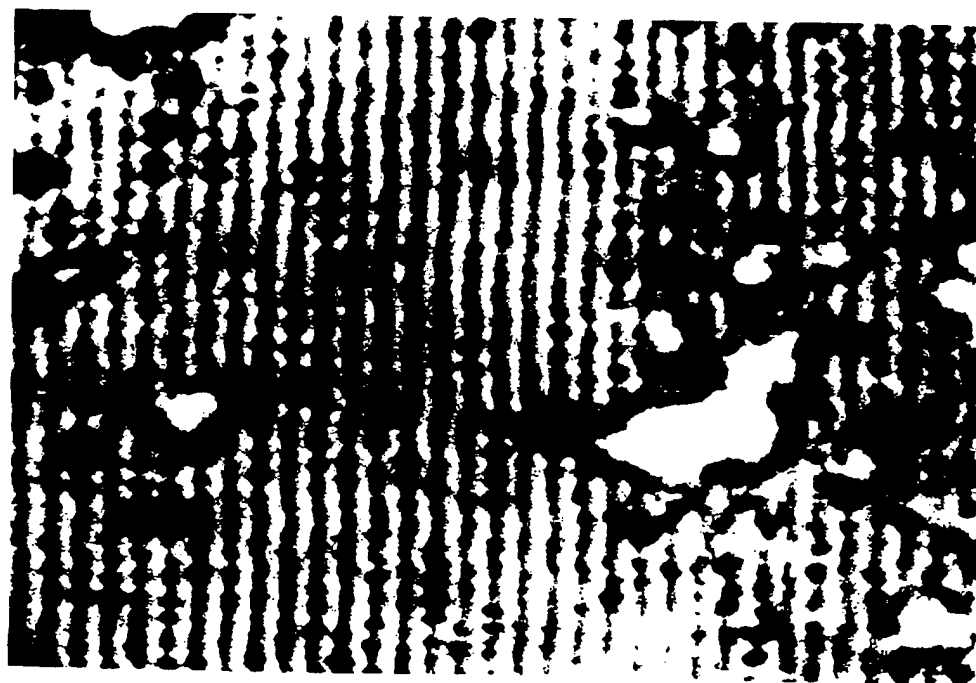
(b)

FIG. 5-1.   Nonlinear resolution of 2000 Å grating in liquid argon.
Image (a) was taken at low input power, while (b) was
taken at higher power

204

Cryogenic acoustic microscopy

J. Heiserman, D. Rugar, and C. F. Quate

Edward L. Ginzton Laboratory, Stanford University, Stanford, California 94305
(Received 17 October 1979; accepted for publication 8 February 1980)

Resolution in the scanning acoustic microscope is determined by wavelength which is in turn limited by attenuation in the acoustic medium. In order to make use of the low attenuation found in cryogenic liquids we have developed an acoustic microscope suited for use at low temperatures. In this paper we present images taken in liquid argon held at 85°K and in superfluid helium at 1.95°K. In liquid argon, wavelengths as short as 0.43 μm were achieved while in preliminary work with superfluid helium a wavelength of 0.36 μm was used. In order to operate in liquid helium, it was necessary to improve the power transfer from the acoustic lens to liquid helium by using double quarter-wave matching layers. Techniques of fabricating and testing these layers are described. Finally, prospects for operating at still shorter wavelengths in superfluid helium held at temperatures below 0.5°K are discussed.

PACS numbers: 43.35.Lq, 43.35.Ns, 43.35.Yb, 43.35.Sx.

INTRODUCTION

Since the introduction of the scanning acoustic microscope by Lemons and Quate in 1974,[1] the resolution has been increased to the point where wavelengths comparable to those of visible light are employed.[2] Progress to shorter wavelengths is hampered by losses in the coupling fluid which limit the maximum frequency of operation and hence the wavelength. Through the use of cryogenic liquids we have extended the limit on minimum wavelength beyond what is presently possible at room temperature. We expect as we develop this technique further, wavelengths much shorter than those of visible light and hence resolution substantially better than the optical microscope will be within reach.

Full descriptions of transmission and reflection mode scanning acoustic microscopes are given elsewhere.[3] Here we will simply summarize the operations of the reflection mode as shown in Fig. 1. A collimated beam of acoustic pulses is generated by a zinc oxide thin film transducer deposited on a sapphire rod. At the front end of the rod a small spherical depression in contact with the coupling liquid (usually water in the room temperature instrument) serves as a lens. Because of the large velocity difference between sapphire and water (a factor of 7.4) acoustic power incident on the interface is focused to a spot in the liquid. The object to be examined is placed at or near the focus. Sound reflected by the object is collected, recollimated by the lens and detected by the transducer which is sensitive to both the phase and the amplitude of the returning acoustic wave. The detected signal is used to modulate the brightness of an oscilloscope display or is stored in an analog scan converter. The object is then scanned in a raster pattern and the motion is synchronized with the x and y axes of the display. The image obtained in this way can be recorded on photographic film.

The key to the acoustic microscope lies in the spherical interface between the sapphire and the liquid. There, the velocity ratio is large and the spherical aberrations—proportional to the square of this ratio[3]—are negligible. The acoustic beam is focused to a spot whose diameter is limited only by diffraction. It is not possible to construct an analogous aberration-free single surface optical lens since optical index of refraction ratios are limited to about two.

Although large velocity differences are beneficial, large acoustic impedance differences limit the power transfer to the coupling liquid and are undesirable. Power transfer efficiency is improved by using one or more quarter wave matching layers. In the sapphire-water case a single glass quarter-wave matching layer is used.

Resolution in the scanning acoustic microscope improves as wavelength in the coupling liquid is reduced. In most liquids at high frequencies attenuation increases as frequency squared and this factor limits the maximum operating frequency. This limitation can be offset by reducing the liquid path by making lenses of smaller radius. Current versions of the acoustic microscope use a lens with a radius of about 30 μm.

For maximum operating frequency one must choose a coupling liquid with low attenuation and low acoustic velocity. In order to assess the usefulness of various fluids as a working medium for the acoustic microscope, a coefficient of merit has been defined which compares each fluid to water.[4] This coefficient is defined as follows:

$$M = \lambda_w/\lambda. \tag{1}$$

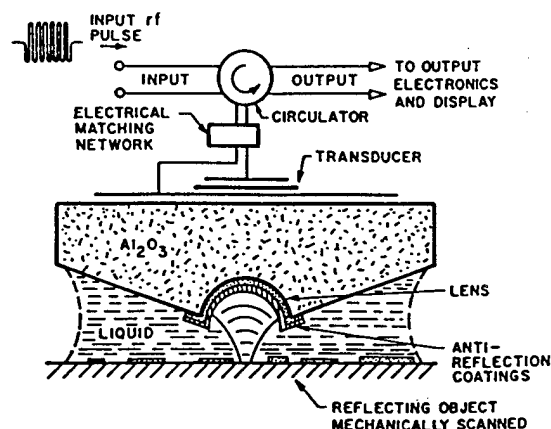Here λ_w is the minimum achievable wavelength in water

FIG. 1. Schematic of a reflection mode scanning acoustic microscope.

TABLE I. Properties of various cryogenic liquids compared to water.

Liquid	Temp °K	$\alpha/f^2 \times 10^{17}$ dB s²/cm	ρ g/cm³	$C \times 10^{-5}$ cm/s	$Z \times 10^{-5}$ g/cm²s	M	Refs.
H_2O	25 °C	191	1.0	1.5	1.5	1.0	6
	60 °C	95	1.0	1.5	1.5	1.4	6
O_2	90	86	1.14	0.90	1.0	2.5	6–8
N_2	77	120	0.80	0.85	0.68	2.2	6–9
H_2	20	49	0.07	1.19	0.08	2.3	6
Xe	.166	191	2.9	0.63	1.8	2.4	6
Ar	87	132	1.4	0.84	1.2	2.2	10
Ne	27	201	1.2	0.60	0.72	2.4	6
He	4.2	1966	0.147	0.183	0.027	2.5	6,11
	1.95	610	0.146	.227	0.033	3.7	11,12
	0.4	15 [a]	0.145	0.238	.035	23	13,14

[a] As described in the text, at this temperature α scales as f. This value is from the data of Ref. 14 extrapolated to 1 GHz.

and λ is the achievable wavelength in the given liquid. We can express M in terms of the absorption coefficient (α/f^2) and the velocity of sound C. We will make our comparison with water at 25°C, where α_w/f_w^2 equals 191×10^{-17} dB s²/cm and $C_w = 1.5 \times 10^5$ cm/s. We take αL, the total attenuation through the liquid cell, and L, the acoustic path length in the liquid, to be constant. We can then write

$$\lambda^2 = \frac{C^2}{f^2} = \frac{C^2}{\alpha L}\frac{\alpha}{f^2} L . \tag{2}$$

In place of Eq. (1) we can now write

$$M = \frac{\lambda_w}{\lambda} = \frac{C_w}{C}\left(\frac{\alpha_w/f_w^2}{\alpha/f^2}\right)^{1/2} . \tag{3}$$

For those liquids where the attenuation scales as frequency squared we let $f_w = f$. For other liquids we must specify the frequency used in the comparison. Since we want liquids with large values of M we look for those liquids with a low value for both the velocity C and the attenuation α. Several fluids of potential interest are listed in Table I. There we find that values for M between 2 and 4 can be realized for a number of cryogenic fluids. For the special case of helium very large values may be possible.

A number of factors can contribute to contrast in the scanning acoustic microscope. Specular and diffuse scattering occur at the liquid–object interface. Some power is converted to longitudinal and shear waves in the sample which can be scattered by boundaries or internal imperfections or absorbed. Some power may be converted to surface waves at the liquid–sample interface. This "leaks" back into the liquid with its phase shifted and may be detected by the transducer. This last mechanism is very sensitive to the lens to object spacing and leads to great sensitivity to height changes and material differences.[5]

I. CRYOGENIC APPARATUS

The mechanical apparatus used in the room temperature acoustic microscope is not well suited to use with cryogenic fluids. It is necessary to use a design in which the complete microscope assembly is immersed directly in a cryogenic bath. It is desirable, however, to adjust the alignment with water at room temperature before cooling down to cryogenic temperatures. Thus strict thermal compensation is needed to ensure that the lens to object spacing does not change between room temperature and the cryogenic temperatures, i.e., during a change of up to 300 °K.

A diagram of the assembly used is shown in Fig. 2. Coarse adjustment of the focal distance is made with a pushrod that is controlled by a differential micrometer head at the top of the assembly. The pushrod is spring loaded by two diaphragms in the cryogenic liquid and thus has negligible backlash. Fine focus is achieved by a piezoelectric positioning element (Burleigh Instruments PZT aligner–translator) located in the
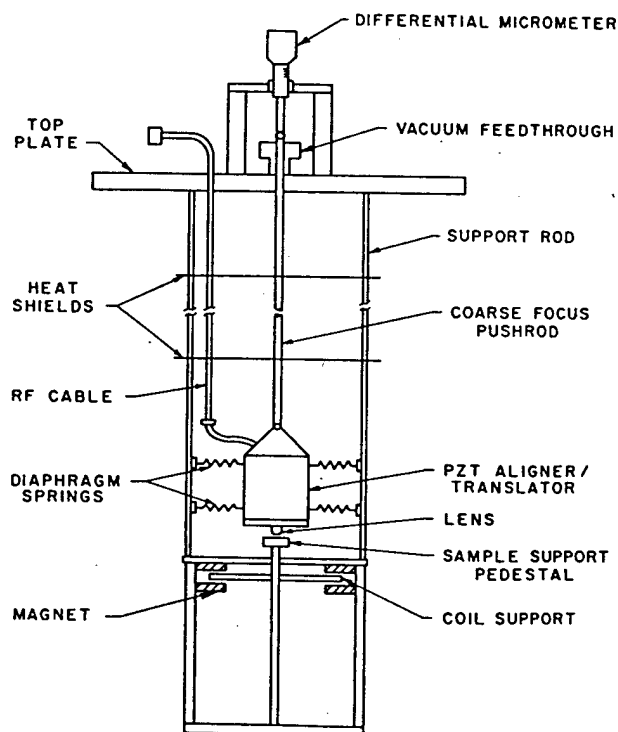

FIG. 2. Schematic of the assembly used in cryogenic liquids.

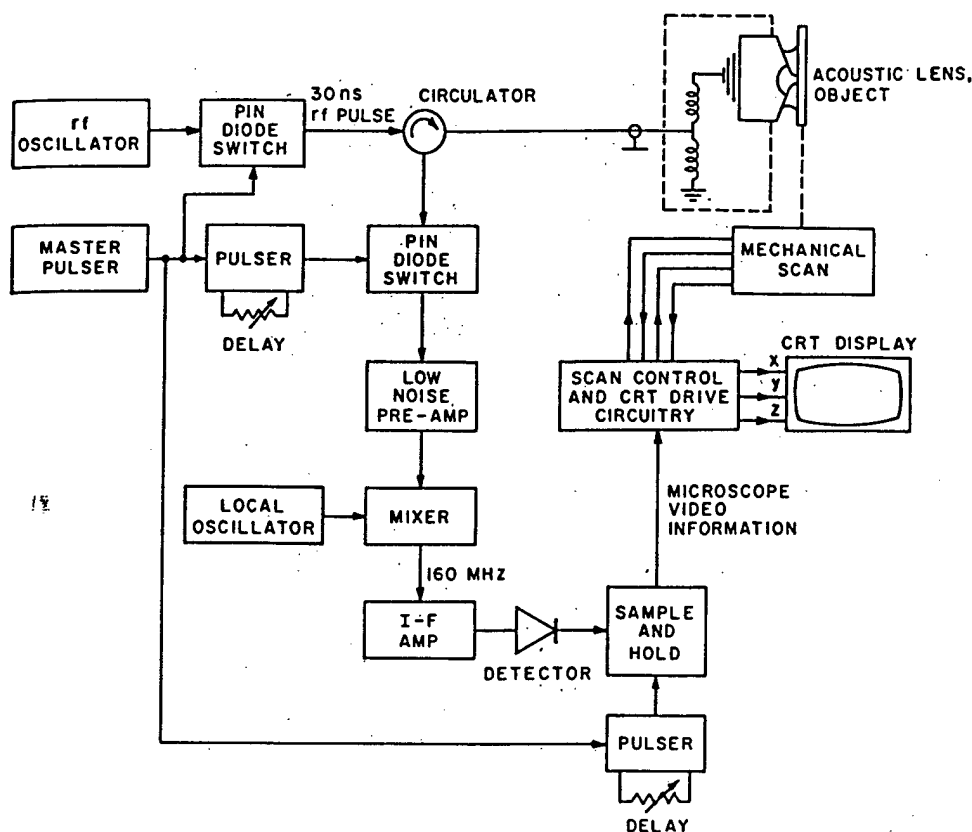Heiserman *et al.*: Cryogenic acoustic microscopy

FIG. 3. Block diagram of typical electronic components used in the cryogenic scanning acoustic microscope.

cryogen and acting in series with the coarse adjust. The fine adjust allows 1 μm total movement (at 2 °K) and about 100 Å position resolution. Electrical connections are made to feedthroughs in the top plate to allow for thermometry and control and monitoring signals for the PZT positioner and scanning stage. The rf drive and signal pulses are carried through the top plate on a stainless steel semirigid coax. Details of typical rf and control electronics are shown in Fig. 3.

An important part of the scanning acoustic microscope is the mechanism that mechanically translates the object through the acoustic beam in a raster pattern. A cross section of the mechanism used at cryogenic temperatures is shown schematically in Fig. 4. The sample is mounted horizontally in a sample holder on top of a flexible pillar made of a 9-cm length of small diameter aluminum tubing. Mounted below the sample holder are four small coils of wire spaced at right angles in a plane normal to the pillar; only two of the four coils are shown in Fig. 4. Situated about each coil is a set of stationary cobalt–samarium magnets. Two of the coils, an orthogonal pair, are used to drive the top of the flexible pillar and sample holder in the two free dimensions. The other two coils are used to detect the velocity of the sample holder. The velocity signals are used in a servo loop to improve the mechanical response of the system and to reduce the effect of any external vibrations coupled into the scanner. Velocity signals are also integrated to give the position of the sample. External vibrations might couple into the mechanical system of the scanner, but with this system the true position of the sample is always accurately

known. This position information is used to control the electron beam of the CRT display (or scan converter) so that the position of the beam is synchronized with the position of the sample. The fast axis of the raster scan is usually driven at a 30-Hz rate. The time required to scan one frame varies from 10 to 30 s.

The top of the flexible pillar travels in an arc rather than a plane and as a result the lens to sample spacing changes as the sample is scanned. The effect is small, however, because the pillar is relatively long and the scanned field is generally less than 250 μm on a side. The piezoelectric positioning element is sometimes
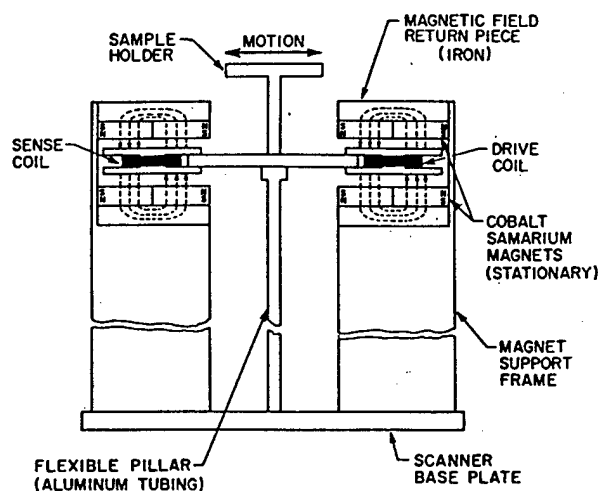

FIG. 4. Schematic of the mechanical scanning stage.

used to control the focal position of the lens to compensate for nonplanar motion of the sample.

Tests in liquid argon indicate that for our present levels of external vibration the scanner can execute a raster scan to within 0.2 μm of an ideal raster pattern. Furthermore, when the scan deviates from the ideal pattern, the position of the sample is at all times known with an uncertainty of less than 0.1 μm. In liquid helium vibrations from the vacuum pump used to maintain the helium bath at 1.95 °K degrade the scanner performance and image quality to some extent. We expect to be able to correct this in the future by further isolation of pump vibrations, increasing the stiffness of the flexible pillar, and improving the scanner electronics.

II. OPERATION IN LIQUID ARGON AND NITROGEN

A benchmark in our efforts to operate the acoustic microscope at cryogenic temperatures was established with liquid argon held at about 85 °K. This liquid was chosen based on its inertness, high figure of merit, and relatively large acoustic impedance. This last feature allowed us to use the same sapphire lenses as we have used in water. A single quarter-wave matching layer of glass provided a good power match. In liquid argon the cryogenic microscope was operated at 2 GHz where the wavelength was 0.43 μm, equal to wavelengths in the violet part of the optical spectrum. The high resolution of the argon microscope is demonstrated in Fig. 5.

Here a comparison is made between acoustic, scanning electron, and optical micrographs of a photoresist grating.[15] The grating was made by exposing a 0.15-μm layer of photoresist on a silicon wafer using two interfering optical beams. The photoresist lines are 0.2 μm wide and the center-to-center spacing between lines is 0.4 μm. The acoustic microscope clearly resolves the lines with a considerable amount of detail on the line edges. The contrast is excellent. After viewing acoustically the sample was coated for viewing in the scanning electron microscope [SEM, Fig. 5(b)]. The resolution is far better in the SEM image, but nevertheless the acoustic image shows a surprising amount of surface detail when compared to the SEM image. Figure 5(c) shows the same object viewed with an optical microscope fitted with a high power, dry objective. The change in the index of refraction between air and photoresist is small and it was also necessary to enhance the optical contrast by coating the sample with a metallic film.

We have examined several other objects in the argon microscope chosen from materials, integrated circuits, and biology. We were able to produce excellent images of a spread of human metaphase chromosomes which are natural objects with submicron structures.[16] Contrast in these objects was high and resolution was at least as good as high quality oil immersion optical micrographs. Figure 6 was chosen to illustrate the imaging performance with integrated circuits. The object here is a microwave dual gate field effect transistor (FET)[17] and the two linear features are the gate electrodes which are 1 μm wide. Figures 6(a) and 6(b) are acoustic micrographs taken at two
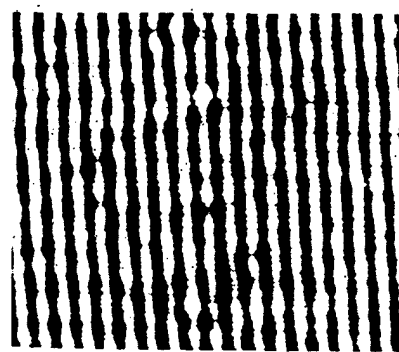
(a)

(b)
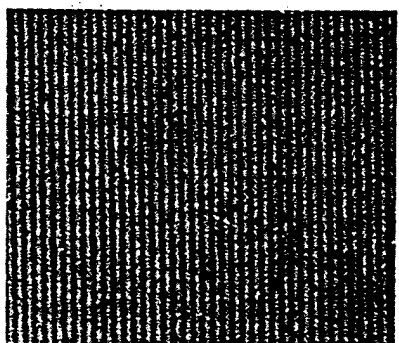
(c)

FIG. 5. Images of a photoresist grating: (a) acoustic micrograph in liquid argon, (b) scanning electron micrograph, and (c) optical micrograph using high power, dry objective. Line period is 0.4 μm. Images show different areas of the sample.

different focuses. Figure 6(c) is an optical image of the same device using a high power oil immersion objective.

We have also operated the microscope in liquid nitrogen at 77 °K using a drive frequency of 1.9 GHz. Reduction of frequency compared to argon was necessary due to the lower acoustic impedance of liquid nitrogen. As predicted in Table I, results were very similar to argon at a comparable frequency. Since liquid nitrogen is less expensive and more readily available than liquid argon, we will continue our investigations with this liquid.
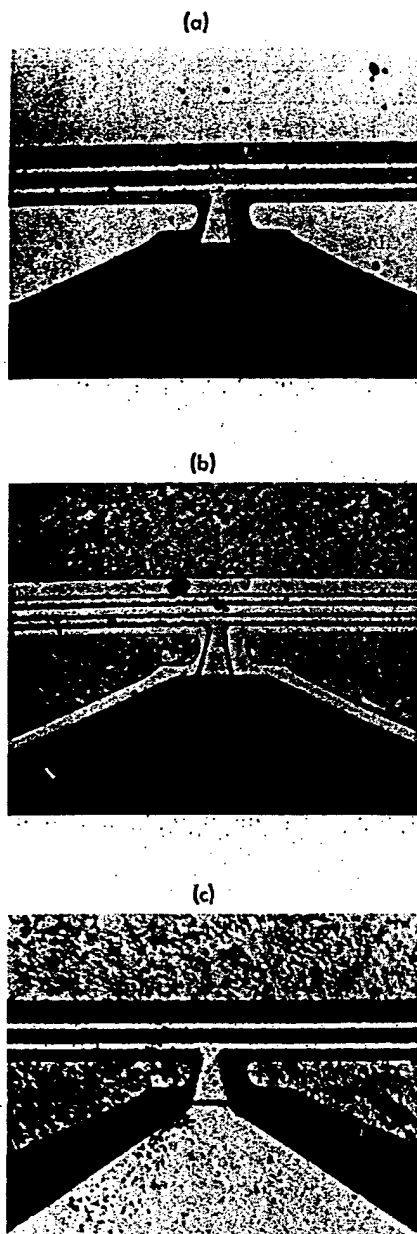
Heiserman *et al.*: Cryogenic acoustic microscopy

208

(a)

(b)

(c)

FIG. 6. Images of a high frequency field effect transistor. Parallel, horizontal lines are gate electrodes and are 1 μm wide: (a) and (b) acoustic micrographs at two different focuses, (c) oil immersion optical micrograph.

III. LIQUID HELIUM

Liquid helium is the most promising cryogen for use in acoustic microscopy and so we have designed our apparatus to be suitable for use in helium at temperatures down to about 1.1 °K.

The velocity of sound in liquid helium is nearly constant from absolute zero to temperatures near the lambda temperature, $T_\lambda = 2.17$ °K, but the attenuation varies considerably in this range. The attenuation α is plotted against T in Fig. 7. The data is from measurements by Imai and Rudnick[11] at 1 GHz and Abraham et al.,[14] extrapolated from 208 MHz. Above 2.5 °K the variation is understood in terms of classical mecha-
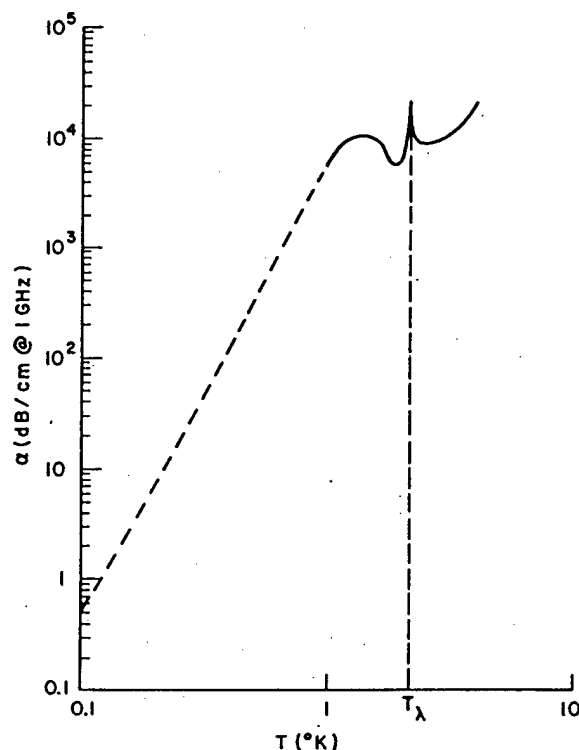
FIG. 7. Attenuation of sound in superfluid helium at 1 GHz as a function of temperature. High temperature (solid line) data from Ref. 11, low temperature (dashed line) data extrapolated from the data of Ref. 14 at 208 MHz.

nisms; viscosity and thermal conductivity contribute to this loss. Attenuation in this regime scales as frequency squared. The sharp peak at T_λ where helium becomes a superfluid is due to the presence of a second-order (lambda) phase transition at this temperature. The broad peak at 1.4 °K followed by a rapid falloff in attenuation at lower T is somewhat surprising. To understand these effects, we recall the behavior of the mean free path of thermal phonons in the liquid at these temperatures.

At temperatures near the peak, the phonon mean free path is determined mainly by scattering between phonons and rotons, the high momentum thermal excitation found in superfluid helium. Figure 8 shows the temperature dependence of the phonon mean free path characterizing viscosity from a calculation by Landau
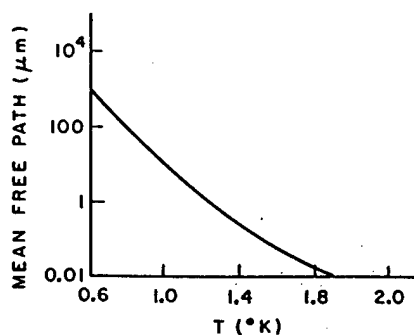

FIG. 8. The viscous mean free path of phonons in superfluid helium as a function of temperature. From data of Ref. 18.

and Khalatnikov.[18] From this it is evident that at the peak in attenuation the viscosity mean free path is approximately equal to the acoustic wavelength. At 1 GHz this is about 0.2 μm. The peak itself results from a relaxation process between the equilibrium populations of rotons and phonons.[19] Well above this temperature (the thermodynamic regime) we can assign values of pressure and temperature to regions small compared to an acoustic wavelength and thermodynamic arguments apply. However at very low temperatures (below 0.6 °K) the mean free path is large compared to an acoustic wavelength and the thermal phonons are not able to establish equilibrium during one acoustic period. We can no longer resort to an equilibrium thermodynamic description of the system. In this limit, known as the collisionless regime, we must consider the statistics of scattering of acoustical phonons by the gas of thermal phonons. From consideration of a three-phonon process a new dependence for the absorption at low temperatures can be obtained where (for $hf \ll kT$) α is found to increase linearly in frequency and according to the fourth power in temperature.[14] Data at 208 MHz supports this interpretation. No data exists at 1 GHz for these low temperatures.

Figure 7 suggests that for the purposes of acoustic microscopy temperatures below 0.5 °K will be very attractive. However, devising an acoustic microscope for use at these very low temperatures is a formidable problem and so as a preliminary step we have operated our present apparatus in liquid helium at 1.95 °K. At this temperature a local minimum in attenuation exists and we have been able to form acoustic images using wavelengths in helium of 0.36 μm, the shortest wavelength used for acoustic microscopy to date. The major problem encountered in operating in helium is the very low acoustic impedance of the liquid, $Z_{He} \sim 0.03 \times 10^5$ g/cm^2 s.

At the interface between a solid and a liquid, the reflection coefficient for acoustic waves is usually large. There is a large difference in acoustic impedance between the two materials. For example, at the interface between sapphire and water, 13% of the power normally incident in a plane wave is transmitted. At the boundary between a solid and liquid helium, the problem is much more severe. At a sapphire–helium boundary only 0.3% of the power is transmitted. Expressed in decibel notation, the transmission coefficient is −25 dB.

In general the power transmission coefficient (T) for transmission of normally incident plane waves from a solid of impedance Z_s to liquid helium can be written[20]

$$T = 4(Z_{He}/Z_s). \tag{4}$$

A standard approach for solving a mismatch problem with coherent waves is to place one or more quarter-wave matching layers between the solid and the liquid. Our problem is complicated in two ways. First, we operate at frequencies far above normal acoustic operating frequencies and our matching layers must be thin (~1 μm). These films are formed by vacuum deposition. This creates the second problem for there is

TABLE II. Properties of various quarter wave matching schemes.

| Lens material | Solid | 1 | 2 | He |
| | | $\lambda/4$ | $\lambda/4$ | |
		Layer 1	Layer 2	$-T$ (dB)
Sapphire (Al$_2$O$_3$)		$\cdots$	$\cdots$	25
Sapphire (Al$_2$O$_3$)		Glass	$\cdots$	14
Sapphire (Al$_2$O$_3$)		Carbon	$\cdots$	6
Sapphire (Al$_2$O$_3$)		Polyethylene	$\cdots$	2
Sapphire (Al$_2$O$_3$)		Au	Al	14
Sapphire (Al$_2$O$_3$)		Au	Glass	11
Sapphire (Al$_2$O$_3$)		Au	Carbon	3 [a]
Sapphire (Al$_2$O$_3$)		W	Glass	7
Sapphire (Al$_2$O$_3$)		W	Carbon	1 [a]
Fused quartz		$\cdots$	$\cdots$	20
Fused quartz		Glass	$\cdots$	19
Fused quartz		Carbon	$\cdots$	11
Fused quartz		Polyethylene	$\cdots$	4
Fused quartz		Au	Al	9
Fused quartz		Au	Glass	5
Fused quartz		Au	Carbon	2 [a]

[a]Based on a value of impedance of 6×10^5 g/cm^2 s as found experimentally in low frequency measurements in a bulk sample of pyrolytic carbon.

only a limited class of materials that can be deposited with vacuum techniques. We must choose from these materials.

The ideal matching layer would have an acoustic im-
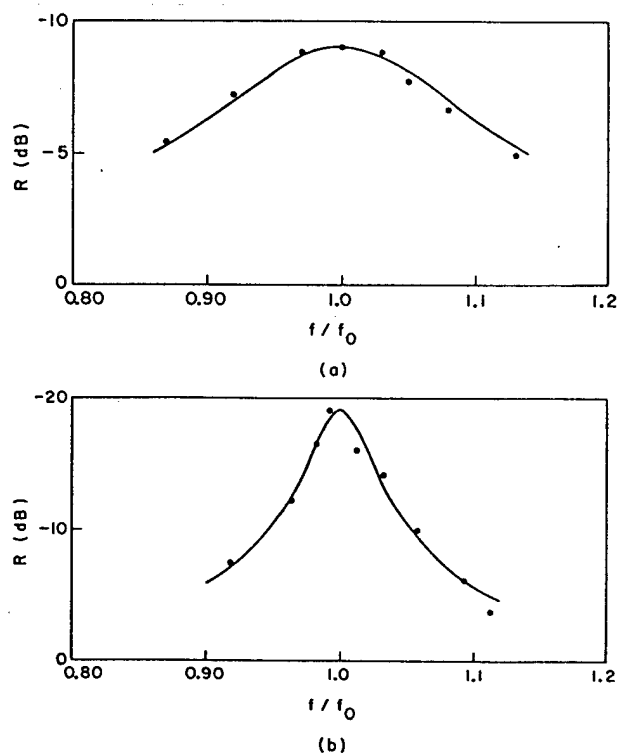
(a)

(b)

FIG. 9. Reflection coefficient of quarter-wave layers between sapphire and water. Points are experimental, curves are from Eq. (6). (a) A single glass layer. For the fit $Z_{glass} = 11.83 \times 10^5$ g/cm^2 s. (b) Gold–glass, with the glass deposited at the same time as the layer measured for (a). Here $Z_{gold} = 57.55 \times 10^5$ g/cm^2 s and Z_{glass} is taken from the fit of (a).

J. Acoust. Soc. Am., Vol. 67, No. 5, May 1980

210
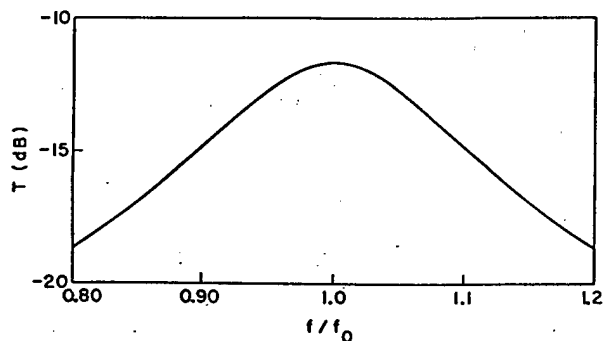
Heiserman et al.: Cryogenic acoustic microscopy

FIG. 10. Calculated transmission coefficient from sapphire to liquid helium using the gold glass quarter-wave matching layers of Fig. 9.

pedance equal to the geometric mean of the values for the solid and the liquid helium. For fused quartz this ideal value is 0.7×10^5 g/cm^2 s. For sapphire it is 1.2×10^5 g/cm^2 s. Unfortunately few solids possess such a low value of acoustic impedance.

It is possible, however, to add a second quarter-wave matching layer and improve the match at the expense

of bandwidth. The first layer is chosen to have a very high acoustic impedance. This effectively increases the impedance of the lens material and makes it possible to choose as a second layer a material of relatively high impedance and still produce a reasonable match into liquid helium. We can compare the different combinations with the following expression for the transmission coefficient. It is written in terms of the impedances of the lens material (Z_s), the first and second layers (Z_1 and Z_2) and the helium (Z_{He}) (see Ref. 20):

$$T = \left(\frac{4 Z_{He} Z_2^2 Z_s / Z_1^2}{(Z_2^2 Z_s / Z_1^2 + Z_{He})^2} \right). \tag{5}$$

This expression is valid for lossless isotropic media at the quarter-wave resonant frequency of the layers. For a single layer $Z_2 = Z_{He}$. In Table II we list a variety of possible matching schemes and compare their transmission efficiencies. We have experimented with several of these schemes and for our initial efforts have chosen a double layer composed of gold and glass on sapphire.

The properties of quarter-wave matching layers for use in helium were initially evaluated using water at
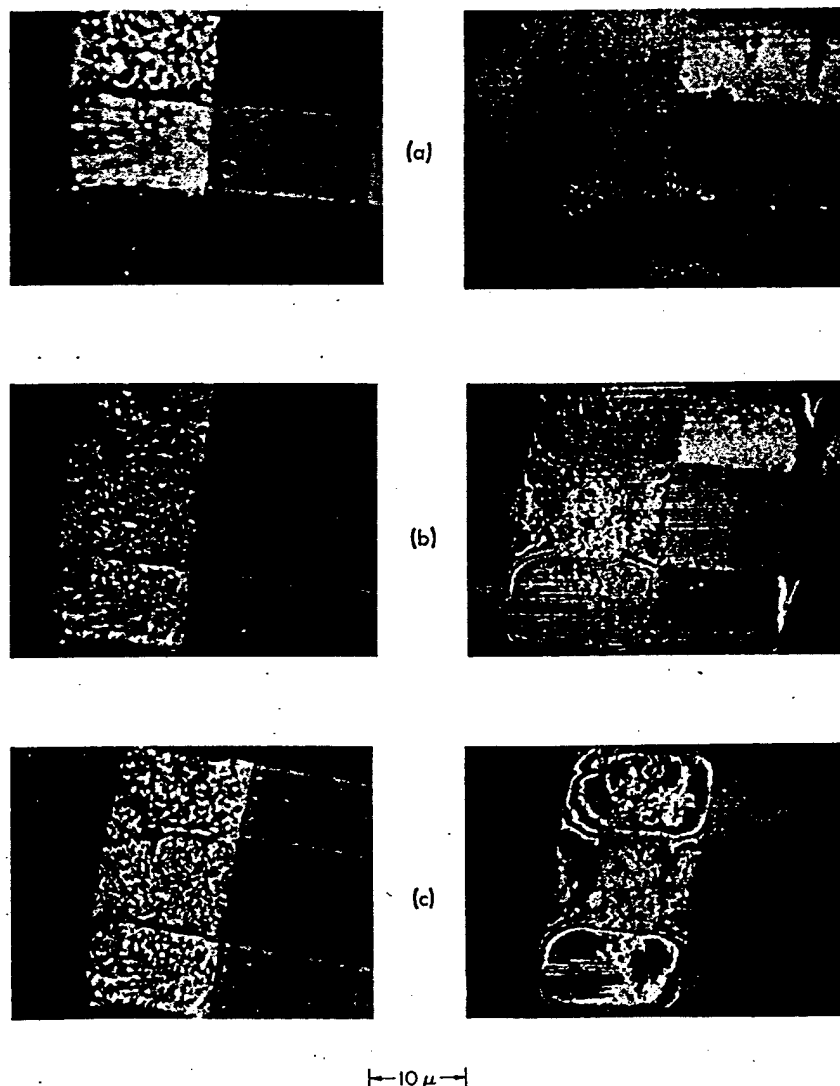

FIG. 11. Optical (left) and acoustic in superfluid helium (right) images of a silicon on sapphire integrated circuit. Acoustic wavelength is 0.36 μm. Lens to object distance increases from (a) to (c). Horizontal bars are polysilicon, 10 μm wide and 0.5 μm thick. Vertical bar is aluminum, 15 μm wide and 1.0 μm thick.

├─10 μ─┤

Heiserman *et al.*: Cryogenic acoustic microscopy

room temperature. Test "flats" were prepared by
depositing ZnO transducers on one end of a short
(1 mm) sapphire rod with the c axis aligned along the
rod axis. The ends of the rod were accurately parallel
and carefully polished. The films to be evaluated were
deposited on the opposite end. Acoustic pulses centered
at frequency f were generated in the rod and the
echo train produced by the multiple reflections ob-
served on an oscilloscope. The amplitude of a parti-
cular echo (normally the first) was measured with the
rod in contact with air (total reflection) and with a drop
of water applied to the matching layer end. The ratio
of these amplitudes squared (or the difference in deci-
bels) is the reflection coefficient of the test piece–
water interface at the frequency f. Measurements at
several frequencies near the quarter-wave resonant
frequency of the layers gives the frequency dependence
of the reflection coefficient. For the case of a single
layer the center frequency together with the measured
thickness of the layer yields a value for the longitudinal
velocity of sound in the layer (C) while the magnitude
of the reflection coefficient gives the acoustic impe-
dance. For two layers of thicknesses d_1 and d_2 the ex-
pected frequency dependence of the reflection coefficient
$R(f)$ can be written in the same approximation as Eq.
(5) as[20]

$$R(f) = \left| \frac{Z - Z_l}{Z + Z_l} \right|^2 , \tag{6}$$

where

$$Z = \frac{Z_s Z_1 - Z_s Z_2 \tan\phi_1 \tan\phi_2 - i(Z_1^2 \tan\phi_1 + Z_1 Z_2 \tan\phi_2)}{Z_1 Z_2 - Z_1^2 \tan\phi_1 \tan\phi_2 - i(Z_s Z_2 \tan\phi_1 + Z_s Z_1 \tan\phi_2)} \, Z_2 ,$$

$$\phi_1 = (2\pi f / C_1) d_1 ,$$

$$\phi_2 = (2\pi f / C_2) d_2 .$$

For two layers the impedance and velocity of the low
impedance, outer layer are separately evaluated. Then
Eq. (6) with $Z_2 = Z_l$ can be inverted to obtain Z_1. From
these measurements the expected transmission coeffi-
cient into liquid helium $T(f)$ can be calculated using
Eq. (6) with $Z_l = Z_{He}$ and recalling that $T(f) = 1 - R(f)$.

Figure 9 shows data taken on a gold glass matching
device evaluated using water. Figure 10 shows the pre-
dicted frequency dependences of the matching section in
liquid helium. At the center frequency, the transmis-
sion coefficient to helium is about 6%. This can be
improved upon using different materials as indicated in
Table II, but the gold–glass matching section was ade-
quate for initial work at 630 MHz with the temperature
held at 1.95 °K. At this temperature and frequency the
wavelength in liquid helium is 0.36 μm.

For our first object in liquid helium we choose a
silicon on sapphire integrated circuit. Optical and
acoustic micrographs are shown in Fig. 11. The verti-
cal bar is aluminum and is 1.0 μm thick. Successive
acoustic micrographs were taken at slightly different
focal positions. Figure 11(c) is the greatest lens to
object spacing and represents focus on the top of the
aluminum stripe. The image of Fig. 11(b) is slightly
closer focus and that of Fig. 11(a) is closest. Depth of
field and phase effects are evident in (c) where only

the thicker aluminum is visible. Due to the imperfect
match (even with the layers most of the acoustic power
is reflected at the sapphire–helium interface and re-
mains in the crystal) the information pulse interfered
with pulses in the sapphire rod resulting in the fringes
and contrast variations evident in the images. Such de-
fects will be eliminated by improving the match. Even
so features with sizes in the micron range are evident,
especially edge roughness apparent in Fig. 11(a).

IV. CONCLUSION

We have obtained preliminary results from an acous-
tic microscope operated in cryogenic liquids. Refine-
ment of these techniques will allow wavelengths well
beyond those of visible light to be employed for acous-
tic microscopy. The improved resolution thus obtained
will extend the range of sizes accessible to acoustic
microscopy to about 0.1 μm. We now have a plan for
constructing a microscope for use in superfluid helium
below 0.5 °K. There the operating frequency can be
increased and we expect a corresponding increase in the
resolving power.

Apart from the increase in resolution acoustic
microscopy at cryogenic temperatures will make
possible some novel observations. Potentially im-
portant applications lie in the study of solids at low
temperatures. We are particularly interested in the
study of the intermediate and mixed states of super-
conducting materials. Because the difference in acous-
tic attenuation between the normal and superconducting
states is typically large at high frequencies and tem-
peratures well below the transition temperature, it
should be possible to obtain sufficent contrast to di-
rectly image the lamellar structure of type I samples
and the vortex array found in type II materials. With
such observations we hope to study the time and mag-
netic field dependent behavior of these effects in detail.

[1] R. A. Lemons and C. F. Quate, Appl. Phys. Lett. 24, 163–165 (1974).

[2] V. Jipson and C. F. Quate, Appl. Phys. Lett. 32, 789–791 (1978).

[3] R. A. Lemons and C. F. Quate in *Physical Acoustics*, edited by W. P. Mason (Academic, New York, 1979), Vol. XIV.

[4] J. Attal and C. F. Quate, J. Acoust. Soc. Am. 59, 69–73 (1976).

[5] A. Atalar, J. Appl. Phys. 49, 5130–5139 (1979).

[6] M. Greenspan in *American Institute of Physics Handbook* (McGraw-Hill, New York, 1972), 3rd ed., Chap. 3e.

[7] B. G. Dudar and S. A. Mikhailenko, Sov. Phys.-Acoust. 22, 287–292 (1976).

[8] A. Van Itterbeck and W. Van Dael, Physica 28, 861–870 (1962).

[9] A. Pine, J. Chem. Phys. 51, 5171–5173 (1969).

[10] J. S. Imai, Doctoral dissertation, UCLA, 1969, Appendix B;

J. Acoust. Soc. Am., Vol. 67, No. 5, May 1980

Heiserman *et al.*: Cryogenic acoustic microscopy

212

D. G. Naugle and C. F. Squire, J. Chem. Phys. **42**, 3725 (1965); D. S. Swyt, J. F. Havlice, and E. F. Carome, *ibid.* **47**, 1199–1200 (1967).

[11]J. S. Imai and I. Rudnick, Phys. Rev. Lett. **22**, 694–697 (1969); M. A. Woolf, P. M. Platzman, and M. G. Cohen, *ibid.* **17**, 294–297 (1969).

[12]J. Heiserman, J. P. Hulin, J. Maynard, and I. Rudnick, Phys. Rev. B **14**, 3862–3867 (1976).

[13]J. Wilks, *Liquid and Solid Helium* (Oxford U. P., Oxford, 1967).

[14]B. M. Abraham, Y. Eckstein, J. B. Ketterson, M. Kuchnir, and J. Vignos, Phys. Rev. **181**, 347–373 (1969). For a recent review of theoretical developments see H. Maris, Rev. Mod. Phys. **49**, 341 (1977).

[15]Courtesy of IBM Corporation.

[16]D. Rugar, J. Heiserman, and C. F. Quate, J. Micros. (Oxford) (in press).

[17]Courtesy of Hewlett–Packard Corporation.

[18]L. D. Landau and I. M. Khalatnikov, Sov. Phys.–JETP **19**, 709–730 (1949). Translation: *Collected Papers of L. D. Landau*, edited by D. Ter Haar (Gordon and Breach, New York, 1965), p. 511.

[19]I. M. Khalatnikov and D. M. Chernikova, Sov. Phys.–JETP **22**, 1336–1346 (1966); **23**, 274–286 (1966).

[20]L. Brekovskikh, *Waves in Layered Media* (Academic, New York, 1960), Chap. 1. See also D. L. Folds and C. D. Loggins, J. Acoust. Soc. Am. **62**, 1102–1109 (1977) for comments on some of the derivations in this book.

J. Acoust. Soc. Am., Vol. 67, No. 5, May 1980   Heiserman *et al.*: Cryogenic acoustic microscopy

213

Ultrasound-Stimulated Vibro-Acoustic Spectrography

Mostafa Fatemi and James F. Greenleaf

An ultrasound method based on radiation force is presented for imaging the acoustic response of a material to mechanical excitation. Acoustic energy was emitted from solids and tissues in response to an oscillatory radiation force produced by interfering focused beams of ultrasound. Frequency spectra of ultrasound-stimulated acoustic emission exhibited object resonances. Raster-scanning the radiation force over the object and recording the amplitude and phase of the emitted sound resulted in data from which images related to the elastic compositions of the acoustically emitting objects could be computed. Acoustic emission signals distinguished tuning-fork resonances, submillimeter glass spheres, and calcification in excised arteries and detected object motions on the order of nanometers.

The mechanical response of objects to external forces is of considerable interest in medical diagnosis, nondestructive inspection of materials, and materials science. An applied force is often used to produce displacement from which elastic constants, like spring constants, can be determined. In resonant ultrasound spectroscopy, an ultrasound source and a detector are used to measure the resonance frequencies of a sample with known size and mass. The resonances are related to mechanical parameters, including the elastic constants of the material (1). Recently, a magnetic resonance elastography method for quantitatively measuring the displacement of tissues in response to externally applied cyclic forces was reported by Muthupillai et al. (2). The method resulted in high-resolution images of the shear modulus of normal and pathologic tissues. Others have used ultrasound to measure tissue displacement associated with externally applied compressive and cyclic forces (3).

We describe an imaging technique that uses acoustic emission to map the mechanical response of an object to local cyclic radiation forces produced by interfering ultrasound beams. Radiation force is generated by changes in the energy density of an acoustic field (4). For instance, a collimated ultrasound beam impinging normally on the surface of an object of arbitrary shape and boundary impedance will produce a radiation force. The radiation force arising from this interaction has a component $F = d_r s \langle E \rangle$ (5) in the beam direction. This component is proportional to the time-average energy density of the incident wave $\langle E \rangle$, the projected area of the object s, and d_r (6), the scattering and absorbing properties of the object.

We probe the object by arranging the intersection of two focused continuous-wave (CW) ultrasound beams of different frequencies at a selected point on the object. Interference in the intersection region of the two beams produces sinusoidal modulation of the ultrasound energy density. Modulation of the energy density creates an oscillatory force, effectively vibrating the object at the selected region. The resulting vibration of the object produces an acoustic field [acoustic emission (7)] that can be measured some distance away.

Ultrasound beams can be constructed in a variety of ways for this purpose (8). We used two coaxial, confocal transducer elements of a spherically focused annular array (consisting of a central disc and an outer annulus) driven by two CW signals at slightly different frequencies ω_1 and ω_2 (Fig. 1). The energy density at a point in this ultrasound field, say at the focus, is proportional to the square of the sum of the ultrasound fields from the two elements. Squaring the sum of two sines gives rise to sum and difference frequency terms. Thus,

high-frequency and low-frequency variations in energy density result at the intersection of the two beams produced by the two elements. Ultrasound-stimulated acoustic emission results from the energy term that produces a low-frequency vibration. The low-frequency force on a target at the focal point can be computed by

$$F_l(t) = d_r \iint_S \langle E_{\text{focal}}(t,x,y) \rangle_T \, dx \, dy$$

$$= C d_r \cos(\Delta \omega t) \qquad (1)$$

where C is a constant, $\Delta \omega = |\omega_1 - \omega_2|$, is the area over which $E_{\text{focal}}(t,x,y)$, the total energy density in the focal plane, has significant value, and $\langle \rangle_T$ represents a short-term time average (9). For focused beams the intersection region can be small enough that $F_l(t)$ can be considered to be an oscillating point force applied to the object at the focal intersection of the beams.
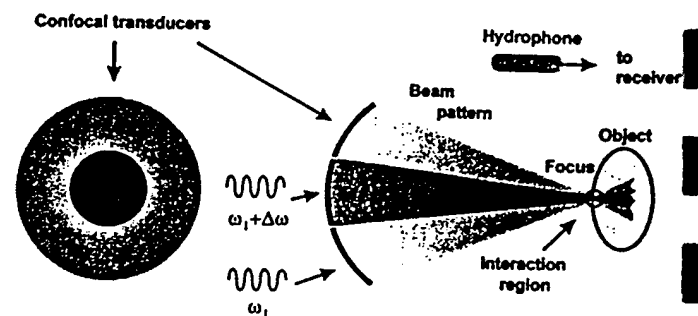
To produce an ultrasound-stimulated vibro-acoustic spectrogram, we vibrate a small region of the object with an oscillating radiation force of varying frequency. The complex amplitude of the resulting acoustic emission field is

$$\Phi(\Delta \omega) = C d_r H(\Delta \omega) Q(\Delta \omega) \qquad (2)$$

where $Q(\Delta \omega)$ is a complex function representing the mechanical frequency response, or admittance, of the object at the selected point, and $H(\Delta \omega)$ represents the combined frequency response, or transfer function, of the propagation medium and receiver and is assumed to be fixed and known (10). Recording $\Phi(\Delta \omega)$ allows us to obtain $Q(\Delta \omega)$ for each point within a constant multiplier (11). We raster-scan the radiation force over the object to produce data, which can

Fig. 1. Experimental system for ultrasound-stimulated vibro-acoustic spectrography: a two-element confocal ultrasound annular array transducer, consisting of a center disc and an outer ring. The elements are driven by two CW sources, at frequencies equal to ω_1 and $\omega_2 = \omega_1 + \Delta \omega$, where these frequencies are very close to the central frequency of the elements, and $\Delta \omega$ is much smaller than (<1%) the center frequency of the ultrasound transducer. The beams interact only in a small region around the joint focal point, where the amplitude of the field oscillates at the difference frequency $\Delta \omega$. The region of interest is placed at the joint focal point and is probed point-by-point by raster scanning. The sound field resulting from object vibrations at each position is received by a hydrophone and recorded. The recorded signal at one or more difference frequencies is used to form an image of the object. The experiments were conducted in a water tank. The transducer center frequency was 3 MHz; its outer diameter was 45 mm; and it was focused at 70 mm. The difference frequencies used in each experiment are mentioned in the corresponding legends.

Ultrasound Research, Department of Physiology and Biophysics, Mayo Clinic and Mayo Foundation, Rochester, MN 55905, USA.

mapped into a pictorial format. The spatial resolution of the resulting image is determined by the region in which significant interference between the ultrasound beams occurs and is of the order of a few wavelengths at the ultrasound frequency.

Experiments were conducted in a water tank, which provided good ultrasonic and acoustic coupling to the object and freedom of movement for the prototype scanner mechanism (Fig. 1). The two-element confocal ultrasound transducer array was positioned such that the beams interfered at the selected region of the object. Sound produced by vibrations of the object is approximately omnidirectional because of the small size of the vibrating portion of the object compared with the wavelength. This sound was detected by a submerged hydrophone placed near the object within the water tank.

To test the hypothesis that ultrasound-stimulated acoustic emission is sensitive to object mechanical properties and to show how such properties can be quantitatively evaluated by this method, we produced an ultrasound-stimulated vibro-acoustic spectrogram of a tuning fork immersed in isopropyl alcohol at two different temperatures. We aimed the focal point of the confocal transducer at a fixed position on one of the tines. The shear viscosity of alcohol changes with temperature, causing a slight, but detectable, shift in the spectrogram (Fig. 2). The shear viscosity η of a liquid is determined by measuring the resonant frequency f_R and the bandwidth δf_R of a tuning fork immersed in this liquid (12)

$$\eta = \frac{\kappa f_R}{\rho}\left(\frac{\delta f_R}{f_R} - \frac{\delta f_{R0}}{f_{R0}}\right)^2 \quad (3)$$

here f_{R0} and δf_{R0} are the resonant fre-

quency and bandwidth measured in vacuum, respectively, and ρ is the liquid density. The constant κ is determined experimentally. The measured values for f_R and δf_R were 1769 and 5.76 Hz at 15°C and 1765.7 and 4.53 Hz at 30°C. The viscosity of isopropyl alcohol is reported to be 2.89 cP (1 centipoise = 1 mPa·s) at 15°C (13). From this value, the constant κ was calculated. The shear viscosity at 30°C was found using Eq. 3 to be $\eta = 1.77$ cP, which is the same as the published data (13).

We tested the ability of ultrasonically stimulated vibro-acoustic spectrography to image the frequency response of different objects with identical d, by scanning three tuning forks with different resonant frequencies. A color acoustic spectrogram was obtained by sweeping the frequency of the radiation force, $\Delta\omega$, in a range covering the resonant frequencies of all forks at each beam position. The acoustic emission signal was filtered by three bandpass filters centered at different frequencies. The outputs of these filters were used to form a three-color composite image (Fig. 3). The forks appear

with three distinct colors because each fork responds primarily at its own resonant frequency. Because the forks were made from identical materials, other ultrasound imaging methods would not be capable of distinguishing these objects.

We tested the hypothesis that the amplitude of acoustic emission at a single frequency can be used to detect small, highly reflective isolated objects. We scanned a 380-μm-diameter glass bead placed on a thin latex sheet and recorded the amplitude of acoustic emission (Fig. 4). The latex sheet produces only a small change in the incident energy because it is almost transparent to the ultrasound beam. This experiment demonstrated the ability of the method to detect isolated regions of hardness with respect to a soft background.

To test the feasibility of using the technique to image mechanical properties of tissues, we measured the phase and amplitude of acoustic emission from calcified and noncalcified excised human iliac arteries. The arteries were scanned in a plane perpendicular to the ultrasound beam axis.

Fig. 3. Vibro-acoustic spectrogram of tuning forks. (A) Three tuning forks made from identical material with identical tine cross sections (lengths are different). Resonance frequencies in water are 407 Hz (right), 809 Hz (middle), and 1709 Hz (left). The forks were scanned in a water tank with the use of the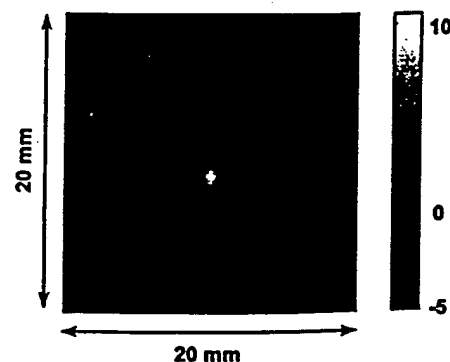 system shown in Fig.1. The scanning plane covers the front tines at the bottom part of the forks. At each position, the difference frequency was swept from 250 to 2250 Hz. The ultrasound-stimulated acoustic emission was detected with the hydrophone and filtered by three overlapping bandpass filters with frequencies centered at 500, 1000, and 1500 kHz, respectively. (B) Color acoustic spectrogram of the forks. The outputs of the bandpass filters were used to produce the red, green, and blue image components. This image displays two characteristics of the object: shape and frequency response. The color associated with each fork indicates its resonance frequency, which can be deduced from the frequency scale.
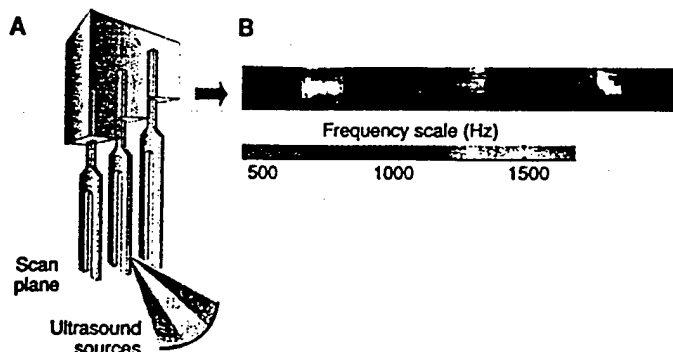
Ig. 2. Vibro-acoustic spectrograms of a tuning fork immersed in isopropyl alcohol at two different temperatures. A point on a tine of the fork was librated with the use of the system shown in Fig. . The difference frequency was swept from 1600 to 2000 Hz. A change of temperature from 15° to 30°C decreases the shear viscosity of the alcohol, which, in turn, changes the resonance frequency and bandwidth of the tuning fork.

Fig. 4. Image of ultrasonically stimulated acoustic emission amplitude from a 380-μm-diameter glass bead placed on a thin latex sheet. The latex surface was scanned at 0.2-mm increments in each direction. The difference frequency was fixed at 7.3 kHz. The amplitude of the acoustic emission of the bead (in relative units) is shown in gray scale. The latex sheet is almost transparent to the imaging system. The glass bead, however, presents a large acoustic impedance discontinuity, resulting in significant oscillatory radiation force. The magnitude of the radiation force gives rise to large-amplitude acoustic emission, thus yielding a high-contrast image of the bead. The image reveals a system resolution of about 700 μm. The confocal beam geometry leads to a negative-amplitude ring around the bright, positive central spot. This effect produces edge enhancement, as seen on artery walls in Fig. 5.

Calcifications within the arteries produced distinctive amplitude and phase values when compared to the normal arterial walls (Fig. 5). The phase of the oscillation of driven mechanical systems relative to the driving force depends on the ratio of mass to stiffness (14). Calcified regions of the diseased artery, identified by an x-ray of the sample, produced phase shifts in acoustic emission completely different from that of the noncalcified, and thus softer or less dense, regions. The amplitude images are highly detailed and exhibit variations in acoustic emission from both calcified and uncalcified regions of the diseased artery. These differences are caused by variations in the product of the reflection properties d_r and the effective mechanical vibration admittance properties $Q(\Delta\omega)$ of the tissue. Thus, vibro-acoustic spectrography is similar to conventional pulse-echo ultrasound imaging, which is sensitive to the ultrasonic parameters of the object, but has the advantage of also being sensitive to the mechanical admittance $Q(\Delta\omega)$ at low frequencies.

Motion induced by ultrasound and measured with ultrasound pulse echo has been used previously to study "hardness" (15). However, the sensitivity of ultrasound pulse
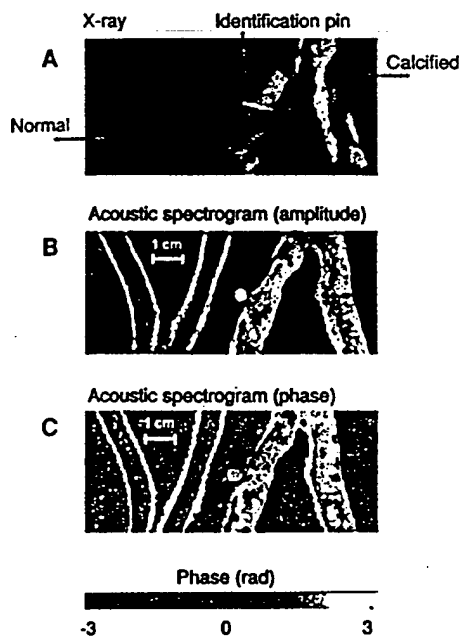
Fig. 5. Vibro-acoustic spectrography of excised human iliac arteries. (A) X-ray image of normal (left) and calcified (right) excised human iliac arteries obtained from a 35-year-old woman and a 67-year-old man, respectively. Bright areas indicate calcifications. (B) Vibro-acoustic spectrogram amplitude image at a fixed difference frequency of 6 kHz. Calcification details appear bright, whereas the arterial walls are dim. (C) Phase image. Calcified regions produce acoustic emission of different phase with respect to regions of the tissue having little calcification, as indicated by the x-ray.

echo to motion at common ultrasound frequencies is limited to several micrometers. The advantage of ultrasound-stimulated vibro-acoustic emission is its high displacement sensitivity. Cyclic displacement of 100 nm at 10 kHz produces an acoustic intensity of about 3.0×10^{-3} W/cm². Hydrophones similar to the one used in these experiments are sensitive to as little as 10^{-15} W/cm² and, therefore, can detect very small cyclic displacements. For example, the hydrophone detected an acoustic pressure of about 15×10^{-3} Pa at a distance of 5 cm from the glass bead shown in Fig. 4. Under the assumption of isotropic vibration, this pressure would be produced by a similarly sized sphere vibrating with a displacement amplitude of about 6 nm. The method will be more sensitive for higher frequency sound because acoustic power is proportional to the square of frequency for constant displacement amplitude. The practical upper limits for the difference frequency produced with modern ultrasound transducers is about equal to their bandwidth. For modern transducers, this value is 80% or more of the central frequency of the transducer. For experiments like those we conducted, emission frequency well in excess of 1 MHz could be produced. The lower limit on the frequency of radiation pressure is zero, that is, static pressure.

Ultrasound-stimulated vibro-acoustic spectrography has potential applications in at least two general areas. The first is nondestructive evaluation of materials, where material characteristics and structural flaws can be identified by measuring changes in the mechanical response to vibration at a point. The object under test could be remotely vibrated, for instance, by beams propagating and interfering in either water or air, or beams propagating within the object could be used to produce acoustic emission from flaws. For medical imaging and detection, the technique appears particularly suitable for noninvasive detection of hard tissue inclusions, such as the imaging of arteries with calcification, detection of breast microcalcifications, visualization of hard tumors, and detection of foreign objects. The stiffness of soft tissues is related to their composition (for example, relative values of fibrotic content), and its change is often related to pathology or therapy. In conventional palpation, physicians estimate tissue stiffness by feeling with the fingers. Because changes of stiffness alter the vibration frequency response or damping of tissue, the present method can potentially provide a noninvasive, remote, high-resolution "palpation" technique that can reach small abnormalities that are otherwise untouchable by conventional methods.

REFERENCES AND NOTES

1. J. Maynard, Phys. Today 49, 26 (January 1996).
2. R. Muthupillai et al., Science 269, 1854 (1995).
3. For a review of elasticity imaging methods, see L. Gao, K. J. Parker, R. M. Lerner, S. F. Levinson, Ultrasound Med. Biol. 22, 959 (1996).
4. The study of radiation force and radiation pressure dates back nearly one century, to the time of Rayleigh [Lord Rayleigh, Philos. Mag. 3, 338 (1902); ibid. 10, 364 (1905)]. More recent analysis of the theory and explanation of the physical significance of the mathematics can be found in G. R. Torr, Am. J. Phys. 52, 402 (1984). Critical and historical reviews on radiation force and radiation pressure are presented in B-T. Chu and R. E. Apfel, J. Acoust. Soc. Am. 72, 1673 (1982) and R. E. Beyer, ibid. 63, 1025 (1978). Some recent analysis of radiation force and pressure in attenuating medium (which may be applicable to biological tissues) are presented in O. V. Rudenko, A. P. Sarvazyan, S. Y. Emelianov, J. Acoust. Soc. Am. 99, 2791 (1996) and Z.-Y. Jiang and J. F. Greenleaf, ibid. 100, 741 (1996).
5. P. J. Westervelt, J. Acoust. Soc. Am. 23, 312 (1951).
6. The complex drag coefficient due to radiation pressure is defined for unit energy density of the incident wave and can be written as (5)

$$d_r = \frac{1}{s}(\Pi_a + \Pi_s - \int \gamma \cos\beta dS) - \frac{j}{s}\int \gamma \sin\beta dS \quad (4)$$

where Π_a and Π_s are the total absorbed and scattered powers, respectively; γ is the scattered intensity (π_a, π_s, and γ are expressed per unit incident intensity); β is the angle between the incident and scattered intensities; s is the projected area of the object; and dS is the area element. The real and imaginary parts represent the components of the force parallel and perpendicular to, respectively, the incident field momentum. In our treatment, we assume a planar object normal to the beam axis; hence, d_r is real, and the force has only a normal component to the object surface.
7. The term "acoustic emission" is used here to describe the acoustic field in response to a cyclic vibration of the object and should not be confused with similar terminology used in the field of nondestructive testing of materials or in opto-acoustic imaging context, where it is used to describe the acoustic field resulting from structural deformation, cracking, or thermal expansion of the object. We note a fundamental difference between the method we present and that of opto-acoustic imaging. In ultrasound-stimulated vibro-acoustic spectrography, the ultrasound energy is converted directly to low-frequency acoustic energy by the object, whereas the opto-acoustic method relies on the conversion of light energy to heat, causing acoustic emission in response to rapid thermal expansion of the object.
8. Modulation of a single beam with the use of a focused transducer driven by an amplitude-modulated signal results in a field that is not spatially confined, producing a radiation force on any object (including the transducer itself) that happens to be in the beam path. The use of two single-frequency beams is advantageous because field modulation occurs only in a confined region, the size of which is controlled by the intersection of the two beams.
9. We define the short-term time average of an arbitrary function of time f(t) around time instance t as

$$\langle f(t) \rangle_T = \frac{1}{T}\int_{-T/2}^{T/2} f(\tau - t)d\tau \quad (5)$$

The long-term time average is obtained when $T \to \infty$. To compute the short-term time average of the acoustic energy density relevant to acoustic emission at $\Delta\omega$, we choose T longer than the ultrasound wave period but much shorter than the acoustic wave period, that is, $2\pi/\omega_2 \ll T \ll 2\pi/\Delta\omega$.
10. In Eq. 2, $H(\Delta\omega)$ can be position-invariant if the geometry of the propagation medium remains unchanged during the scan. In our experiments, we minimized position dependency by fixing the position of the transducer relative to the hydrophone and moving, instead, the object in raster-scanning motion.
11. Changing $\Delta\omega$ by shifting the frequency of the ultrasound

beams does not change the beam amplitude because Δω is a small fraction of the bandwidth of typical ultrasound transducers. For instance, the bandwidth of the ultrasound transducer may be greater than 1 MHz, whereas Δω might be around 10 kHz.

12. M. R. Fisch, R. P. Moeller, E. F. Carome, J. Acoust. Soc. Am. 60, 623 (1976).
13. D. R. Lide, Ed., CRC Handbook of Chemistry and Physics (CRC Press, Boca Raton, FL, ed. 72,

1991), pp. 6–158.
14. P. M. Morse and K. U. Ingard, Theoretical Acoustics (McGraw-Hill, New York, 1968).
15. T. Sugimoto, S. Ueha, K. Itoh, in 1990 IEEE Ultrasonics Symposium Proceedings, B. R. McAvoy, Ed. (Institute of Electrical and Electronic Engineers, New York, 1990), vol. 3, p. 1377

Flower-Associated Brachycera Flies as Fossil Evidence for Jurassic Angiosperm Origins

Dong Ren

Pollinating insects played a decisive role in the origin and early evolution of the angiosperms. Pollinating orthorrhaphous Brachycera fossils (short-horned flies) collected from Late Jurassic rocks in Liaoning Province of northeast China provide evidence for a pre-Cretaceous origin of angiosperms. Functional morphology and comparison with modern confamilial taxa show that the orthorrhaphous Brachycera were some of the most ancient pollinators. These data thus imply that angiosperms originated during the Late Jurassic and were represented by at least two floral types.

The ancestors and time of appearance of angiosperms remain obscure (1–5). The earliest fossil evidence of nectar secretory tissue is provided by the Santonian-Campanian flowers from Sweden (6). The oldest angiosperm pollen grains have been found in Israel, in strata of Early Cretaceous time (Late Valanginian to Early Hauterivian) (7). The earliest recognized angiosperm inflorescences have been recovered from rocks of Late Hauterivian Age at Jixi, China (8).

The origin and early evolution of flowering plants are probably related to the coevolution of insect pollinators (9–11). Cretaceous and Tertiary flower-visiting insects were diverse and include an impressive variety of Coleoptera (beetles), Diptera (true flies), Lepidoptera (moths), Hymenoptera (wasps and bees), and other less diverse taxa, such as Thysanoptera (thrips). Some highly faithful pollinators such as butterflies and cyclorraphan flies appeared in the middle Tertiary (12). Few pre-Cretaceous pollinating insects are known. Small insects, especially flies and parasitoid wasps, may have been important then and thus in the origin and evolution of angiosperm pollination (13). Here I describe Late Jurassic pollinating orthorrhaphous Brachycera with well-preserved nectaring mouthparts.

Early pollinating insects have long tubular mouthparts designed for feeding on or extracting nectar from long tubular flowers (9–11). Other examples of Jurassic insects having this type of mouthpart include nemonychid weevils, which probably fed on bennettitaleans or cycads (14), and a monotrysian Lepidopteran with a siphonate proboscis (15, 16).

I collected the fossil Brachycera at a locality near Beipiao City, Liaoning Province, China, from nonmarine sedimentary rocks of the Yixian Formation (17). These rocks contain abundant remains of insects (18, 19), fishes, conchostracans, reptiles, birds, and mammals of Late Jurassic (approximately Tithonian) age (20).

Extant Brachycera comprise a wide variety of flower visitors (9, 10). Most orthorrhaphous Brachycera feed on flowers as adults. The new fossil orthorrhaphous Brachycera (19) include deer flies (Pangoniinae of Tabanidae), flower-loving flies (Apioceridae), and tangleveined flies (Nemestrinidae).

Most extant pangoniines are exclusively flower feeders (21). They often hover over flowers on the borders of dense vegetation (9, 10). Both males and females subsist on nectar and on the juice of flowers. The female proboscis of some species is flexible and suitable only for imbibation of nectar (22, 23), and is three or four times the length of the body. One of the Jurassic fossils, described as Palaepangonius eupterus Ren, 1998, includes a complete body and an associated well-developed long proboscis (Fig. 1) (19). These fossils provide direct evidence for the mid-Mesozoic diversification within Tabanidae of subclades with nectaring mouthparts. Palaepangonius not only provides evidence for the extraction of nectar from flowers or flowerlike structures but also demonstrates that the Pangoniinae have existed since the Late Jurassic. Another
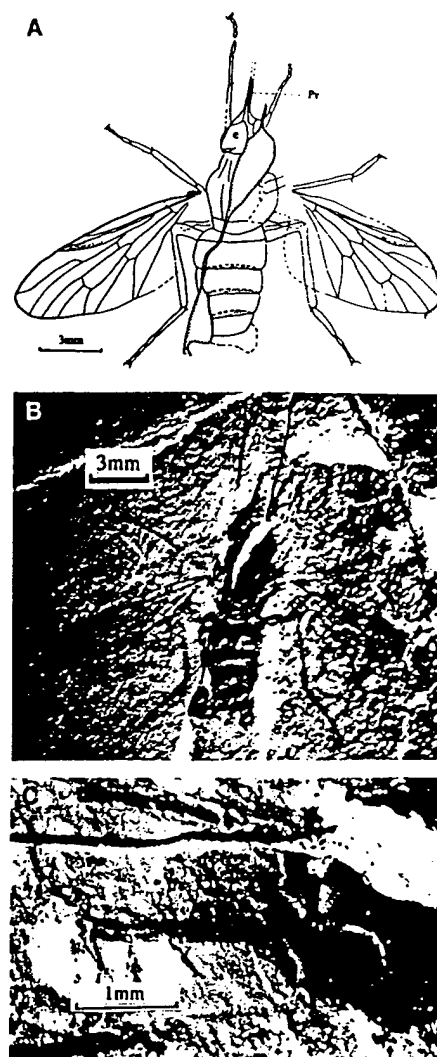
er brachyceran clade, the Nemestrinidae, are important pollinators of flowers (9, 10). Modern members are often collected when feeding on blossoms or hovering over them while imbibing nectar (24). Many Late Jurassic examples were collected and were described as Protonemestrius jurassicus Ren, 1998. These had a proboscis about 5.2 mm long, which would have been especially suitable for visiting long tubular flowers (Fig. 2). Florinemestrius pulcherrimus was also an important flower visitor. Its long stout proboscis seems to have been suited to extracting nectar from open or short tubular flowers (Fig. 3). Similar proboscides have been reported from the Late Jurassic of Karatau, Kazakhstan (25).

A representative of the stem group of the Apioceridae has also been found (Fig. 4) and called Protapiocera megista Ren, 1998. Its body bears dense hairs, a feature used in

Fig. 1. *Palaepangonius eupterus* Ren, 1998. (A) Camera lucida drawing of specimen LB97017. (B) Photograph of body, LB97017. (C) Photograph of proboscis, LB97017. Abbreviations: e, compound eye; Pr, proboscis.

National Geological Museum of China, Xisi, Beijing 100034, China.

PHOTO ACOUSTIC IMAGING

Photoacoustics on a microscopic scale

H. K. Wickramasinghe, R. C. Bray, V. Jipson, C. F. Quate, and J. R. Salcedo

Edward L. Ginzton Laboratory, Stanford University, Stanford, California 94305

We have modified a transmission acoustic microscope by replacing the input lens with an optical counterpart. The input to this system comes from a mode-locked and Q-switched Nd:YAG laser at 1.06 μm. Acoustic signals arising from thermoelastic coupling are generated in metallic films at frequencies determined by the modulation envelope of the optical beam. The output acoustic lens and transducer responsive to sound waves at 840 MHz detect the acoustic energy which comes from the heated volume (~2-μm diameter) near the focal region. The sample is mechanically scanned through the focal point in a raster pattern to record the photoacoustic images. We suggest this as a method for collecting new information on microscopic structures.

Materials with optical-absorption bands can be used to transfer energy from an optical beam to acoustic radiation—the photoacoustic effect. The change in temperature and the thermal expansion which accompanies the absorption of optical energy generates acoustic waves at frequencies corresponding to the amplitude modulation of the optical beam. A definitive experiment that has strongly influenced our thinking was carried out in 1971 by Kreuzer and Patel.[1] They used the acoustic energy generated by the photoacoustic effect to monitor trace impurities in gases. Since 1973, Rosencwaig[2] has carried out extensive photoacoustic work with solids, liquids, and biological materials. Several others[3] since that time have used photoacoustic spectroscopy (PAS) to study a variety of problems, and the results of all this has been summarized in two reviews by Maugh[4] and Farrow.[5]

The results to be reported here, where we discuss the feasibility of a photoacoustic microscope, follow more directly from a report by Brienza and De Maria.[6] In 1967 they demonstrated that mode-locked lasers with Q-switching could be used to generate intense sound beams through surface heating of metal films deposited on piezoelectric crystals. With picosecond pulse trains they were able to generate sound in solids at frequencies above 2 GHz. This was preceded by the early work of White,[7] who demonstrated in a classic paper that elec-

tromagnetic energy of various forms could be used to heat materials and generate acoustic waves. He predicted that this could be used as a tool for "... study of the high-frequency thermal properties of materials". Ghizoni *et al.*[8] have used periodic current flow to heat silicon and generate useful thermoelastic signals. Von Gutfeld and Melcher[9] considered this problem in 1977 and found that the strength of the acoustic signal could be increased by a large factor if the impedance of the material on either side of the absorbing layer was carefully controlled. They suggested that photoacoustic signals generated in this manner could be used in a system for imaging material defects. Wong[10] has used photoacoustic signals in a gas cell to image defects in ceramics of silicon nitride.

The imaging proposal of Von Gutfeld combined with photoacoustic spectroscopy provides the basis for a photoacoustic microscope. The first experimental results as reported here give us confidence in the feasibility of the system. Our previous experience with acoustics[11] indicates that the resolving power of this new instrument will be at least as good as that of the optical microscope. It will require microwave frequencies in the acoustic path, but we anticipate no difficulty on this point. In the analysis of thermoelastic signals,[7,12] one finds that in some cases—namely, in metal films where the thermal skin depth is greater than the
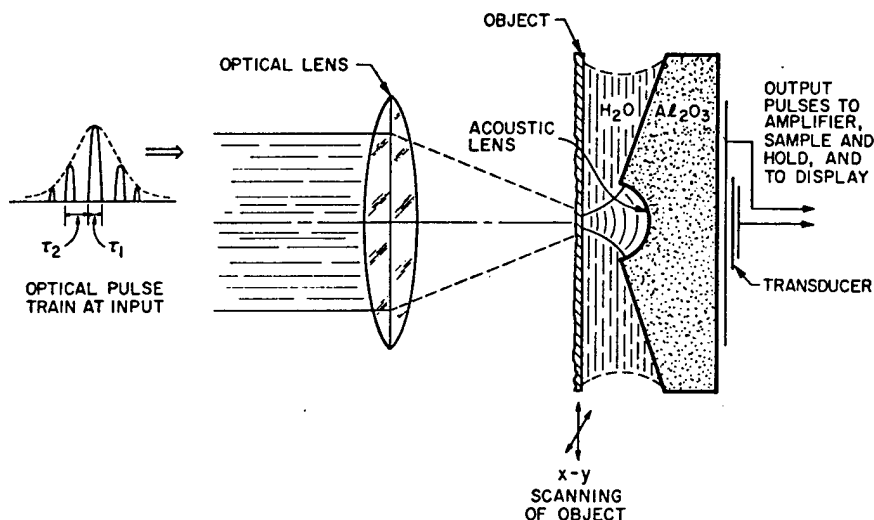

FIG. 1. Schematic of photoacoustic apparatus. The optical pulse train on the left at 1.06 μm has a peak power of 1 kW. The mode-locked pulses of width $\tau_1 = 200$ psec repeat at 210 MHz ($1/\tau_2$). The Q-switched pulses have a width of 200 nsec with a repetition frequency of 2.7 kHz. The acoustic lens has a radius of 200 μm and the transducer is tuned to 840 MHz.

→| |← 10 μm

(a) OPTICAL IN REFLECTION
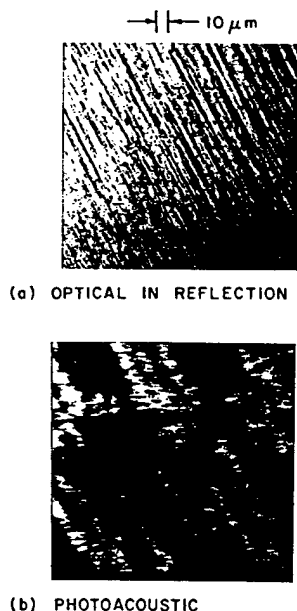

(b) PHOTOACOUSTIC

FIG. 2. A comparison of the optical and photoacoustic image for aluminum foil 25 μm thick.

skin depth of optical absorption and less than the sound wavelength—the conversion efficiency is directly proportional to the acoustic frequency. Cachier's experimental results at 800 MHz confirm this.

In this report we will present our initial results on this form of imaging. We have modified a conventional acoustic microscope[11] used in the transmission mode. The input acoustic lens has been replaced with an optical objective lens (NA = 0.25), and in this way the optical input beam is focused to a diameter of 2 μm. It is this focused optical beam, properly modulated, that is the source of our photoacoustic signal. The output lens and the transducer from the original acoustic microscope were used to detect the acoustic energy as radiated from the heated specimen. The schematic for the overall system is shown in Fig. 1.

In that diagram we see the incoming optical beam on the left. It emanates from a Nd : YAG laser mode locked and Q switched. The optical signal consists of packets 200 nsec in duration with a repetition frequency of 2.7 kHz. Within each packet we find the mode-locked pulse train. The pulses are 0.2 nsec in width (τ_1) and repeat at 210 MHz ($1/\tau_2$). The thermal response of the sample is fast enough to permit the generation of sound at frequencies corresponding to the nth harmonic of 210 MHz. Our experiments were performed at 840 MHz, although signals were observed at 1050 MHz. The acoustic lens used for receiving had a radius of 200 μm. It was coated with a quarter-wave layer of glass to minimize the acoustic reflection at the lens surface. The transducer was made of sputtered ZnO with a response centered near 800 MHz and a bandwidth of 100—200 MHz.

The first image from the photoacoustic arrangement is shown in Fig. 2. The object was a sheet of aluminum foil (25 μm thick), as shown in the optical image of Fig. 2(a). The lines that appear come from the rolling process used in producing the foil. The photoacoustic image is shown in Fig. 2(b). Although somewhat imperfect by optical standards, we do see the gross features attri-

butable to the rolling marks. For this image the peak power in the optical beam was limited to 1 kW in order to avoid sample damage. The signal-to-noise ratio at the output for the acoustic signal was 20 dB. The acoustic loss in the water path was estimated to be 32 dB.

In the second image of Fig. 3 we used a chromium pattern on a glass cover slip. The chrome was 200 nm thick and it was overcoated by 150 nm of aluminum. The chrome pattern was the hexagonal finder grid shown in the optical images of Figs. 3 (a) and 3 (b). Thus, the object alternates between a layer of aluminum directly on the glass and a layer of chrome overlaid with the aluminum. The contrast in the photoacoustic image of Fig. 3(c) comes from two factors: (1) the photoelastic constant for the aluminum-glass interface is different from that for the chrome-glass interface and (2) the acoustic impedance of the chrome-aluminum double layer is different from that of the single layer of aluminum; as a result, the amount of generated sound transmitted into the water differs for the two regions.

The ragged appearance in these first photoacoustic images of Figs. 2(b) and 3(c) results from improper synchronization between the motion of the electron beam on the monitor scope and the motion of the sample. We were using photodiodes for sensing the motion of the object—a carryover from the scanning acoustic microscope—and the laser interfered with the smooth operation of these sensors.

The theory for the configuration of Fig. 1 has not been worked out, but we do know something about the parameters that will come from this theory. The coupling coefficient relating the generated acoustic power to the square of the absorbed optical power is given by White.[7] We know that the photoelastic coupling constants are large in those materials with a small heat capacity, a

(a) OPTICAL (TRANSMISSION)   (b) OPTICAL (REFLECTION)
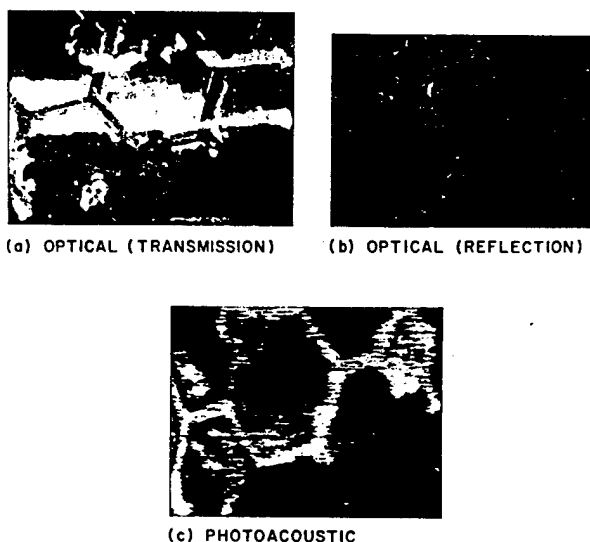

(c) PHOTOACOUSTIC

FIG. 3. A comparison of the optical and photoacoustic image of a metallized pattern deposited on a glass cover slip. The hexagonal grid pattern is formed with a 200-nm layer of chrome. The cover slip is then overlaid with a 150-nm film of aluminum. The bar width of the hexagonal grid is 25 μm. The quality of the photoacoustic image is degraded through improper synchronization of the CRT and the sample motion.

Appl. Phys. Lett., Vol. 33, No. 11, 1 December 1978

Wickramasinghe et al.

220

large expansion coefficient, and a high value for the thermal conductivity. Furthermore, we know that the thermoelastic effect in many metals is the major source of acoustic attenuation. Because of this similarity we expect high values for the photoacoustic coupling in those materials such as gold or silver, where the acoustic attenuation is large, and small photoacoustic coupling constants in materials such as nickel or aluminum, where the attenuation is small, but this is not the full story since we want to maximize the ratio of the acoustic power in the liquid cell to the incident optical power. In gold the optical reflectivity is high, so that very little of the incident energy (5%) is absorbed, and the acoustic impedance is high, so that very little of the generated acoustic power (10%) is transmitted into the liquid. A material such as aluminum with a lower value for both the reflectivity and acoustic impedance is a better choice, even though the photoelastic coupling is inferior to that for gold.

Although it is not demonstrated here, we plan in later work to scan the optical frequency and record the absorption spectra over the area of the sample. A suitable format for the output could be multiple images in pseudocolor. If all of this can be properly implemented, it would lead to a system for identifying the constituents of the specimen point by point across the imaging field.[13]

In summary, we claim to have demonstrated that a new form of microscopic examination is possible, and this is attained by combining modulated laser light with acoustic emission at high frequencies. The present numbers are interesting. They can be improved in several ways. The modulation of the laser beam in the form of picosecond pulses distributes the acoustic energy over a much wider bandwidth than that covered by the transducers. Sinusoidal modulation of the amplitude of the laser beam at 800 MHz would concentrate the optical energy into the acoustic band and improve the conversion efficiency by almost two orders of magnitude. Furthermore, we note that the sample did not have an antireflection coating, and much light was reflected rather than being absorbed. Finally, we note that the large mismatch between the metallic layers and the liquid reduced the available acoustic power. A proper coating for matching the acoustic energy into the liquid would correct this. We estimate the conversion efficiency from light to sound to be near 10^{-4}. This is consistent with the calculations of Cachier[12] if we consider the metal-glass interface to be clamped in the manner described by Von Gutfeld.[9]

[1] L. B. Kreuzer and C. K. N. Patel, Science **173**, 45 (1971); L. B. Kreuzer, J. Appl. Phys. **42**, 2935 (1971).
[2] A. Rosencwaig, Opt. Commun. **7**, 305 (1973); A. Rosencwaig, Phys. Today **28**(No. 1), 23 (1975).
[3] L. C. Aamodt, J. C. Murphy, and J. G. Parker, J. Appl. Phys. **48**, 927 (1977); F. A. McDonald and G. C. Wetsel, Jr., J. Appl. Phys. **49**, 2313 (1978).
[4] T. H. Maugh II, Science **188**, 38 (1975).
[5] M. K. Farrow, R. K. Burnham, M. Auzanneau, S. L. Olsen, N. Purdue, and E. M. Eyring, Appl. Opt. **17**, 1093 (1978).
[6] M. J. Brienza and A. J. De Maria, Appl. Phys. Lett. **11**, 44 (1967); Proc. IEEE **57**, 5 (1969).
[7] R. M. White, J. Appl. Phys. **35**, 3559 (1963).
[8] C. C. Ghizoni, M. A. Siqueira, H. Vargas, and L. C. M. Miranda, Appl. Phys. Lett. **32**, 554 (1978).
[9] R. J. Von Gutfeld and R. L. Melcher, Appl. Phys. Lett. **30**, 257 (1977).
[10] W. H. Wong, R. L. Thomas, and G. F. Hawkins, Appl. Phys. Lett. **32**, 538 (1978).
[11] R. A. Lemons and C. F. Quate, Science **188**, 905 (1975); V. Jipson and C. F. Quate, Appl. Phys. Lett. **32**, 789 (1978).
[12] G. Cachier, J. Acoust. Soc. Am. **49**, 974 (1971).
[13] It has been suggested to us by S. E. Harris that in addition to the optical-absorption spectra, the spontaneous Raman spectra could equally well be used in a system of this type to identify the molecular components.

Acoustic microscopy at optical wavelengths

V. Jipson and C. F. Quate

Edward L. Ginzton Laboratory, Stanford University, Stanford, California 94305
(Received 27 February 1978; accepted for publication 4 April 1978)

Recent advances in the field of acoustic microscopy have allowed the instrument to be operated at wavelengths that correspond to the center of the optical band. Experimental results in the form of acoustic micrographs are presented and compared to their optical counterparts. It is apparent that the resolving power of the instrument is similar to that of the optical microscope. Also it is seen that the acoustic micrographs yield information on the subsurface region. This information is not available in the optical images.

In the scanning acoustic microscope that was first introduced by Lemons and Quate,[1] the resolving power is determined by the beam diameter at the focal point. Heretofore it has been inferior to the resolution of the optical microscope.

We believe that this limitation has been overcome and we want to report on our recent work with an instrument operating at higher frequencies where the acoustic wavelength is equal to the wavelength of visible light.

The resolving power of the microscope is determined by the wavelength of the acoustic radiation in the liquid cell which fills the region between the lens and the specimen under study. In water heated to 60 °C the velocity of sound is 1.55 km/sec. Recently, we have been able to record acoustic micrographs near 3 GHz where the corresponding acoustic wavelength is 520 nm. This is a factor of 3 improvement over previously reported results and it is less than the optical wavelength of 550 nm which corresponds to the center of the optical band (i.e., green light). The resolving power should be comparable to an optical microscope since the lens that we are using is nearly free of aberrations.

The basic components of the instrument can be seen in Fig. 1. In this scanning version the object is translated point-by-point and line-by-line in a raster pattern past a focused diffraction-limited acoustic beam. The beam is brought into focus by a single surface spherical lens ground into a sapphire disk. The back surface of the sapphire contains a piezoelectric transducer which is centered on the axis of the lens. The front surface of the lens is coated with a glass layer—one-quarter wavelength in thickness. This serves as an antireflection coating between the mechanically stiff sapphire and the water.

The instrument used here is operated in the reflec-

tion mode with rf pulses 20 nsec in duration. The returning pulse from the object under study which contains all of the information is separated from the input pulse (and other spurious reflections) with the combined use of a circulator and time-gating circuits.

A single surface lens of this type can be used as the imaging element in the microscope since there is a large velocity change between the acoustic wave that propagates in the sapphire ($V = 11.1$ km/sec and the wave that propagates in the water ($V = 1.55$ km/sec). Beam distortion due to spherical aberration is inversely proportional to the square of this velocity ratio and for the acoustic case[2] it can be ignored. The large velocity ratio produces a correspondingly large angle of refraction at the lens-liquid interface. Because of this factor the acoustic energy leaves the lens close to the normal and converges to a focal point that is only 13% greater than the radius of the lens ($f = 1.13r$).

r_0 = RADIUS OF LENS = 40 μm
f = FOCAL LENGTH OF LENS = 1.13 r_0
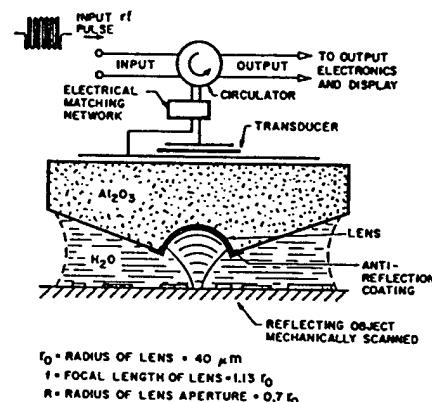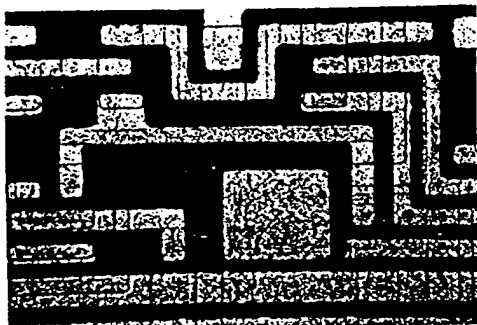R = RADIUS OF LENS APERTURE = 0.7 r_0

FIG. 1. The configuration of the scanning acoustic microscope as used in the reflection mode.

(a)

(b)

FIG. 2. Comparison of the optical (a) and acoustic (b) images of an integrated circuit. The aluminum line width is 7 μm.

image as recorded on a Zeiss instrument we see the surface features and contours with great clarity. In the acoustic image the aluminum lines have a very different appearance. The reflectivity of these lines is a strong function of the layering structure beneath the aluminum. The details of how this contrast comes about are contained in a previous publication.[4] If the subsurface layering is disturbed, there is a change in the acoustic reflectivity and this results in a change of the contrast. Thus the dark spots that appear on the aluminum lines at various points in Fig. 2(b) are attributed to subsurface structure. In some instances these spots correspond to regions where the oxide has been removed and emitter contacts have been formed.

Similar features can be seen in Fig. 3(b). This is another region of the silicon chip that is imaged with a higher magnification. Here again we can see mottled regions on the aluminum lines which are not evident in the optical image. These regions are also not discernible when the image is formed acoustically when the object at the focal point of the lens. The micrograph shown in Fig. 3(b) was taken at a slightly out of focus position to enhance this structure. This procedure, of course, results in a loss of resolution. We conclude from this comparison that these features are not on the surface. It is quite likely that they represent changes in depth profile of elastic materials beneath the surface.

The fundamental limit on the operating frequency, and therefore the resolving power of the instrument, is the acoustic absorption in the liquid cell over a path length d ($d = 2.26r$). The acoustic absorption for water is known to exhibit a square-law dependence over frequencies extending from a few kilohertz to well over 6 GHz. Because of this it is common to use the ratio α/f^2 to characterize the acoustic absorption in liquids. Here α is the attenuation coefficient (cm^{-1}) and f is the frequency (sec^{-1}). The value of α/f^2 for water heated to 60 °C is reduced by a factor of 2 over the room-temperature value. It is approximately 10^{-16} sec^2/cm and for a frequency of 3 GHz the associated acoustic loss is 7.8 $\times 10^3$ dB/cm. The magnitude of this value is the basic reason that we use the small lens radius as depicted in Fig. 1.

The possibility of using liquids other than water as the acoustic coupling media has been examined by Attal and Quate.[3] They pointed out that carbon disulfide, heavy metals and cryogenic liquids provide a substantial improvement over heated water but each of these liquids present some problems of specimen compatibility. Liquid media have not been extensively studied at microwave frequencies and it is still possible that improved liquids will be found with properties that will allow us to further reduce the wavelength.

The results of this work are shown in the micrographs of Figs. 2—4. In Fig. 2 we present the comparison between the acoustic and optical images for an IC circuit fabricated on silicon. The aluminum lines connecting to the transistors are 7 μm in width. In the optical

(a)

(b)

FIG. 3. Comparison of the optical (a) and acoustic (b) images of a different region of the same wafer as Fig. 2. Of interest are the dark blotches which appear on the aluminum lines in the acoustic micrograph but are not visible on the optical image. The acoustic micrograph was taken at a slightly out of focus position to enhance the subsurface detail.

V. Jipson and C.F. Quate

FIG. 4. Comparison of the optical (a) and acoustic (b) images of a dual-gate GaAs FET. The gate lines are 1 μm wide with a 1-μm spacing.

The final micrograph of Fig. 4 shows the comparison for an FET device as fabricated on gallium arsenide with a dual gate. This configuration is interesting to us since the two lines forming the gate are spaced by 1 μm and they are each 1 μm in width. It is included to demonstrate the resolving power of this instrument.

We believe that the evidence as presented in the comparison micrographs indicates that the acoustic microscope operating near 3 GHz can be used to resolve detail that is comparable in size to the features that appear in optical micrographs. In addition, we find that some regions which are inaccessible to optical radiation such as those beneath thin layers of metallization can be imaged with the acoustic instrument. Many of the defects in solid-state devices occur in underlying layers and it is our hope that the acoustic instrument can be used to characterize some of these defects without requiring that the surface be removed.

In summary the resolution of the acoustic microscope has been improved by approximately a factor of 3 over previous versions. This has been accomplished by (a) accurately centering the acoustic transducer behind the lens to ensure proper lens illumination, (b) utilizing quarter-wave matching layers to reduce interface losses, (c) heating the water path to further reduce the acoustic losses, and (d) reducing the lens size and, therefore, the acoustic attenuation in the liquid cell.

[1] R.A. Lemons and C.F. Quate, Appl. Phys. Lett. 24, 163 (1974).
[2] R.A. Lemons and C.F. Quate, 1974 *Ultrasonics Symposium Proceedings*, IEEE Cat. 73CHO807-SU (IEEE, New York, 1974), Paper E-6, p. 18.
[3] J. Attal and C.F. Quate, J. Acoust. Soc. Am. 59, 69 (1976).
[4] A. Atalar, C.F. Quate, and H.K. Wickramasinghe, Appl. Phys. Lett. 31, 791 (1977).

Appl. Phys. Lett., Vol. 32, No. 12, 15 June 1978
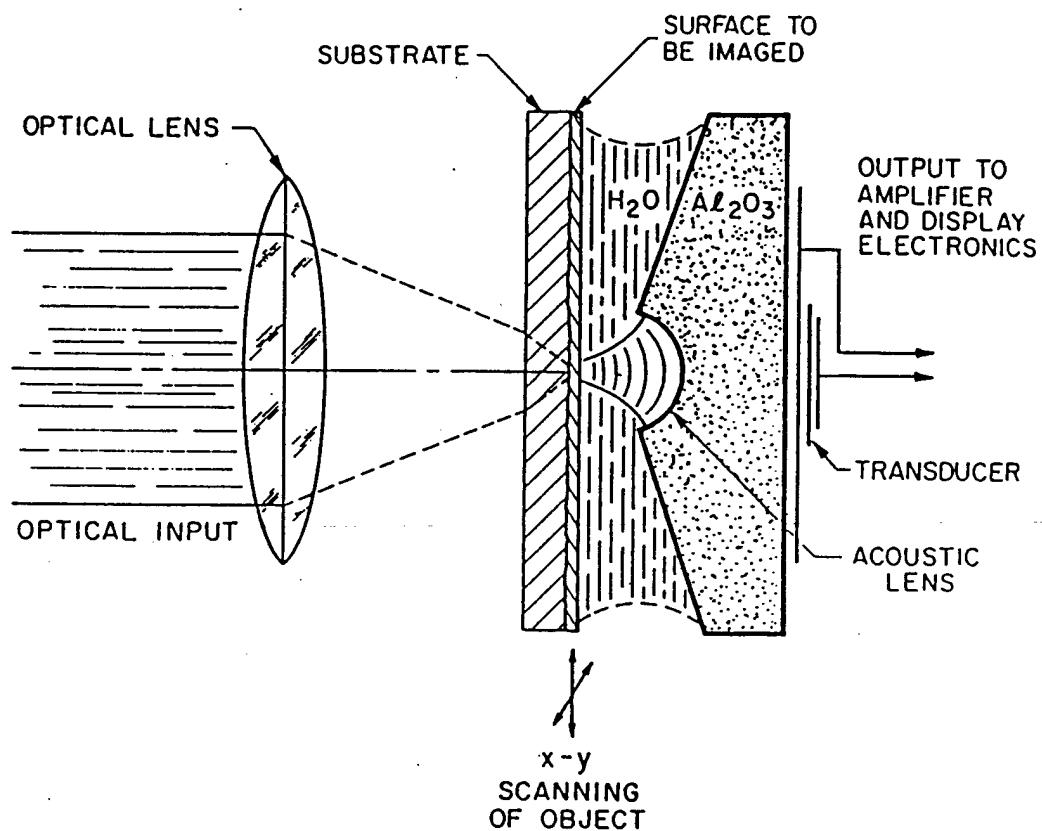
V. Jipson and C.F. Quate

224

Figure 4-1. The experimental arrangement for photoacoustic imaging of thin films on optically transparent substrates using an acoustic lens to collect the sound generated by the modulated optical power.
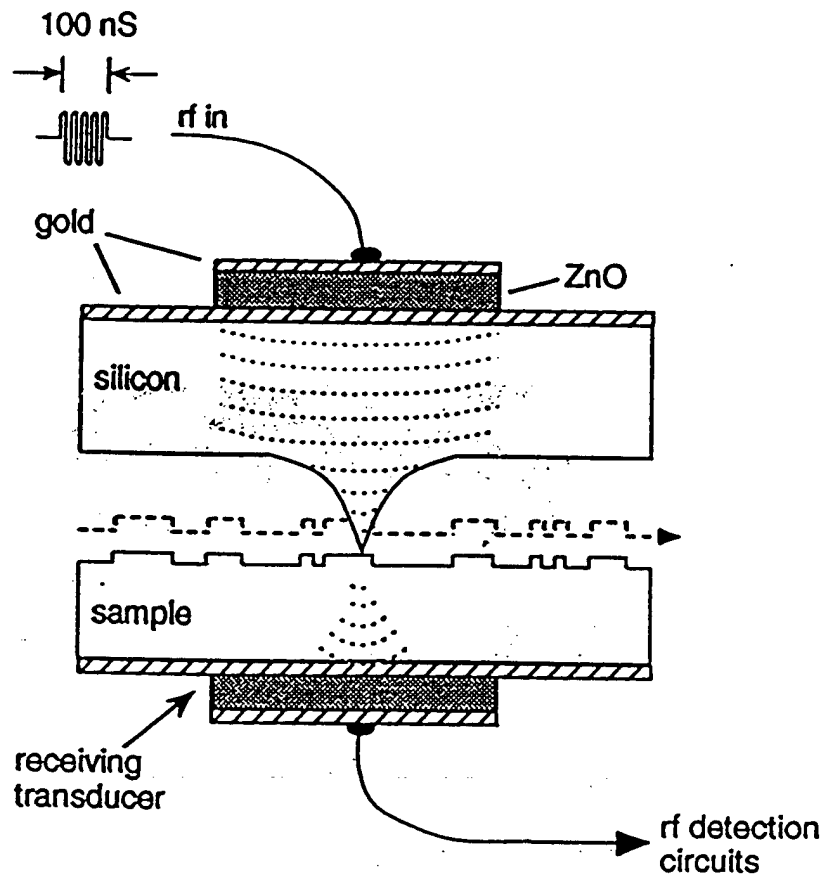
225

100 nS

rf in

gold

ZnO

silicon

sample

receiving
transducer

rf detection
circuits

Figure 5.1 In transmission mode NSAM, acoustic waves are generated by the tip transducer, coupled to the sample through the tip and detected by a second transducer on the back of the sample.

From the tip transducer up to the point in the tip where the cross-section is approximately equal to $\lambda/2$, the loss is proportional to the ratio of the tip cross-section at $x = \lambda/2$ and the area of the transducer. Since the acoustic impedance mismatch of the silicon tip to air is 19.7×10^6 to 0 kg/m^{2}s, virtually nothing is transmitted from the silicon into the air. In this region, the power density is constant along the longitudinal axis of the tip.

(a) PHOTOACOUSTIC

(b) ACOUSTIC (IN FOCUS)

(c) OPTICAL

Figure 4-7. A comparison of photoacoustic, acoustic and optical images of a gold film on a glass substrate.

227
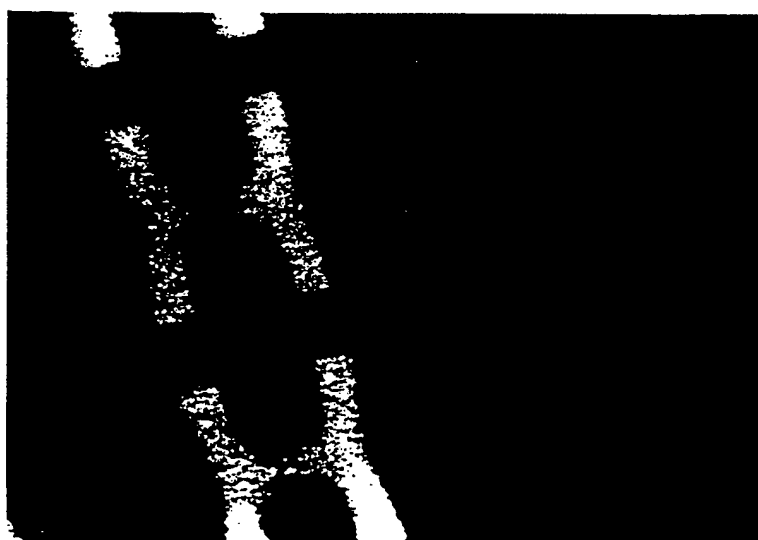
Figure 5-8. (a) Photothermal image of boron implanted silicon. The darker area is the implanted region, the lighter region is not implanted. (b) Photothermal image of implanted silicon with 300 Å of titanium deposited on the silicon in the lelft half of the image.
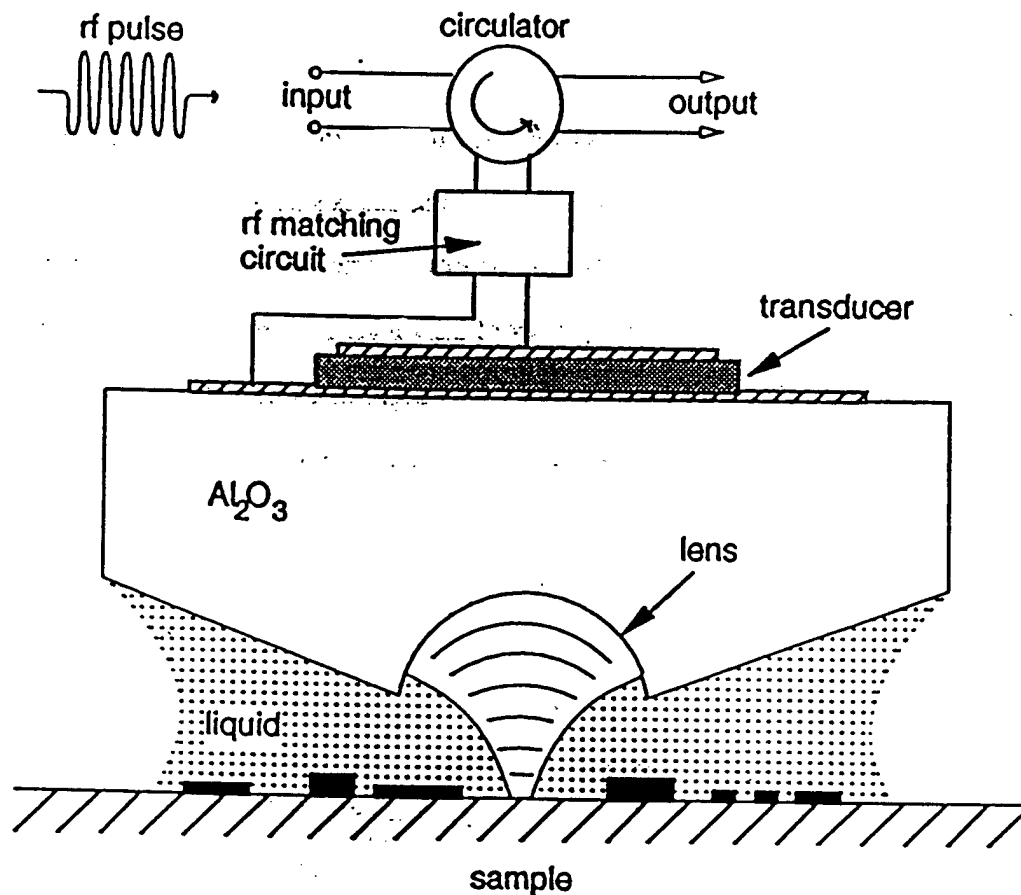
NEAR FIELD ACOUSTIC

Figure 4.1 The far-field scanning acoustic microscope utilizes a hemispherical acoustic lens to focus acoustic waves in a liquid to a diffraction limited spot. (After Quate et al. [2].)

Since it is a far-field instrument, the resolution of the scanning acoustic microscope is limited by the spot size. In order to reduce the spot size the instrument can be operated at higher frequencies to reduce the acoustic wavelength; however, the frequency can not be increased without bound since the attenuation of sound in most liquids is proportional to the square of the frequency. Hence, if the frequency is increased, the acoustic path length must be reduced by a corresponding factor in order to maintain acoustic signal power

4.2.1 Resonance NSAM

There are two major types of NSAM instruments. The first type, exemplified by acoustic microscope of Güthner and coworkers, utilizes some form of an acoustic resonator (Figure 4.2) [11]. In this case, one arm of a quartz tuning fork acts as a probe. The tuning fork is driven at its resonance frequency of 32.7 KHz and the interaction of the tuning fork with the sample is monitored by measuring changes in the resonance frequency or vibration amplitude of the tuning fork. This configuration has yielded images of insulators with lateral resolution on the order of 1 μm.
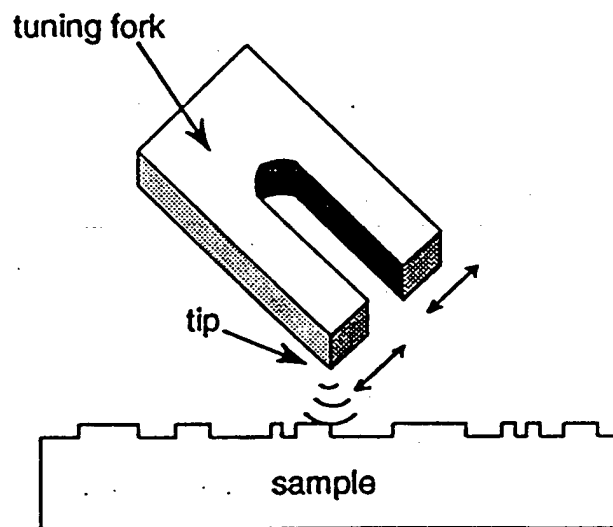
Figure 4.2 Güthner and coworkers use one leg of a quartz oscillator as a tip to probe a sample acoustically. The image is obtained by monitoring the amplitude of the oscillator's vibration.

Among several benefits of the resonator-type instrument is that when forces between the sample and the tip are sufficiently large, the vibration of the resonator stops.
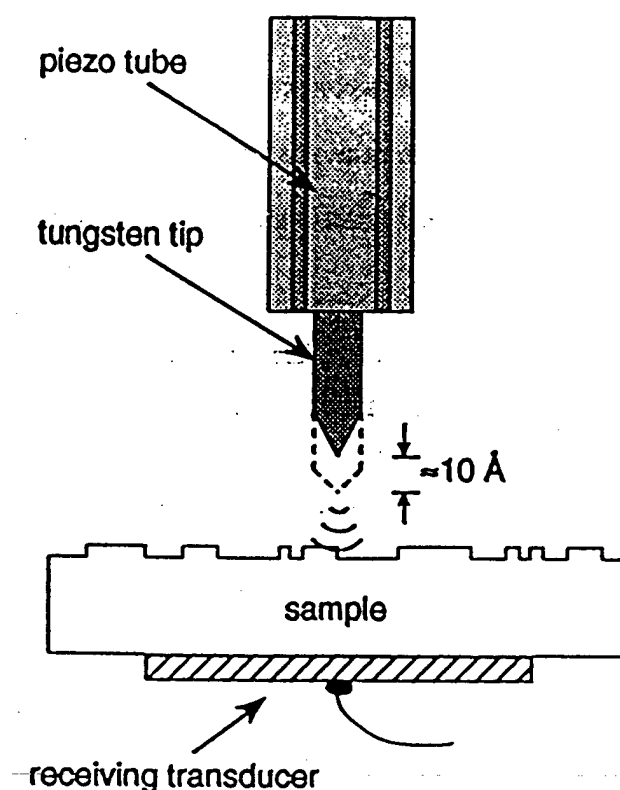
Figure 4.3 Transmission mode NSAM has been demonstrated by several researchers using modified STMs. Takata and coworkers modulate the z piezo of a standard tube scanner to excite acoustic waves into the sample.

Advantages of two-port NSAM instruments include the fact that since the tip need not be an integral part of an acoustically resonator, common tungsten or platinum/iridium tips similar to those used in STMs can be used. If the tip is damaged during use, it can be replaced easily and quickly. Damage to the tip is of particular concern for non-resonator-type instruments since there is no inherent sign, in the absence of a tunneling current signal, when the tip and sample come into contact.

In the resonator-type instruments, when the tip and sample come into contact, the resonator stops resonating. In transmission instruments however, if the tip were to make

were not designed specifically for this project and therefore the tips and transducers were not aligned to each other across the entire wafer. For expediency, the substrates were each diced and individual tip dice were bonded to individual transducer dice.
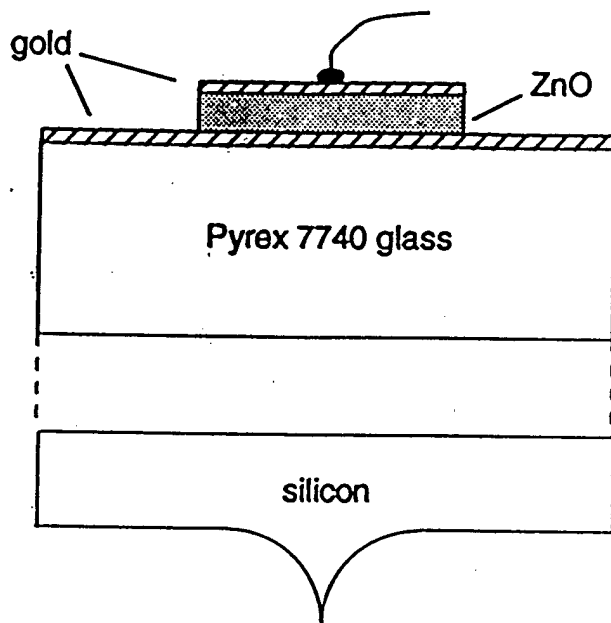

Figure 4.11 In the hybrid process, the tip is fabricated on a silicon wafer while the acoustic transducer is fabricated on a 7740 Pyrex glass substrate. To form the probe, the tip and transducer substrates are anodically bonded together.

The bonding of the tip and transducer substrates was accomplished using a low temperature anodic bonding process [18]. To begin the process, the die with the tip is placed in a special jig that holds the die with the tip facing downward into a cavity. The Pyrex die with the transducer is placed on top of the tip die and aligned under a direct view optical aligner which has coaxial optical objectives to view both sides of the sample. The alignment is accomplished by hand using tweezers to adjust the position of the dice. Approximately 25 μm accuracy in alignment is achievable using this technique. When the
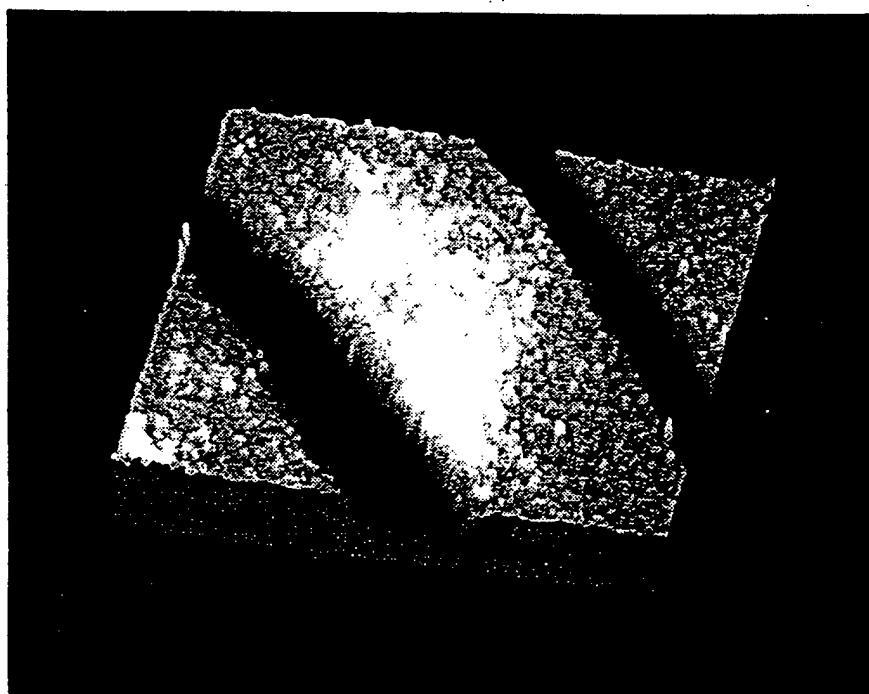
Figure 6.5 The scan direction was rotated to yield this 3-dimensional transmission NSAM image of a 6.5 μm period silicon grating with 940 Å step height. The image was taken using a tetrahedral tip and an acoustic frequency of 135 MHz.

The scan angle was rotated to produce the three-dimensionally rendered image of the same grating shown in Figure 6.5 These images were taken using 135 MHz acoustic

234

Acoustic microscopy by atomic force microscopy

U. Rabe and W. Arnold

Fraunhofer Institute for Nondestructive Testing, Bldg. 37, University, D-66123 Saarbrücken, Germany

We have constructed an atomic force microscope enabling one to image the topography of a sample, and to monitor simultaneously ultrasonic surface vibrations in the MHz range. For detection of the distribution of the ultrasonic vibration amplitude, a part of the position-sensing light beam reflected from the cantilever is directed to an external knife-edge detector. Acoustic images taken on the surface of a wafer show a lateral resolution of about 100 nm at an ultrasonic frequency of 20 MHz.

High-resolution acoustic imaging is a powerful tool for materials investigation.[1,2] By using such techniques, elastic properties and defects in materials can be determined with a resolution given by the wavelength. Various optical schemes have been developed in order to detect the displacements of surface, longitudinal, and shear acoustic waves,[3] either for local analysis or combined with a scanning technique such as scanning laser acoustic microscopy.[4] But due to Abbe's principle a lateral resolution not better than about a wavelength is obtained in techniques where focused beams are used. This limit can only be surmounted by using near-field techniques. Recently, it has been successfully demonstrated that the high lateral resolution of near-field techniques can be exploited for the detection of acoustic waves by using a scanning tunneling microscope (STM),[5–9] and even images have been obtained,[9,10] but up to now detection of ultrasound by atomic force microscopy (AFM) has not been reported. Vibrating the sample at frequencies below the resonance frequency of the cantilever is used for elasticity mapping with AFM,[11] and is a standard technique in noncontact scanning force microscopy.[12] With the technique used in our experiment, however, vibrations at ultrasonic frequencies (5–20 MHz) are detected by an AFM well above the cantilever resonance frequency of several kHz.

A sample to be examined is attached by a coupling medium onto a delay line of an ultrasonic transducer (20 MHz) which is fixed on top of the AFM scanner (Fig. 1). Longitudinal ultrasonic waves are generated which are reflected at the sample surface causing vertical displacements. The AFM (Nanoscope III, manufactured by Digital Instruments, Santa Barbara, CA) works in contact mode with a beam deflection position sensor. The cantilever is a standard Si_3N_4 microfabricated triangular cantilever with pyramidal tip of approximately a 50-nm radius. We modified the instrument; by an optical beam splitter, half of the intensity of the light beam reflected from the cantilever is coupled to an external knife-edge detector using a fast photodetector (Si-*pin* diode, rise time 1 ns), while the undeviated part is used for topography imaging. Knife-edge detection is well known for optical detection of surface acoustic waves.[13] Other types of sensors which allow an absolute calibration of the cantilever vibration amplitude, like interferometers or capacitive detectors could also be used.

The A scan after 60-dB amplification (Fig. 2) shows first a small signal (1) due to direct electrical pick-up of the exciting spike (400 V, 15-ns rise time). The first ultrasonic signal (2) follows after a delay given by the delay line, the sample, and the tip of the cantilever. It is followed by a series of echoes caused by multiple reflections and mode conversion. Signal-to-noise ratio is 22 dB in a detection bandwidth of 27 MHz. To evaluate the amplitude of the first ultrasonic signal, it is fed into a boxcar integrator (Fig. 3) yielding a voltage proportional to the amplitude of the selected pulse. A gate from a pulse generator was used and was delayed such that it coincides in time and duration with this first ultrasonic signal [(2), Fig. 2]. The integrated signal is stored in the second input channel of the microscope electronics. It would also be possible to digitize the ultrasonic signal and display the peak value or any other suitable signal.

The experimental procedure is as follows: in absence of tip–sample interaction, the laser diode beam is centered onto the cantilever. The position sensor and the knife-edge detector are adjusted to their most sensitive working point. The sample is approached to the tip and the feedback loop is activated. Then an image is taken. The feedback loop maintains a constant low interaction force during the scan, also keeping the working point of the external knife-edge detector constant. A special vibration isolation is not necessary because mechanical vibrations are well below the 3-dB bandwidth of the knife-edge detector (1–28 MHz).

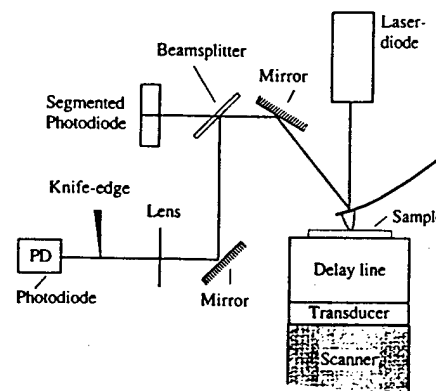Figure 4(a) shows an unfiltered image of the ultrasonic

FIG. 1. Modified AFM for detection of ultrasonic vibrations. By the beam splitter half of the intensity of the light beam reflected from the cantilever is directed via a mirror to an external knife-edge detector with a photodiode (PD) which is used for detection of the ultrasonic signal.
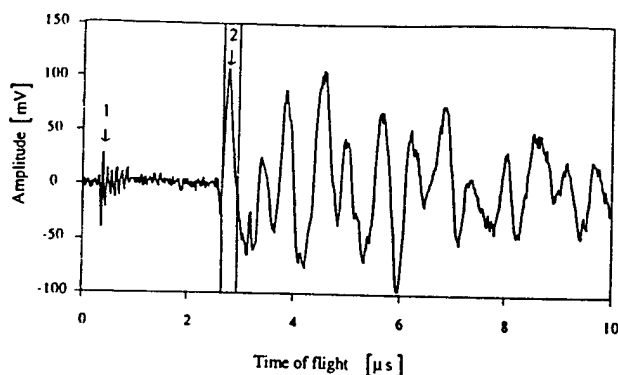
FIG. 2. Signal obtained by the cantilever knife-edge detector. After a delay time an ultrasonic signal and various echoes are received. Signal-to-noise ratio is 22 dB. The first signal, received (2) was used for image buildup (indicated by the thin solid lines. Electrical pick-up signals from the transmitter are also present. (1).

amplitude distribution obtained on a Si wafer. An offset of 9 V calculated from the mean value of the raw data was subtracted. The gray scale is from 0 V (black, smallest ultrasonic amplitudes) to 0.3 V (white). The scan size is $100 \times 100 \ \mu m^2$, the scan rate was 1 Hz and 512 points per line were sampled. The repetition frequency was adjusted to be at least $1/\Delta t$, where Δt is the time interval between two image points. the horizontal lines are typical artefacts of AFM scans probably due to material transfer between tip and sample. All other features are reproducible. The corresponding topography image is shown in Fig. 4(b). The gray scale covers 1 μm of corrugation. As can be seem from an image taken with a smaller scan size ($3 \times 3 \ \mu m^2$, in Fig. 5) the smallest features which can be resolved are about 100 nm large, which is comparable to the tip diameter. We also took a 10×10-μm^2 image on a freshly cleaved mica surface and the mean ultrasonic amplitude was comparable to the one on the wafer. As the surface was absolutely flat and homogeneous, we observed no change in amplitude except noise and artifacts.

High-frequency ultrasonic vibrations can be transmitted

(a)   25 μm

(b)

FIG. 4. Images taken of a Si wafer. Signals used for imaging are (a) integrated ultrasonic amplitude, the gray scale covers 0.3 V, color of the largest amplitudes is white, (b) topography, gray scale covers 1 μm from dark (low) to light (high).

into the soft cantilever because of tip-sample interaction forces. When the tip is in contact with the sample surface it is attracted by adhesion forces, and the repulsive forces also
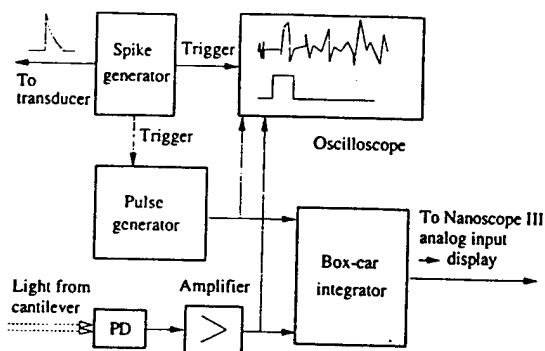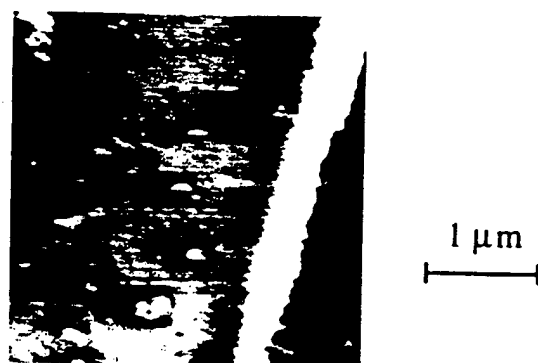

FIG. 3. Block diagram of the electronics. The signal from the photodiode is amplified by 60 dB and fed into a boxcar integrator. Its output signal is used for image buildup. The gate provided by a pulse generator is delayed such that it coincides with the first ultrasonic signal.

1 μm

FIG. 5. Details of the wafer surface showing a lateral resolution of approximately 100 nm obtained, the gray scale covers 0.2 V.

U. Rabe and W. Arnold

236

present result from the deformation of the tip and the sample surface. The tip with the cantilever can be regarded as a point mass m suspended between two fixed surfaces by a soft spring with the spring constant k of the cantilever, and a spring with spring constant k^* given by the derivative of the tip-sample interaction forces.[12] In the AFM contact mode the tip-sample interaction force is slightly repulsive. The force derivative k^* is negative and under ambient conditions larger than k ($k \approx 0.1$ N/m). Therefore, the resonance frequency ω_0' of the cantilever with the tip in contact with the surface is larger than the resonance frequency ω_0 of the free cantilever: $\omega_0'^2 = (k - k^*)/m \gg \omega^2 = k/m$. The forces increase the effective resonance frequency of the cantilever.

Apart from the wafer and mica samples mentioned above, we also carried out measurements on thin glass samples. As all samples had a thickness of less than 1 mm, the ultrasonic attenuation can be neglected and, hence, the amplitude was approximately the same for all materials. Measurements of the absolute amplitude of the detected signal would only be possible after a calibration of the knife-edge detector. However, in a separate experiment, the amplitude of the surface vibration was measured by interferometry and was approximately 5 nm on the wafer. As the scanned surface area was smaller than the ultrasonic wavelength (0.15–0.3 mm at 20 MHz), all parts of the surface region can be thought to move uniformly with the same phase and amplitude. Therefore, the changes in the detected amplitude are caused by a change of the coupling between the surface and the tip of the AFM. These differences in coupling can be due to the geometry of the surface—all edges show a decrease of the ultrasonic amplitude—a change in elastic or chemical surface properties, or even layers of adsorbate which are not well bound to the surface, as well as differences in thickness of the adsorbed water layer. A rough estimation shows that the kinetic energy which the tip and the cantilever gain during the ultrasonic vibration, might be large enough so that the tip can jump out of the water layer.

Common adhesion potentials[14] indicate that a linear approximation of the interaction force should not be sufficient for a vibration amplitude of several nm. Such large amplitudes should lead to nonlinearity in the signal. However, we could not detect such a behavior so far. This might be due to the additional adhesion caused by the adsorbed water film always present in usual laboratory air conditions, or by the large ultrasonic amplitudes employed in our measurement. Further investigations will be necessary in order to fully un-

derstand the transmission of the ultrasonic signal to the cantilever via the tip, and the image contrast obtained. It is of particular interest to clarify wether the features obtained in our images stem from local variations of elasticity, as it was observed in scanning microdeformation microscopy.[15]

In summary, we should like to stress that the monitoring of the transmission of ultrasound is an attractive method for probing of tip-sample interaction forces because of the large signal/noise ratio obtained and in order to construct acoustic microscopes yielding a resolution beyond the diffraction limit.[16,17]

Note added in proof: Recently, we became aware of research work carried out by two groups discussing the expected nonlinear transmission of ultrasound in an AFM.[18,19]

[1] See, for example, collected papers by *Proceeding of the 19th Symposium on Acoustical Imaging (1992)*, edited by H. Ermert and H. P. Harjes (Plenum, New York, 1992).

[2] G. A. D. Briggs, *Acoustic Microscopy* (Oxford University Press, Oxford, 1992).

[3] J. P. Monchalin, IEEE Trans. UFFC-33, 485 (1986).

[4] L. W. Kessler, in *Proceedings of the 18th Symposium on Acoustical Imaging (1991)*; edited by H. Lee and G. Wade (Plenum, New York, 1992).

[5] J. Heil, J. Wesner, and W. Grill, J. Appl. Phys. 64, 1939 (1988).

[6] W. Rohrbeck, E. Chilla, H. J. Fröhlich, and J. Riedel, Appl. Phys. A 52, 344 (1991).

[7] K. J. Strozewski, S. E. McBride, and G. C. Wetsel, Ultramicroscopy, 42–44, 388 (1992).

[8] A. Moreau and J. B. Ketterson, J. Appl. Phys. 72, 861 (1992).

[9] E. Chilla, W. Rohrbeck, H. J. Fröhlich, R. Koch, and K. H. Rieder, Appl. Phys. Lett. 61, 3107 (1992).

[10] K. Takata, T. Hasegawa, S. Hosaka, S. Hosoki, and T. Komoda, Appl. Phys. Lett. 55, 1718 (1989).

[11] P. Maivald, H. J. Butt, S. A. C. Gould, C. B. Prater, B. Drake, J. A. Gurley, V. B. Elings, and P. K. Hansma, Nanotechnol. 2, 103 (1991).

[12] D. Sarid, *Scanning Force Microscopy* (Oxford University Press, Oxford, 1991).

[13] R. L. Whitman and A. Korpel, Appl. Opt. 8, 1567 (1969).

[14] J. Israelachvili, *Intermolecular and Surface Forces* (Academic, London, 1992).

[15] B. Cretin and F. Sthal, Appl. Phys. Lett. 62, 829 (1993).

[16] S. Akamine, B. Hadimioglu, B. T. Khuri-Yakub, H. Yamada, and C. F. Quate, in *Proceedings of the International Conference on Solid State Sensors and Actuators*, edited by S. Middelhoek (IEEE, New York, NY, 1991), pp. 857–859.

[17] A. Kulik, J. Attal, and G. Gremaud, in *Proceedings of the 20th International Symposium on Acoustical Imaging*, edited by Y. Wei and B. Gu (Plenum, New York) (to be published).

[18] O. Kolosov and K. Yamanaka, Jpn. J. Appl. Phys. 32, 22 (1993).

[19] E. Chilla (private communications).

Sensor Physics: Signals and Noise

Thomas B. Gabrielson
Applied Research Laboratory
The Pennsylvania State University
P.O. Box 30
State College, PA 16804

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 traveled 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. Also, 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 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 roughly as f^1 (where f is 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

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

$$\ddot{x}_r + \frac{R}{M}\dot{x}_r + \frac{1}{MC}x_r = \ddot{x}_i \tag{1}$$

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

$$(\ddot{x}_i)_{ne}^2 = \frac{4RkT\Delta f}{M^2} = \frac{4kT}{M}\frac{\omega_0}{Q} \tag{2}$$

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
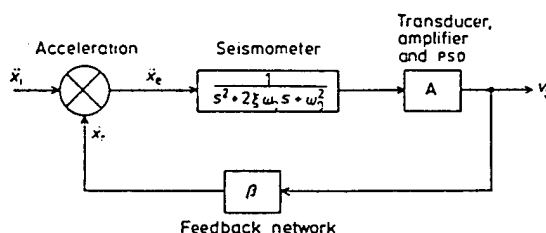

Figure 1 A force-feedback seismometer.

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

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^{-8} 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 Bremaecker 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 apparently not used in practice. Systems have been

described by Willmore (1959), Russell et al (1968) and Koleznokov 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

$$(\ddot{x}_1)_{ne} = \frac{(s^2 + 2\xi\omega_0 s + \omega_0^2)}{r} (4 R_{nv} kT\Delta f)^{1/2} \tag{3}$$

$$= \frac{\omega_0^2}{r} (4 R_{nv} kT\Delta f)^{1/2} \text{ for } \omega < \omega_0$$

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 $(\ddot{x}_1)_{ne}^2 < 10^{-19}$ (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 detector (PSD), filters, etc) negligible. This can be done by ensuring that there is sufficient gain prior to each element.

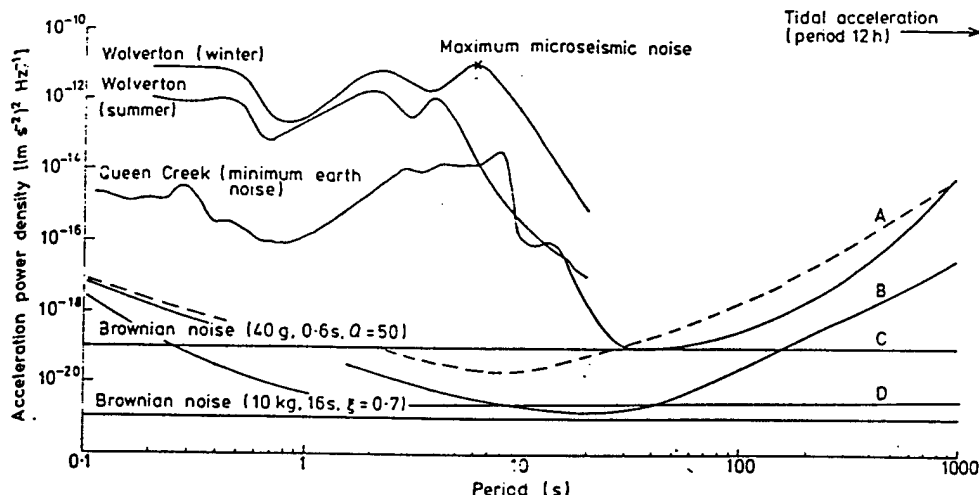Finally, to achieve the desired wideband response of defined

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
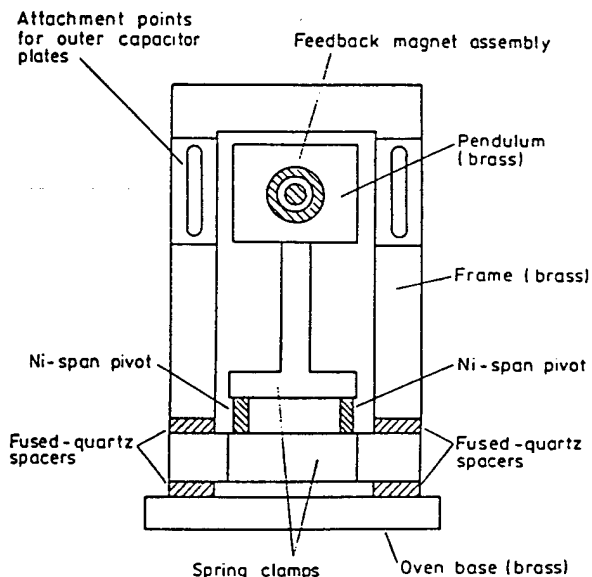

Figure 3 Diagram of seismometer with heater and cover removed.

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 an
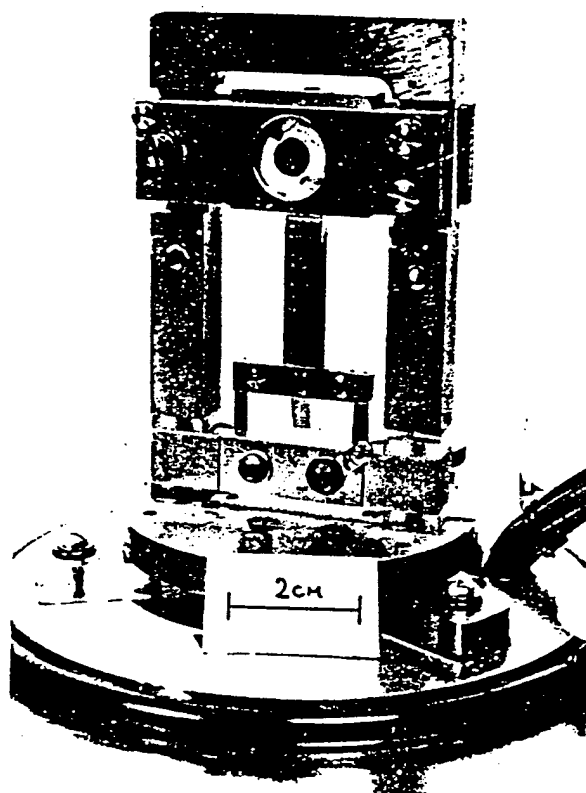
M J Usher, I W Buckner and R F Burch

Figure 4 Photograph of seismometer with heater and cover removed.
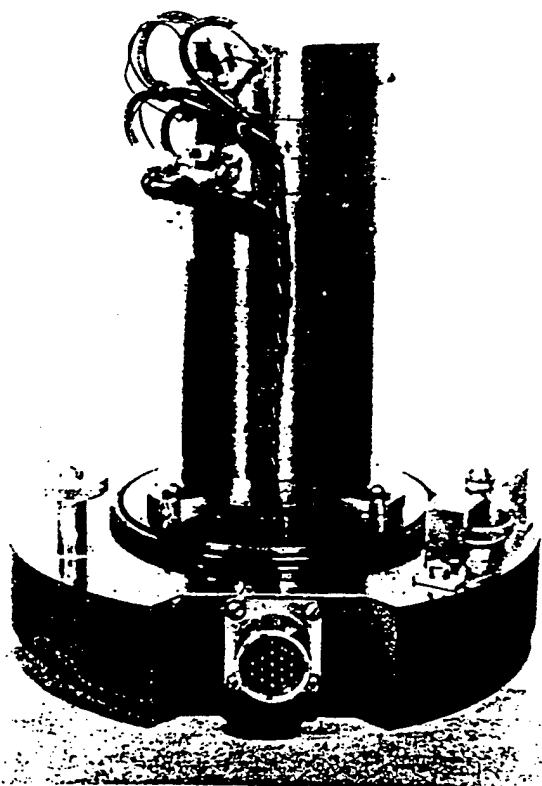

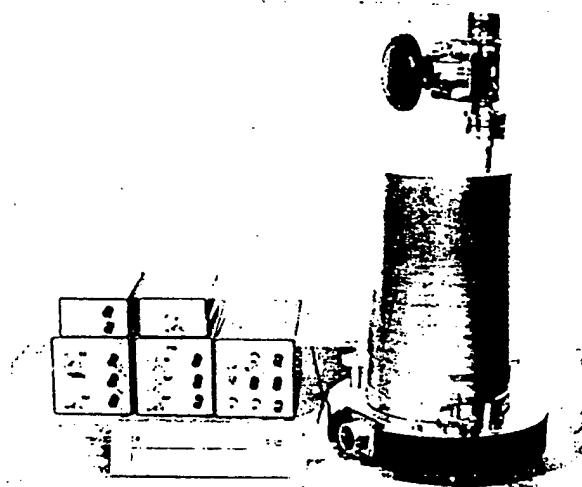Figure 5 Photograph of seismometer showing heater and internal electronics.

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 r_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.
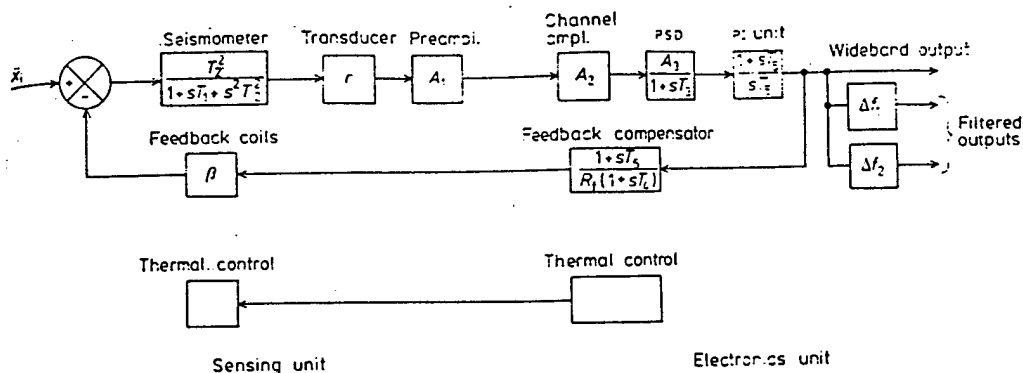
245

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
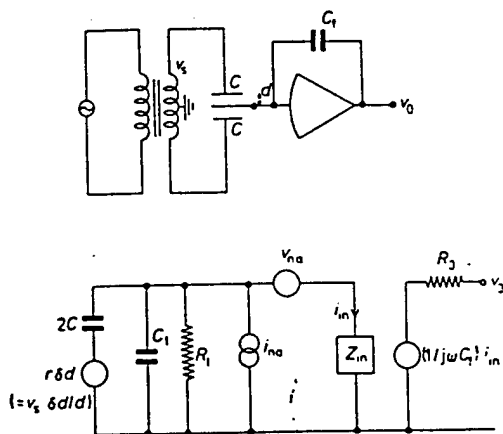

Figure 8 The Blumlein bridge and its equivalent circuit.

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.

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

$$R_n = R_{nv}\left[1+\left(\frac{C_1+C}{C}\right)+\frac{1}{\omega C R_1}\right]^2 + \frac{1}{\omega^2 C^2}\left(\frac{1}{R_{nl}}+\frac{1}{R_1}\right)$$

where R_{nv} and R_{nl} are the noise-equivalent resistances of the input device, which was a low-noise FET. Using the values given above and assuming that $R_{nv} \approx 500\Omega$, $R_{nl} \approx 10$ 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 $R_n = 1·5$ 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_{nl}. 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.

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 ± 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

$$K = \frac{A_1 A_2 A_3 r \beta}{\omega_0^2 T_5 R_f}, \qquad T_1 = \frac{2\xi}{\omega_0}, \qquad T_2^2 = \frac{1}{\omega_0^2};$$

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}, $\beta = 1·4$ m s^{-2} A^{-1}, $\omega_0 = 10$ 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 $1/\beta C_6 = 2·4 \times 10^5$ 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 $\xi = 0·7$. 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 $(\ddot{x}_i)_{ne} \approx 3 \times 10^{-10}$ m s^{-2} Hz^{-1}
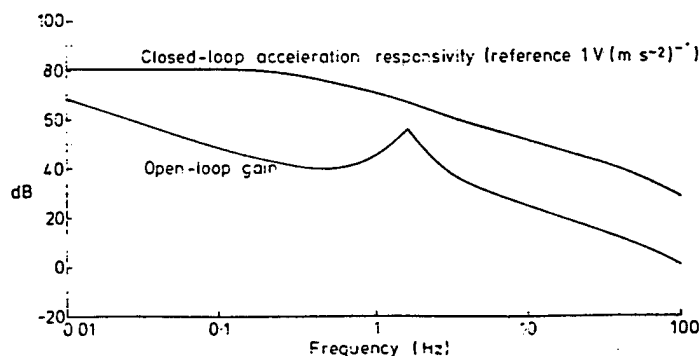

Figure 9 Open- and closed-loop responses of the feedback seismometer. The open-loop gain is greater than 40 dB over most of the seismic range (0·01 to 10 Hz). The closed-loop response is flat to acceleration from DC to 0·35 Hz (responsivity 10^4 V (m s^{-2})$^{-1}$) and flat to velocity from 0·35 to 100 Hz (responsivity $2·4 \times 10^5$ V (m s^{-1})$^{-1}$).

for the Brownian contribution. The transducer/amplifier contribution with $R_n = 1·5$ kΩ and a natural period of 0·6 s gives $(\ddot{x}_1)_{ne} = 5 \times 10^{-11}$ m s^{-2} 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 3×10^{-6} 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 3×10^{-10} 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} 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-
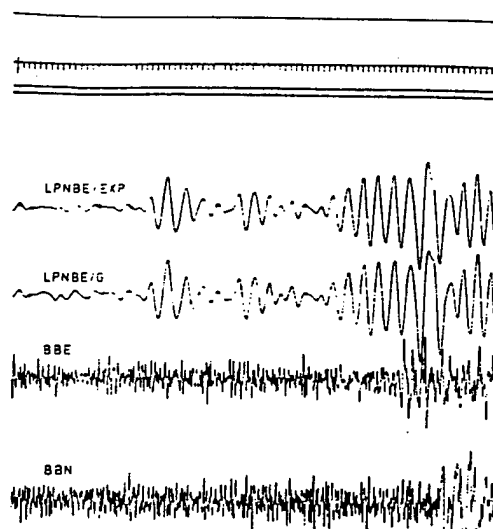

Figure 10 Comparison of seismometers (small event). A small event is shown at a magnification of 8200 at period 20 s using a long-period narrow-band (LPNB) response. LPNBE/EXP is from the feedback seismometer in the east-west orientation and LPNBE/G from the standard Geotech instrument. Two broadband records are also shown. (Event recorded at 03.30.00, 1 October 1975; scale, 8 s/division.)
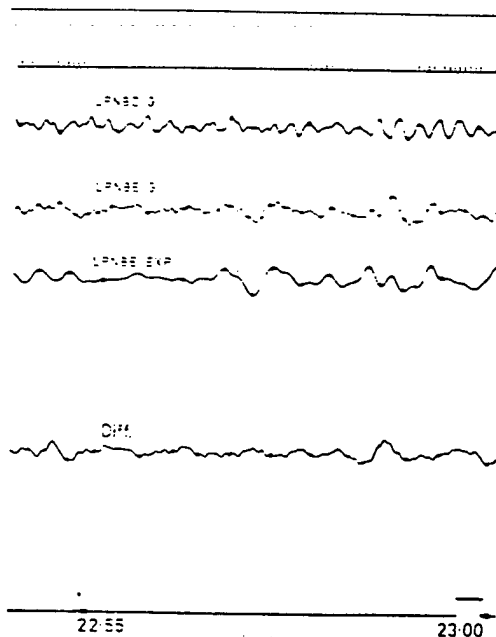

Figure 11 Comparison of seismometers (long-period noise). Long-period noise is shown at a magnification of 200 000 at 20 s using the LPNB response. LPNBZ/G is from a vertical-component Geotech instrument, and LPNBE/G and LPNBE/EXP are from the horizontal-component Geotech and feedback instruments. (Recorded at 22.55.00, 30 September 1975; scale, 4 s/division.)

248

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 ((0·6)^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.

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.

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

$$V_1 = ai_1 + bi_2, \tag{1}$$

$$V_2 = ci_1 + di_2.$$

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]

$$b = \pm c. \tag{2}$$

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

$$V_1 = ai_1 + bi_2,$$

$$V_2 = bi_1 + ci_2. \tag{3}$$

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

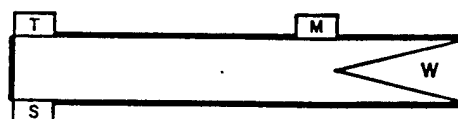S_s = speaker output in pressure produced at the microphone divided by volts across the speaker.

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.
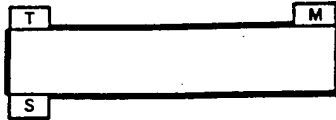
FIG. 2. Same as Fig. 1 except that the nonreflecting termination is replaced by a rigid termination. All transducers are right at the rigid terminations.

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

$$M_o = (V_2/i_1 Z)^{1/2} \times 10^{-7/2}, \tag{5}$$
$$S_o = (V_2 Z/i_1)^{1/2} \times 10^{-7/2}.$$

(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

$$V_1 = ai_1 + bU_1,$$
$$0 = bi_1 + cU_1. \tag{6}$$

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.
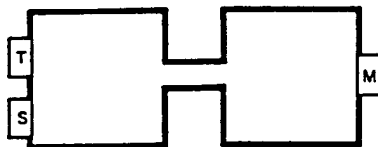

FIG. 3. A double cavity Helmholtz Resonator with three transducers, (T), (M), and (S). The volume of each cavity is V. The length of the neck is l and its cross-sectional area is A. Its angular resonant frequency ω is given by $\omega = C(2A/lV)^{1/2}$, where C is the sound velocity.
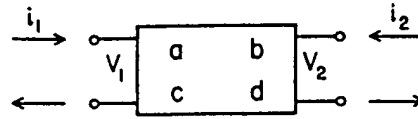

FIG. 4. A linear passive four pole network.

For the transducer (2) being used as a short circuited microphone we can write

$$0 = ai_2 + bU_2,$$
$$p_2 = bi_2 + cU_2, \tag{7}$$

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

$$M_o = \left(\frac{1}{Z} \frac{V_T}{i_T} \frac{V_{M_1}}{V_{M_2}}\right)^{1/2} \times 10^{-7/2},$$
$$\tag{10}$$
$$S_o = \left(Z \frac{V_T}{i_T} \frac{V_{M_1}}{V_{M_2}}\right)^{1/2} \times 10^{-7/2},$$

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-
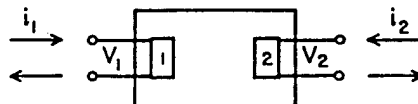

FIG. 5. Transducers 1 and 2 are coupled by an acoustic field.

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

Fig. 1: $Z = \rho C / A$, (11)

Fig. 2: $Z = 2\rho C Q_n / A n \pi$, (12)

where Q_n is the quality factor of the nth resonant harmonic:

Fig. 3: $Z = \rho C^2 Q / 2 V \omega$, (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

[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.

J. Acoust. Soc. Am., Vol. 63, No. 6, June 1978

252

A PROBLEM IN BROWNIAN MOTION

By G. E. Uhlenbeck and S. Goudsmit

Department of Physics, University of Michigan

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^{-6} 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:

$$\overline{\delta_p{}^2}=\frac{16}{\pi}\cdot\frac{1}{n}\cdot\frac{1}{\bar{c}\Delta t\Delta o}$$

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 being 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:

$$\omega_k = 2\pi k/\tau$$

$$\left. \begin{array}{l} A_k = \dfrac{2}{\tau}\displaystyle\int_0^{\tau} T(t) \cos \omega_k t \cdot dt \\[3mm] B_k = \dfrac{2}{\tau}\displaystyle\int_0^{\tau} T(t) \sin \omega_k t \cdot dt \end{array} \right\} \tag{9}$$

We can now replace these integrals by the following sums:

$$A_k = \frac{2}{\tau} \sum_1^z T(t_i) \cos \omega_k t_i \cdot \Delta t_i$$

$$B_k = \frac{2}{\tau} \sum_1^z T(t_i) \sin \omega_k t_i \cdot \Delta t_i \tag{10}$$

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 ith 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:

$$\phi(t) = \sum_k \phi_k(t) = \sum_k \frac{1}{(\omega^2-\omega_k^2)^2+r^2\omega_k^2}[\{A_k(\omega^2-\omega_k^2)-B_k r\omega_k\}\cos\omega_k t \\ + \{A_k r\omega_k+B_k(\omega^2-\omega_k^2)\}\sin\omega_k t] \tag{11}$$

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:

$$\overline{T^2(t)} = 4m\bar cp/I\rho\Delta t. \tag{16}$$

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:*

$$\overline{\phi_k^2} = \frac{\pi^{1/2} m^{1/2} (8kT)^{3/2} \cdot p}{\rho I \cdot \tau} \cdot \frac{1}{\pi kT(\omega^2 - \omega_k^2)^2 + 32p^2\rho^2\omega_k^2} \tag{18}$$

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 kth 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 kth 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):

$$\overline{\phi^2} = \sum_{k=0}^{\infty}\overline{\phi_k^2} = \frac{2}{\tau}\overline{T^2(t)} \cdot \Delta t \sum_{k=0}^{\infty}\frac{1}{(\omega^2 - \omega_k^2)^2 + r^2\omega_k^2}. \tag{20}$$

The last sum we now replace by an integral, substituting:

$$x_k = \omega_k/\omega = 2\pi k/\tau\omega$$

which gives:

$$\sum_{k=0}^{\infty}\frac{1}{(\omega^2 - \omega_k^2)r^2 + ^2\omega_k^2} \simeq \frac{\tau}{2\pi\omega^3}\int_0^{\infty}\frac{dx}{(1 - x^2)^2 + (r^2/\omega^2)x^2}$$

257

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]

$$W\Delta t\Delta o\Delta G = \frac{\pi}{2}\cdot\frac{1}{(2\pi mkT)^{3/2}}\cdot\frac{kT}{V}\cdot G e^{-G^2/8mkT}\Delta t\Delta o\Delta G. \tag{23}$$

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:

$$\overline{G^2(t_i)} - \overline{G(t_i)}^2 = \overline{(\Sigma_j G_j n_{ij})^2} - (\overline{\Sigma_j G_j n_{ij}})^2 = \Sigma_j G_j^2(\overline{n_{ij}^2} - (\overline{n_{ij}})^2) \tag{24}$$

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:

$$\overline{n_{ij}^2} - (\overline{n_{ij}})^2 = \overline{n_{ij}} \tag{25}$$

and $\overline{n_{ij}}$ 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:

$$\overline{G} = (2\pi mkT)^{1/2}$$

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:

$$M_1 = px\Delta o - m\bar{c}Nux\Delta o V^{-1}. \tag{30}$$

In the same way, we find for the moment of momentum given to Δo per second by molecules striking from the right:

$$M_2 = -px\Delta o - m\bar{c}Nux\Delta o V^{-1} \tag{31}$$

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.

SINGLE BUBBLE SONOLUMINESCENCE

ANTHONY ATCHLEY
Graduate Program in Acoustics
The Pennsylvania State University
State College, PA

Abstract

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 optical spectrum is broadband and increases in intensity in the ultraviolet. The duration of the SL pulse is between approximately 50 and 350 ps, depending upon the experimental parameters, but independent of optical wavelength. The emission can be extremely stable and repetitive, lasting for hours, or it can be chaotic. 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 substantially isotropic and unpolarized. Water is the most sonoluminescence-friendly host liquid, although SL has been observed in other liquids. SL has been observed from hemispherical bubbles attached to solid surfaces. 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.

The following references are suggested. Additional material may be found in the references provided in these articles.
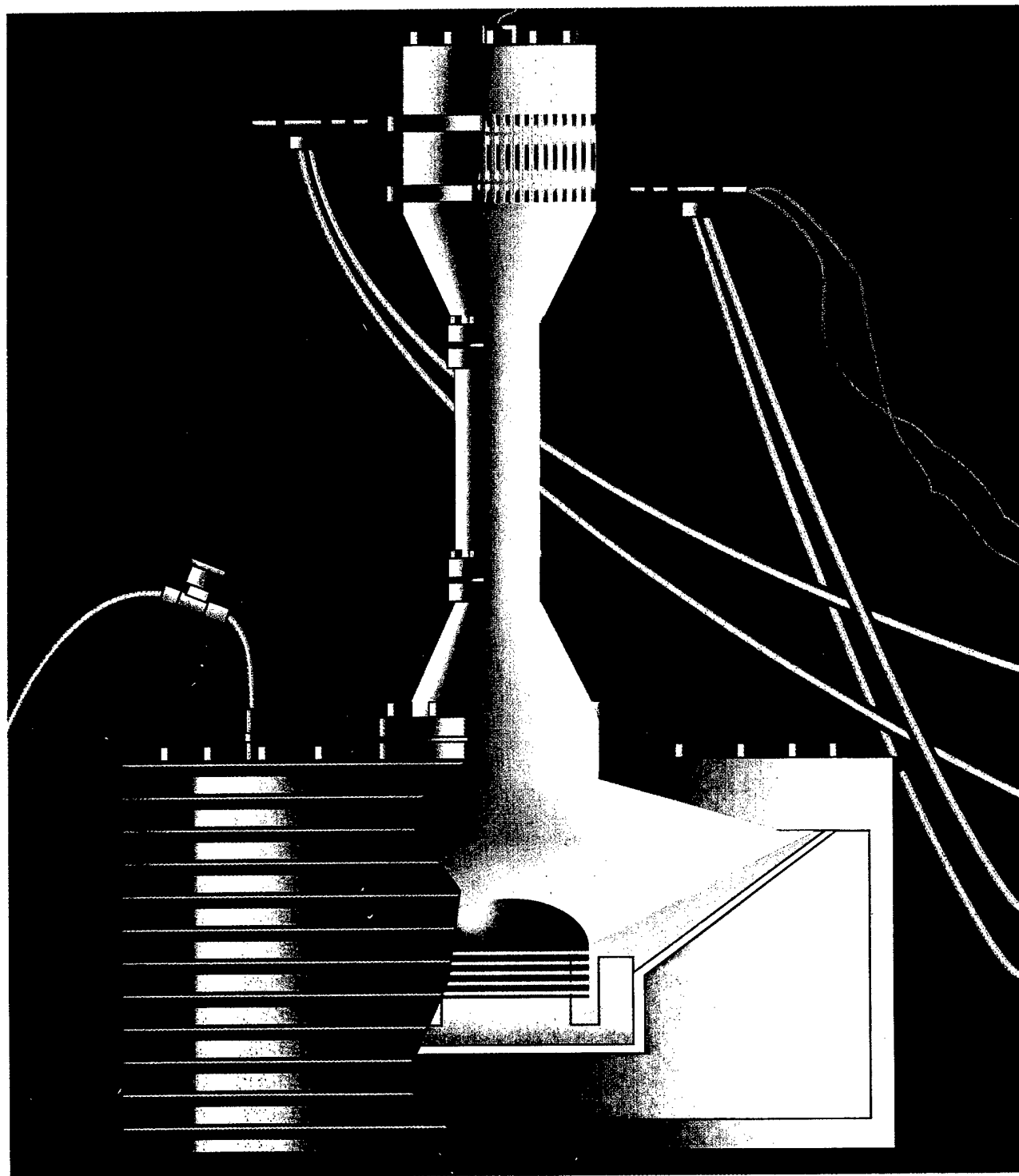
1. B. P. Barber, R. Hiller, R Loefstedt, and S. J. Putterman, "Defining the unknowns of sonoluminescence," Physics Reports, Vol. 281, 65-143, 1997.

2. S. J. Putterman, "Sonoluminescence: Sound into Light," Scientific American, 46-51, February 1995.

3. L. A. Crum, "Sonoluminescence," Physics Today, 22-29, September 1994.

4. R. G. Holt and D. F. Gaitan, "Observation of stability boundaries in the parameter space of single bubble sonoluminescence," Physical Review Letters, Vol. 77, 3791-3794, 1996.

5. B. Gompf, R. Gunther, G. Nick, R. Pecha, and W. Eisenmenger, "Resolving sonoluminescence pulse width with time-correlated single photon counting," Physical Review Letters, Vol. 79, 1405-1408, 1997.

THERMOACOUSTICS

STEVEN L. GARRETT
Pennsylvania State University

PHYSICS TODAY

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.*
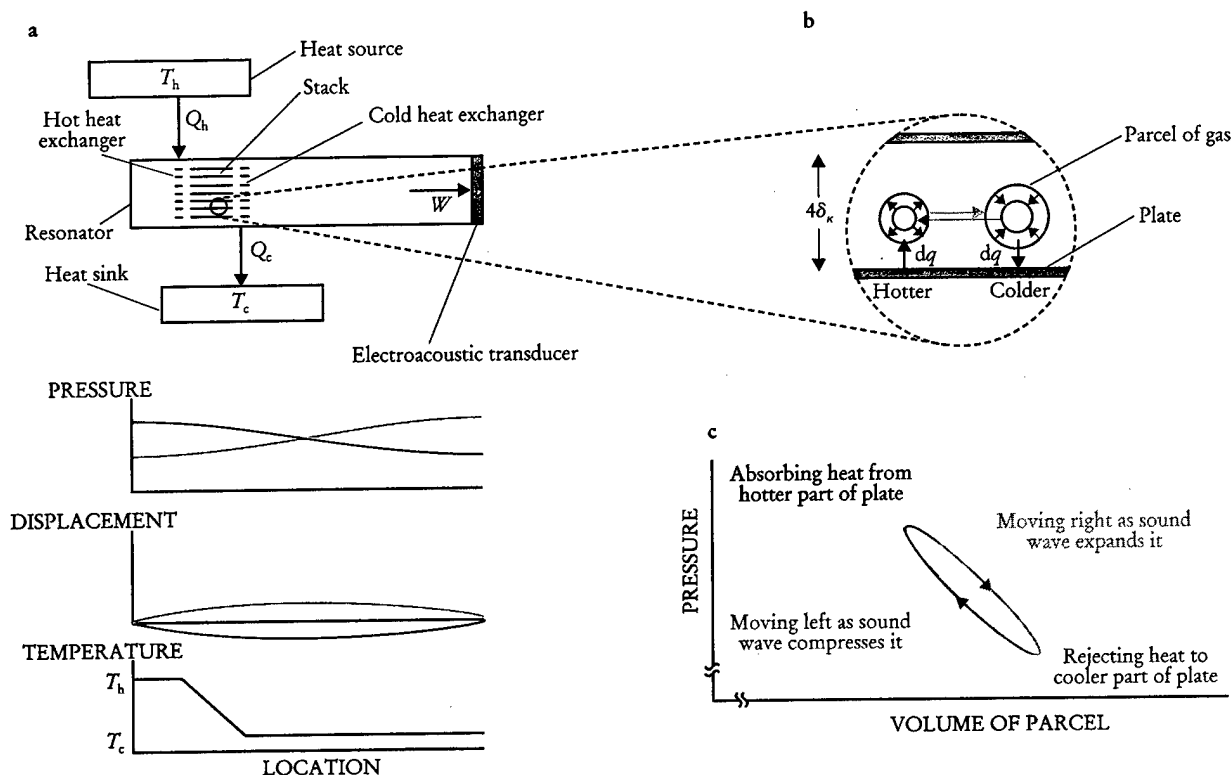
SIMPLE THERMOACOUSTIC ENGINE. a: Heat exchangers and a stack in a half-wavelength acoustic resonator convert some of the heat power Q_h from a thermal reservoir at temperature T_h into acoustic power W, rejecting waste heat power Q_c to another reservoir at T_c. The acoustic power is delivered to an 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

265
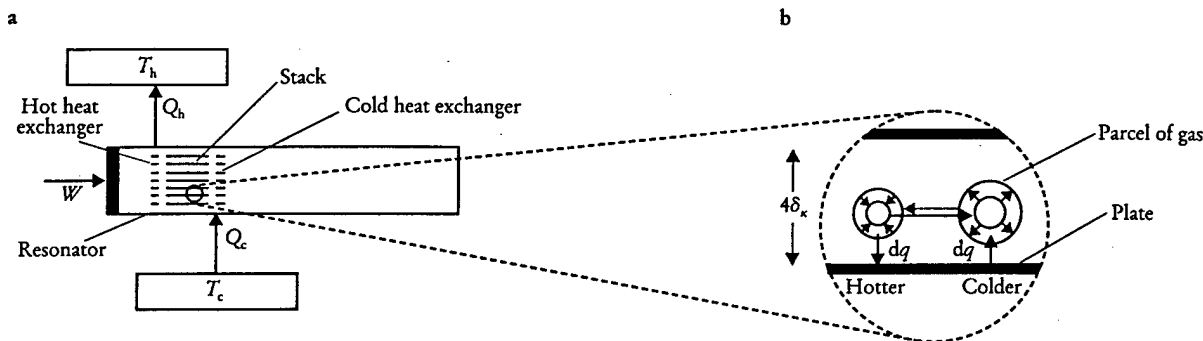
SIMPLE THERMOACOUSTIC REFRIGERATOR. a: Electroacoustic transducer at the left end delivers acoustic power W to the resonator, producing refrigeration Q_c at low temperature T_c and rejecting waste heat power Q_h to a heat sink at T_h. As in figure 1, this is a half-wavelength device with a pressure node 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.
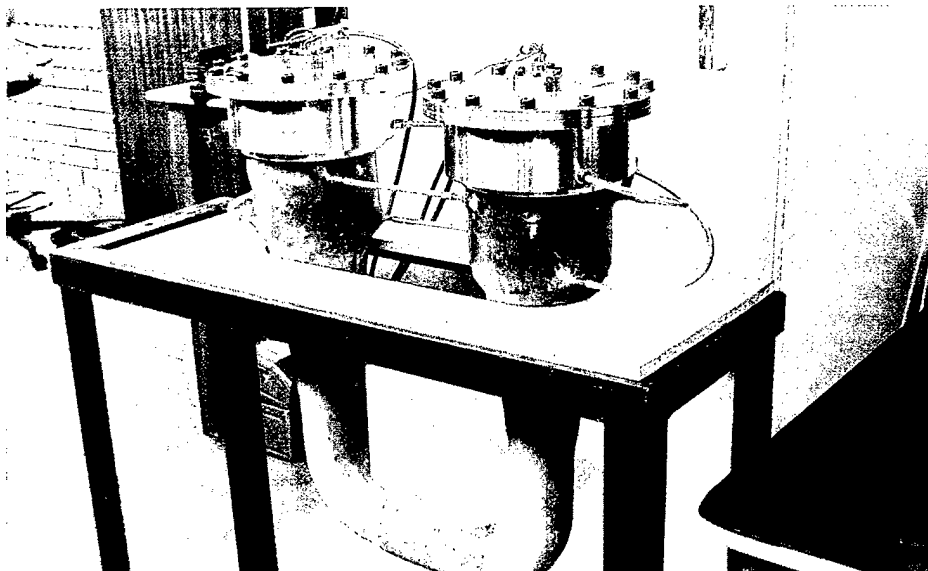
The largest thermoacoustic engine to date, under construction at Cryenco Inc in Denver, Colorado, will also drive a pulse-tube refrigerator. The heat source for the engine will be natural gas combustion. Intended for industrial and commercial liquefaction of natural gas it will measure 12 meters long, 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.
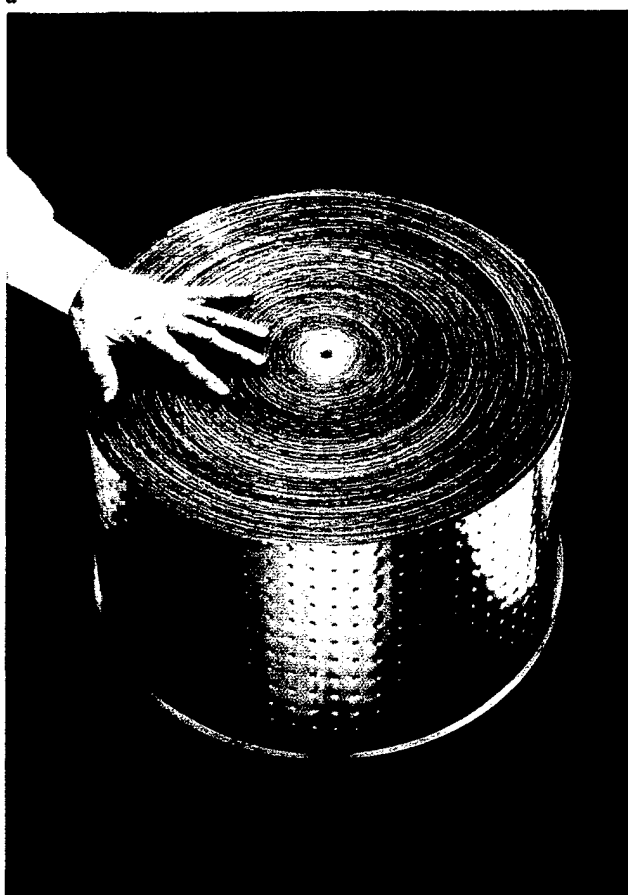
Power and efficiency

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
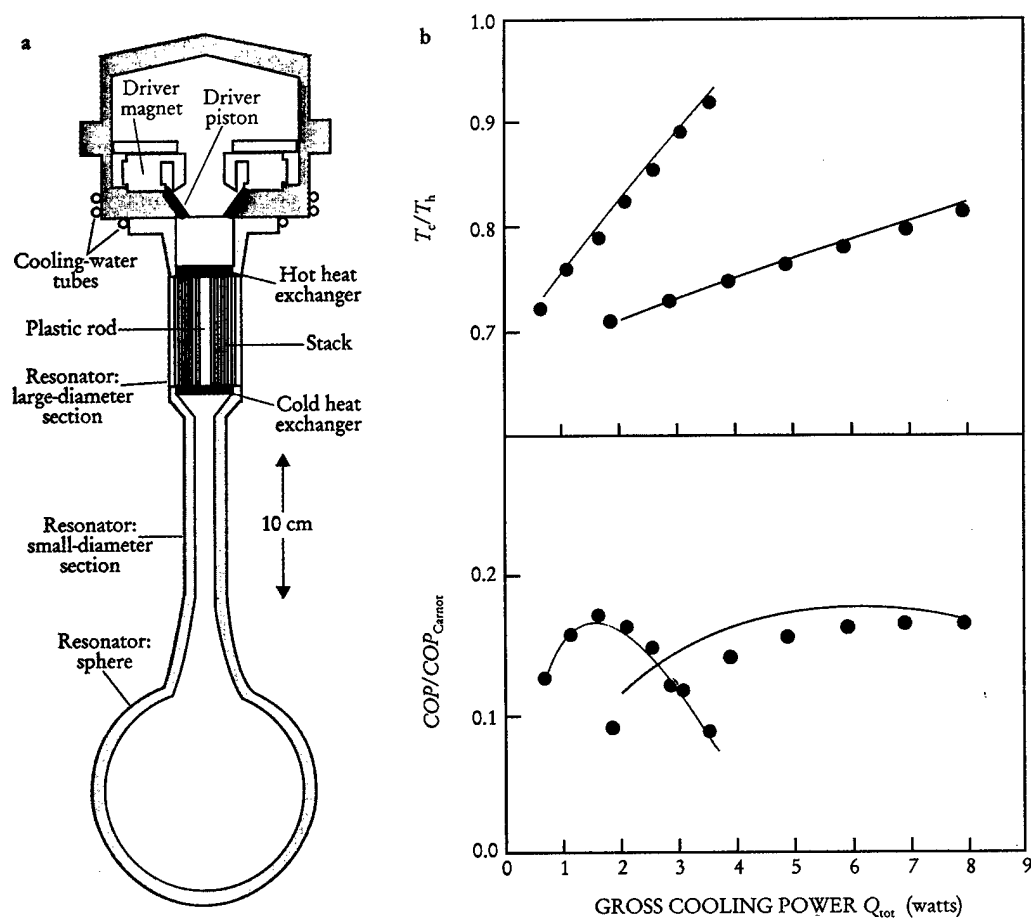
FIRST EFFICIENT THERMOACOUSTIC REFRIGERATOR (a) and some of its performance parameters (b) as measured[4] (data points) and calculated[5] (curves) for operation with 500-Hz pressure oscillations in 10-bar helium gas, and with $T_h = 300$ K. Blue circles are data for 1.5% pressure oscillations; red circles, 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

▷ 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 d$q\,\delta T/T^2$. 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.
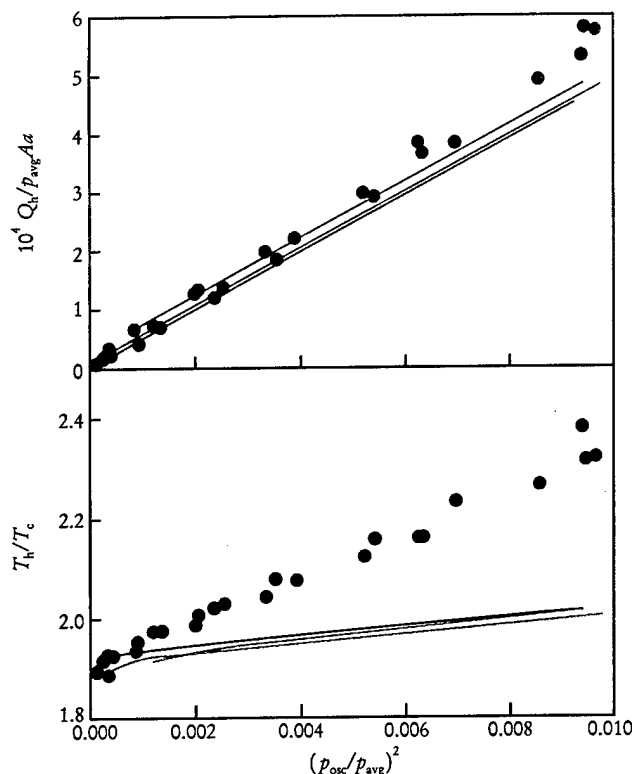
▷ Viscous loss in the stack occurs because work must be done to overcome viscous shear forces as the gas oscillates between the plates. The viscous penetration depth $\delta_\mu = \sqrt{\mu/\pi f\rho}$ (where μ is viscosity) is comparable to the thermal penetration depth, so most of the space between 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

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:

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.
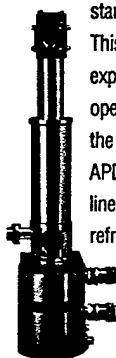
Figure 5 illustrates the amplitude dependence of the accuracy of our current understanding of thermoacoustic phenomena. The data shown as points were taken from a no-load thermoacoustic engine.[16] The plots can be interpreted as displaying the required heater power Q_h and temperature T_h at the hot 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.

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

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 15 Dec 99	3. REPORT DYPE AND DATES COVERED Final 01 Feb 97 - 21 June 98

4. TITLE AND SUBTITLE Proceedings of the 1998 Physical Acoustics Summer School: Volume III. Background Materials	5. FUNDING NUMBERS PE 61153N G N00014-98-1-0044
6. AUTHOR(S) Henry E. Bass	

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Jamie L. Whitten National Center for Physical Acoustics The University of Mississippi 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 1998 Physical Acoustics Summer School contains abstracts and copies of publications that were useful to graduate students who attended the summer school. These materials helped the students prepare for the lectures that 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, Acoustic Resonators, Chaos, Nonlinear Bubble Dynamics, Nonlinear Acoustics, Periodic, Random, and Quasiperiodic Media, Porous Media, Quantum Mechanics, Resonant Ultrasound Spectroscopy, Scanning Acoustic Microscopy, Sensor Physics, Sonoluminescence, Thermoacoustics	15. NUMBER OF PAGES 273
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT